The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review

The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review

- Golnar Karimian,
- Elena Petelos &
- Silvia M. A. A. Evers

Abstract

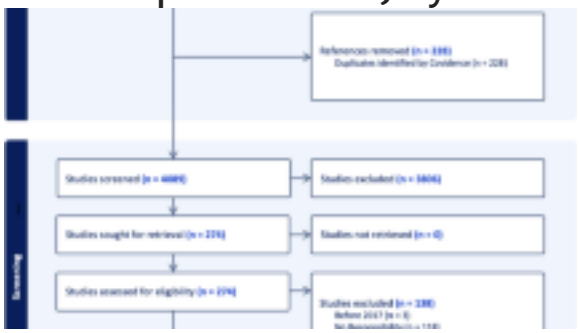
Artificial intelligence (AI) is being increasingly applied in healthcare. The expansion of AI in healthcare necessitates AI-related ethical issues to be studied and addressed. This systematic scoping review was conducted to identify the ethical issues of AI application in healthcare, to highlight gaps, and to propose steps to move towards an

evidence-informed approach for addressing them. A systematic search was conducted to retrieve all articles examining the ethical aspects of AI application in healthcare from Medline (PubMed) and Embase (OVID), published between 2010 and July 21, 2020. The search terms were “artificial intelligence” or “machine learning” or “deep learning” in combination with “ethics” or “bioethics”. The studies were selected utilizing a PRISMA flowchart and predefined inclusion criteria. Ethical principles of respect for human autonomy, prevention of harm, fairness, explicability, and privacy were charted. The search yielded 2166 articles, of which 18 articles were selected for data charting on the basis of the predefined inclusion criteria. The focus of many articles was a general discussion about ethics and AI. Nevertheless, there was limited examination of ethical principles in terms of consideration for design or deployment of AI in most retrieved studies. In the few instances where ethical principles were considered, fairness, preservation of human autonomy, explicability and privacy were equally discussed. The principle of prevention of harm was the least explored topic. Practical tools for testing and upholding ethical requirements across the lifecycle of AI-based technologies are largely absent from the body of reported evidence. In addition, the perspective of different stakeholders is largely missing.

Similar content being viewed by others

Public views on ethical issues in healthcare artificial intelligence: protocol for a scoping review

Article Open access15 July 2022



Scoping Review Shows the Dynamics and Complexities Inherent to the Notion of "Responsibility" in Artificial Intelligence within the Healthcare Context

Article 11 June 2024

Implementing Ethics in Healthcare AI-Based Applications: A Scoping Review

Article 03 September 2021

1 Introduction

Alongside the increasing use of big data, artificial intelligence (AI) is undergoing rapid growth, finding new applications across sectors, including security, environment, research, education, health and trade [1,2,3,4,5,6]. AI-based technology applications in healthcare provide new opportunities [4]. Many hospitals are using AI-enabled systems in the context of decision-support systems (DSSs) for medical staff in the context of diagnosis and treatment. AI systems also have an impact on organisational aspects for the delivery of care, as for example, to improve the efficiency of different workflows, including nursing and managerial activities in hospitals [4]. The introduction of AI is accompanied by ethical questions that need to be identified and adequately addressed in the best possible evidence-informed manner. The ethical issues surrounding AI in the field of healthcare are both broad and complex. Although AI may have the potential to improve the health of people, as well as to contribute to the resilience and the sustainability of health systems, recent analyses of the implications of AI in public health have suggested a more cautious approach to the introduction of AI in healthcare whilst more research is conducted to ensure ethical design and deployment of AI [7].

Despite the increasing rate of applying AI in healthcare, there is currently no universally accepted comprehensive framework to inform

the development and implementation of AI-based decision support in healthcare. Most critically, ethical issues that may be applicable across a spectrum of technological advances and uses of algorithms remain largely unaddressed. Intrinsic to AI are issues such as biases (e.g., poor or negative outcomes due to the use of inadequate or poor testing and training datasets for developing AI algorithms), protection of patient privacy, and gaining the trust of patients and clinicians. Key challenges for the integration of AI systems in healthcare include those intrinsic to the science of machine learning (ML), logistical difficulties in implementation, and planning encompassing due consideration of the barriers to adoption, as well as the necessary sociocultural or clinical pathway changes. All of these aspect compromise clinical applicability and relevance [8]. Developers of AI algorithms must be vigilant regarding potential dangers, including dataset shift, accidental fitting of confounders, unintended discriminatory bias, the challenges of generalization to new populations, and the unintended negative consequences of new algorithms on health outcomes [8]. It is important to build information systems that are capable of detecting unfairness and dealing with it in an adequate manner.

There are many uncertainties about the advantages and disadvantages of AI applications in healthcare, including the degree of difficulty to communicate the level of uncertainty to practitioners and patients alike. Developing trustworthy AI is of utmost importance in overcoming ethical issues of AI in healthcare and gaining the trust of the users. In Europe, legislative frameworks on key aspects such as data protection have led to regional, local, and national approaches to addressing how data are handled, i.e., the emergence of the General Data Protection Regulation (GDPR) in Europe, and a more unified approach across the European Union (EU). Given, also, the cross-border care directive (Directive 2011/24/EU) [9], it is worth examining developments in the EU region in terms of AI, particularly in relation to ethics for new technologies, including AI. Provisions regarding information systems ought to also respect the regional and national legal framework related to the Personal Data Protection (lawfulness, fairness, and transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, accountability). It is also important to consider economic development and international competition, the role of multilateral bodies and forms of global governance to determine how existing regulation

could inform convergence in terms of defining and addressing ethical challenges. There is also a clear need to consider cross-sectoral frameworks related to systems security, for example, the Network Information and Security (NIS) Directive 1148/2016 & NIS 2 Directive of the EU, and cybersecurity, as they determine cross-border collaboration within, but also beyond the EU, including ex-ante supervision in critical sectors such as health and digital infrastructure, and ex-post supervision for digital service providers.

The call of the European Group on Ethics in Science and New Technologies (EGE) to launch a process that would pave the way towards a common, internationally recognized ethical and legal framework for the design, production, use and governance of AI, robotics, and for 'autonomous' systems was another step towards developing ethical AI. The statement of EGE proposed a set of fundamental ethical principles, based on the values laid down in the EU Treaties and the EU Charter of Fundamental Right to guide its development [10]. The European Commission published a set of non-binding ethics guidelines for trustworthy AI. The core principle of this guideline is that the EU must develop a human-centric approach to AI that is in line with European values and principles [11].

The Ethics Guidelines for Trustworthy AI is a non-binding framework presented by the High-Level Expert Group on Artificial Intelligence in 2020 to the EC aiming at providing a framework to develop trustworthy AI. According to this guideline, one of the three components to achieve trustworthy AI is adherence to ethical principles and values. These ethical principles are respect for human autonomy, prevention of harm, fairness, and explicability. The guideline emphasizes the importance of paying attention to the more vulnerable groups, thorough risk assessment of AI systems and adaptive measures to mitigate the risks when appropriate. The protection of privacy is also an important aspect of trustworthy AI [11].

Several review articles have previously discussed the ethical concerns of applying AI-based technology in medicine and healthcare [12,13,14,15,16,17,18,19,20]. A recent scoping review explored the ethical issues of the application of AI in public health [7]. Currently, to the best of our knowledge, there is no systematic review (SR) that

examines qualitative and quantitative evidence about the ethical issues in healthcare. There is a clear need for a collective, wide-ranging and inclusive process that would pave the way towards a common, internationally recognized ethical framework for the design, production, use and governance of AI, robots and 'autonomous' systems. The EGE is of the opinion that Europe should play an active and prominent role in this. Overseeing the debates on moral responsibility for AI and so-called 'autonomous' technology, the EGE calls for more systematic thinking and research about the ethical, legal and governance aspects of high tech-systems that can act upon the world without direct control of human users, to human benefit or human detriment [10].

This study will focus on extracting the evidence regarding the ethical principles that have been emphasized by the Ethics Guidelines for trustworthy AI, and on extracting data regarding practical solutions for adherence to ethical principles and on stakeholder opinions. Drawing upon the Ethics Guidelines for trustworthy AI, this review aims to identify the ethical issues of AI application in healthcare, highlight gaps and propose steps to move towards an evidence-informed approach for addressing ethical issues.

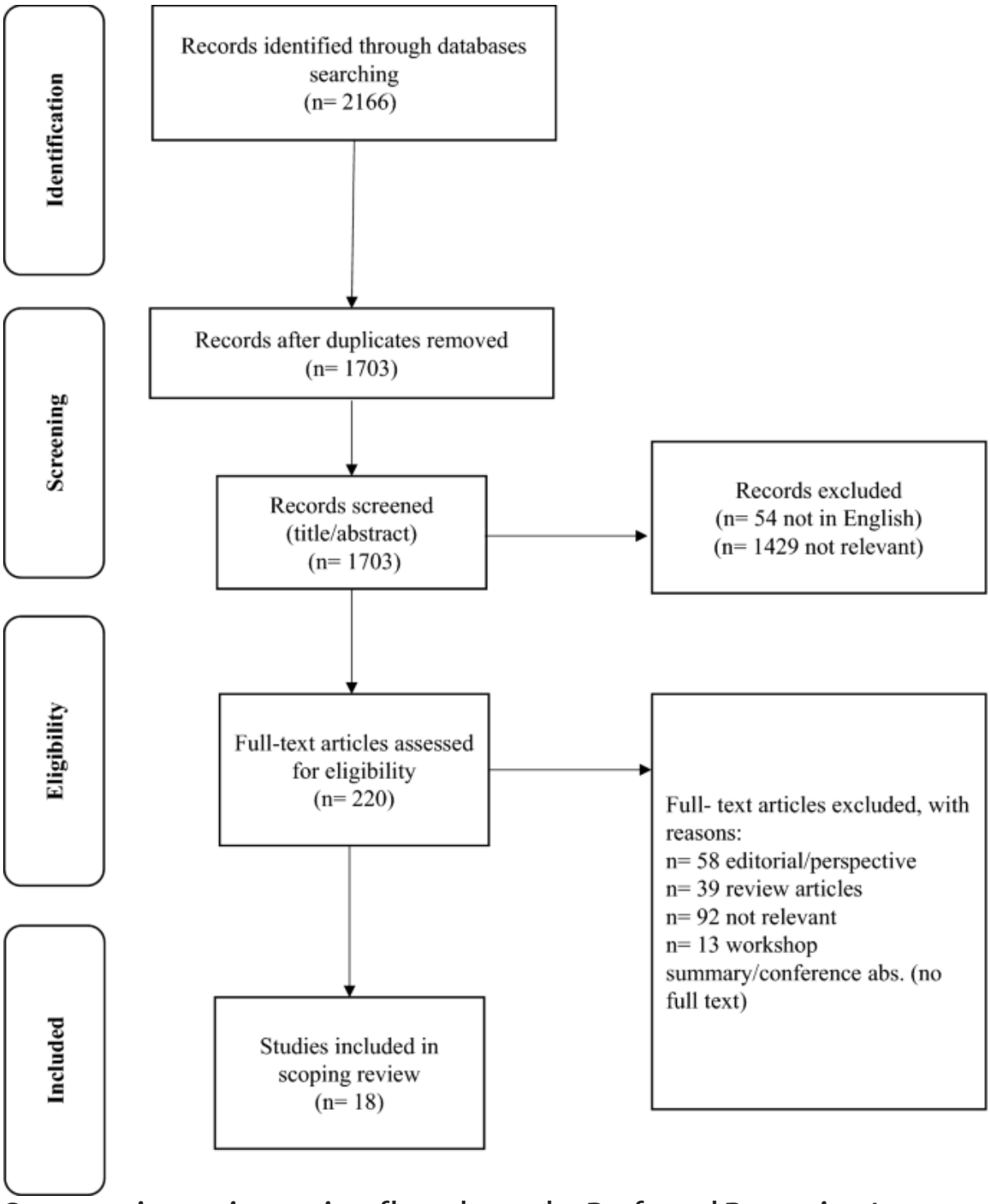
AI refers to systems that show intelligent behavior by analyzing their environment and taking actions (with some degree of autonomy) to achieve specific goals. These systems combine ML techniques, robotics, algorithms and automated DSSs [21]. ML is a domain in which a machine performs repetitive loops to improve executing a specific task. ML produces algorithms to analyze data. These algorithms can autonomously improve their performance (without the need for direct human input) by training themselves on data and learn descriptive and predictive models. ML algorithms find their patterns in the data and apply what they learn to the new data, so as to produce outcomes without reprogramming. ML is divided into unsupervised (i.e., to identify groups within data based on commonalities) and supervised methods. Supervised ML algorithm trains on specific data pairs in the form of input–output data and learns predictive models that subsequently can link new input to outputs. One of the supervised ML models is the Artificial Neural Networks (ANN), which is inspired by the neuroanatomy of the brain [22, 23]. Each computing unit acts as a neuron and all computing units are connected to build a network like

the neural network of the brain. Each input to the first layer travels through many (hidden) layers to reach the last layer and results in an output. The Deep Learning (DL) concept refers to complex neural network architecture including a variety of deep neural networks (DNN). These models apply a sequence of filters allowing the automatic detection of relevant characteristics of input data. The DL models are intrinsically dependent on the training dataset. If the training dataset does not include enough diversity or contains bias, the outputs may not be generalizable to real-life [22, 23]. Different forms of AI-based technology are currently being used in healthcare. AI in medicine can be categorized into two subtypes: virtual and physical [24]. The virtual part ranges from applications such as electronic health record systems to neural network-based guidance in treatment decisions (i.e., ML, DNN, AI-driven clinical decision support systems (CDSS), embodied AI and AI prediction algorithms). The physical part deals with robots assisting in performing surgeries, intelligent prostheses for handicapped people, and elderly care [25, 26].

2 Methods

This scoping review was conducted in five steps according to the methodological framework for scoping studies [27] and using a PRISMA flow chart (Fig. 1) as follows: identifying the research question (as described in Sect. 1), identifying relevant studies, selecting studies, charting the data, summarizing and reporting the results. The steps of the review are listed below:

Fig. 1



Systematic scoping review flow chart: the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart detailing the records identified and screened, the number of full-text articles retrieved and assessed for eligibility, and the number of studies included in the review

Full size image

3 Identify relevant studies

This study identified, retrieved, and evaluated information from peer-reviewed articles that examined the ethical aspects of artificial intelligence in healthcare from two databases, Medline (through PubMed) and Embase (through OVID). Search terms employed in database searches included 'artificial intelligence' or 'machine learning' or 'deep learning' in combination with 'ethics' or 'bioethics' (as stated in Table 2). The focus of the study was on studies published between 2010 and 21 July 2020 (the last date of database search), and, indeed most of the literature regarding AI has been published in the past 5 years. The grey literature entries were not searched, as the aim of the study was to map evidence from peer-reviewed articles.

Two independent researchers (GK and EP) searched the databases using a combination of search terms related to AI and ethics and generated an overview of retrieved output. The third author (SE) reviewed the results.

4 Selecting studies

Inclusion criteria were as follows:

- The topic is AI and/or ML and/or DNNs and/or DDS/CDDS and there is mention of ethics/ethical considerations or a dimension thereof;
- The topic should have been about the application of AI for humans in healthcare;
- Primary and secondary research, incl. qualitative and/or quantitative studies;
- Articles were written in English.

Exclusion criteria.

- Editorials and perspectives, summaries of workshops and conference abstracts;
- Articles written in languages other than English.

Regarding the criterion "primary and secondary research", this was independent of the type of research methodology utilized (i.e.,

inclusion irrespective of whether the methodology was qualitative or quantitative methodology). Some SRs that were included in the data charting did not primarily aim to investigate the ethical concerns of AI in healthcare, but some of their data addressed ethical issues in AI. However, these studies did not get a higher weight than other primary studies for data charting.

5 Charting the data

The item list used for charting data is presented in Table 1.

Table 1 Data charting list

Full size table

Information on first author, year of publication, journal, type of research (quantitative or qualitative), methodological design, study setting (i.e., country in which the study was performed or specifics of healthcare organization in which the study was carried out), participant characteristics such as specific patient group or professionals (where relevant) and the specifics of AI-based technology (i.e., ML, DNN, CDSS, AI augmentation, etc.) were charted. In this scoping review, study finding regarding the ethical principles of respect for human autonomy, prevention of harm, fairness, explicability and patient privacy were extracted. These specific principles were chosen drawing upon the Ethics Guidelines for trustworthy AI [11].

Responsible data science centers around four challenging topics: fairness, i.e., data science without prejudice; accuracy, i.e., data science without guesswork; confidentiality, i.e., data science that ensures confidentiality and transparency, i.e., data science that provides transparency [28]. Training data to inform data science approaches carries concrete potential to contribute towards better outcomes. Nevertheless, the benefit gained in terms of a fair outcome is only as good as the quality of the data used for training. In other words, cases of individual discrimination or lack of adequate representativeness, and, even, data collection conducted in a manner addressing different source of bias, may result in lack of representation of minority groups or to lead to further stigmatization of such group. Unintended discrimination and profiling also represent

important challenges, and technical and regulatory aspects ought to be carefully considered to safeguard fairness in decision-making [29].

Using mathematical notions of fairness can offer a step in this direction. There has been intense debate about the extent to which ML/AI algorithms may result in unfair discrimination, including for ethnic groups, race, gender, and demographic characteristics. Also, in terms of patient and consumer harms, oftentimes with broader societal implications. To assess whether algorithms are resulting in fair outcomes, as well as to mitigate such potential effects, various efforts have been deployed focusing on mathematical definitions of fairness. These, however, are starkly different to real-life determination of fairness, grounded in shared ethical beliefs and values. Furthermore, for decision- and policymaking to be informed by evidence and context-relevant, there need to be sound frameworks and robust methodologies to assess tradeoffs. As it is often impossible to fulfil multiple conditions at the same time, it is critical that those affected by automated decisions participate across the lifecycle of the products and solutions utilizing ML/AI algorithms, for example stakeholder input is key to ensure diversity-in-design [30].

Different definitions have been put forward that formalize fairness in AI mathematically. These can be grouped into concepts of fairness across groups or individuals [31,32,33]. Fairness in AI is violated by so-called biases. Bias can be defined as a systematic deviation of an estimated parameter from true value. Biases can emerge along the complete AI pipeline namely with regard to (1) data, (2) modeling, and (3) inadequate applications [31,32,33]. To prevent harm and to respect the principle of fairness, it is important to validate AI algorithms correctly and to ensure that the outcomes are sufficiently reliable and generalizable to be applicable in clinical practice. Also, the accuracy of the AI algorithm, as well as the safety and equity of the outcomes compared to the standard of care, are all key aspects that need to be considered. Transparency of algorithms, protection of patient privacy (through data protection and data governance) and sustainability of technical robustness are important aspects to respect human autonomy, the principle of explicability and prevention of harm [34]. Data were extracted from the articles in the form of phrases that matched these ethical concepts or were related to them. Additional information was extracted if any practical idea or tool was presented in

the study evaluating adherence to the ethical principles. Data regarding stakeholder opinions such as patients, healthcare providers or managers were also extracted (usually these groups formed the participants of the study).

6 Results

The search strategy yielded 2166 articles using the search queries that are listed in Table 2.

Table 2 Search queries employed in database searches

Full size table

When duplicates were removed, 1703 articles were left for screening. We excluded 85% of articles based on the screening of the title and the abstract. Full-text screening of the remaining 220 articles resulted in 202 articles (Appendix 1) being excluded for different reasons on the basis of predefined criteria, incl. the type of article, i.e., editorials and perspectives, review articles, lack of relevance based on inclusion criteria and summaries of workshops and conference abstracts that did not have a full text (Fig. 1). There was only one article published before 2015. This article was a literature review entitles 'Recommendations for the ethical use and design of artificial intelligence care providers''. All other studies were published after 2015. The key ethical principles were largely absent in terms of consideration for the design or the deployment of the many retrieved articles. 18 articles were selected for data charting in the scoping review. Figure 1 summarizes the PRISMA flowchart for study selection.

7 Summarizing and reporting the findings

There were 18 eligible studies based on the literature search strategy [25, 35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51]. A list of the eligible studies and the study characteristics are reported in Table 3.

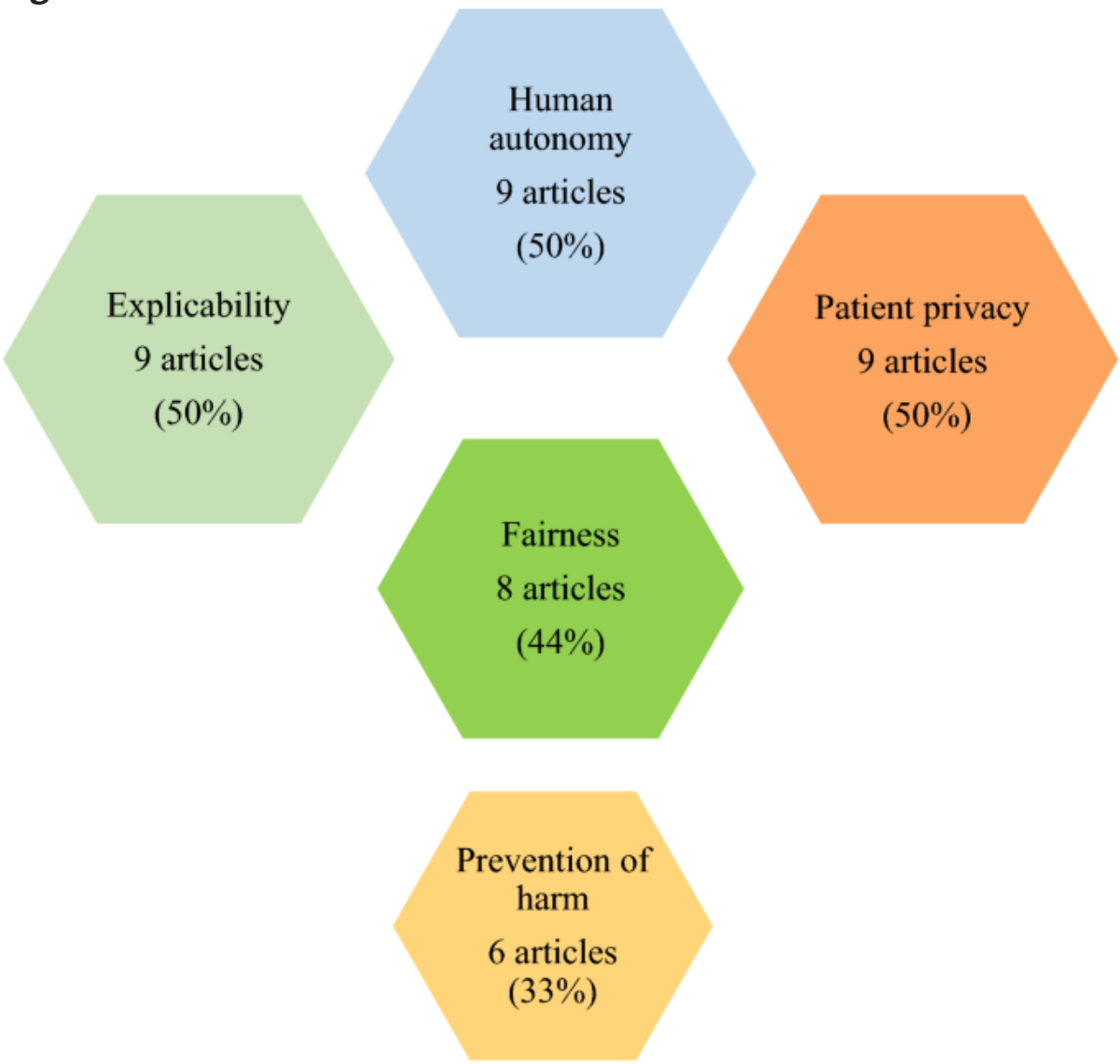
Table 3 List of eligible studies and the study characteristics

Full size table

Most of the studies were in recent years and used a qualitative research approach with different methodological design such as case analysis,

online or web-based surveys, semi-structured interviews and document analysis. The studies were performed (where stated) in a variety of settings such as primary health care, mental healthcare, ICU and community hospitals. In this scoping review, data from systematic reviews were extracted [25, 37, 41, 48, 51] and not from the primary source of the studies that were included in these systematic reviews, because the primary aim of these reviews were not ethical concerns of AI in healthcare. However, some of their data addressed the ethical issues of AI in healthcare. Study participants were healthcare professionals, care providers, some groups of patients and health informatics leaders .

Fig. 2



Map of outcomes (i.e., ethical principles) measured by the number of studies in the scoping review

Full size image

The study-specific findings regarding the ethical principles of AI are listed in Table 4.

Table 4 Extracted data regarding ethical principles

Full size table

Preservation of human autonomy when applying AI in healthcare was one of the most frequent ethical principles to have been discussed in the literature. The concerns regarding this principle were the lack of patient centeredness [36], the lack of shared decision making, [38] and the potential to ignore individual preferences when applying AI-driven DSS/CDSS, as an AI-tool or algorithm may have been based on limited epidemiological evidence and may not correspond to the specific population characteristics in other settings than the ones from which data were collected for said algorithm and/or, indeed, may not match individual patient needs, but also wishes and preferences [42].

The principle of explicability of AI algorithms was also frequently discussed in the literature [35, 38, 41, 45, 46, 48,49,50,51]. The black-box issue referring to the complexity of algorithms was considered one of the major roadblocks in the preservation of explicability of AI algorithms in healthcare as it would hinder comprehensive communication between the healthcare providers and patients regarding the advantages and disadvantages of following an AI-driven decision.

The principle of patient privacy was another frequent topic of research in the literature [25, 39,40,41, 43,44,45,46, 48]. Given that AI-based technology needs training data to provide outputs, the willingness of individuals to share health data is an essential precondition to the successful development of AI tools. In one study [45], authors argued that meaningful individual control of data calls for new models of control. This could be implemented by regarding patients as co-managers of their data and of the processes into which their information is channeled, rather than regarding patients as mere data subjects whose data can be analyzed under the GDPR in the EU. In another study [43], the current regulations regarding the principle of patient privacy were explored further by interviewing health

information management leaders and with a systematic review of the literature. According to this article, the protection of privacy and confidentiality of health information in the USA is subjected to the Health Insurance Portability and Accountability Act (HIPAA) [52], which allows for sharing protected health information without patient consent specifically for the purposes of "treatment, payment and operations". However, in the United Kingdom, governmental regulations are stricter and patient consent must be obtained prior to sharing information with any third party that is not in a direct care relationship with the patient. Researchers need to obtain permission from Health Research Authority's Confidentiality Advisory Group (CAG) to access confidential patient information without patients' consent. In another study [40], when a group of meningioma patients (a neurological disease) and their caregivers were interviewed about the ethical issues of using AI in healthcare research, some thought that loss of privacy was an acceptable sacrifice for the greater good.

Different studies discussed the principle of fairness [25, 41, 42, 45,46,47,48, 51]. In these studies, various type AI-based technology was studied including AI-driven clinical decision support systems, AI prediction models, Embodied AI and ML algorithms in medical diagnosis. These studies suggested that bias in the development phase of an algorithm could lead to discrimination, lack of equity, lack of diversity inclusion and lack of just provision of care. Embedded unconscious bias by CDSS could lead to unfair care of patients [42] and unfair predictions based on factors such as race, gender and the type of patient health insurance policy. In a retrospective cohort study [47], the authors tested two AI prediction models retrospectively in two independent cohort data sets in a psychiatric unit and an intensive care unit (ICU) to investigate whether any ethical issues would arise. Their study showed that when an AI model was used to predict the rate of readmission in the psychiatric unit, it would lead to significant bias due to including the type of patient insurance policy as one of the factors in the prediction model. Similarly, when another AI model was used to predict the rate of mortality in ICU, there was a significant difference when race, gender or the type of patient insurance policy were included in the prediction model. This study demonstrated how some unforeseen factors during the development phase of an AI algorithm could lead to discrimination. The development of discrimination conscious

algorithms was suggested to be beneficial in reducing bias and prejudices in healthcare [51].

The principle of prevention of harm was the least explored topic among the literature [25, 48, 49, 51].

Ethical principles as outcomes are depicted in Fig. 2. They were mapped by measuring the number of studies referencing them in the scoping review. Looking at the application of AI in healthcare, the principles of fairness, preservation of human autonomy, explicability and patient privacy were equally the most frequent ethical aspects that were studied in the literature as is shown in the map Study participants were healthcare professionals, care providers, some groups of patients.

Looking at the evidence regarding stakeholder opinions about the ethical issues of AI application, only two articles were exploring the opinions of patients about the ethical issues of applying AI in healthcare [37, 40]. One article was a modified scoping review focusing on identifying the evidence regarding AI ethics and disabled patients. This study showed that these vulnerable groups are largely underrepresented in the discussion about AI and ethics [37]. The other article [40] was a qualitative interview with a group of meningioma patients and mostly focus on the application of AI in healthcare delivery and research. This group of patients were very accepting of the possibility of errors using AI-based technology and loss of privacy for the greater good. Five articles focused on the opinions of healthcare providers regarding ethical issues of AI in healthcare [35, 36, 38, 40, 42] and only one article focused on health informatics managers [43].

Looking at the evidence regarding practical methods for evaluating adherence to ethical principles, it was clear that this information is largely missing from the evidence. Despite the frequent discussion of the principle of privacy in the literature, only two studies presented practical methods for AI-developers to preserve patient privacy while developing AI- algorithms [39, 44].

8 Discussion and conclusion

The aim of this study was to provide an overview of the current body of evidence regarding ethical issues of AI, to identify the ethical issues of AI application in healthcare, to highlight gaps and to propose steps for moving towards an evidence-informed approach for addressing ethical issues. This study focused on extracting the evidence regarding the ethical principles that have been emphasized by the Ethics Guidelines for trustworthy [11] AI, as well as data regarding practical solutions for adherence to ethical principles and stakeholder opinions.

The current published literature about the ethical issues in applying AI-based technology in healthcare showed the principles of fairness, preservation of human autonomy, explicability and patient privacy were equally the most frequent ethical aspects that were discussed. The principle of prevention of harm was the least researched issue in the literature. Similar to our study, another scoping review about the application of AI in public health [7] highlighted a number of common ethical concerns related to privacy, trust, accountability, and bias in AI application in healthcare in public health.

One of the limitations of this scoping review is that it is limited to the date of the last database search in July 2020, however, the field of AI-based technology and its application in healthcare is expanding. In the coming years, new evidence may be generated on the topics discussed in this review. Another limitation in this study is that we did not specifically extract data regarding the geographical location where the study was performed. Therefore, it is not clear whether these data are only representative for the high-income countries or also the low- and middle-income countries (LMICs). LMICs also present additional challenges in terms of infrastructure and ability to inform patients, communicate uncertainty, administer consent, and generate robust data, therefore, further ethical considerations may apply. Another scoping review published in the beginning of 2021, indeed, highlights the critical need for exploring the ethical implications of AI within LMICs [7].

The literature identified in this scoping review places emphasis on the general discussion about the ethical principles of AI in healthcare, whilst identifying the ethical problems that are linked with different types of AI-based technologies (such as AI-driven clinical decision

support systems, AI prediction models, Embodied AI and ML algorithms in medical diagnosis). However, practical methods or frameworks for testing whether an AI-based technology upholds ethical principles were largely missing in the current published literature. Ethics Guidelines for trustworthy AI suggests a checklist with seven key requirements for trustworthy AI including (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination, and fairness, (6) environmental and societal well-being and (7) accountability. However, evidence about the application of these requirements when applying AI in healthcare was largely missing from the published literature. Only a few articles provided a practical method to solve the issue of patient privacy in healthcare [39, 44]. One study called for meaningful individual control of data by regarding patients as co-managers of their data and the processes into which their information is channeled [45]. However, this would imply all patients have adequate literacy, and information to be able to counter information asymmetry aspects and to ensure they have a full understanding of both the purposes their data would be used for and the processes their data were undergoing. This suggestion calls for approaches to improve patient literacy in information technology and data science.

There were no practical tools or frameworks for testing whether an AI technique upholds the principle of fairness, prevention of harm, human autonomy or explicability.

The development of trustworthy AI in healthcare and implementing the wide application of such AI in healthcare requires thorough identification of different stakeholders, understanding their point of view regarding the ethical issues of AI, and capturing their needs, wishes and preferences. This scoping review shows that there is a gap in the current evidence about the stakeholder perspective on these issues, particularly as there were very few articles approaching a limited number of healthcare providers and patients. Other groups of stakeholders in healthcare such as regulatory authorities, healthcare facility managers, AI developers and vulnerable groups of people were not considered in the current evidence. Similar to our findings, another scoping review [7] showed that those leading the discussion on the ethics of AI in health seldom mentioned engagement with the end-

users and beneficiaries whose voices they were representing. Interestingly, some end-users of AI (i.e., patients) may give a lower weight to the ethical problems such as errors and loss of privacy than expected [40] in the discussion about AI ethics in healthcare.

In conclusion, AI-based technology is expanding rapidly and is applied in many areas of health care. The ethical aspects and issues of applying AI are extendedly discussed in the literature, however, no universally accepted comprehensive framework for ethical AI development and implementation in healthcare has been developed. Ethics Guidelines for trustworthy AI is a comprehensive but non-binding framework that provides guidance and recommendations for AI developers. However, the lack of practical methods to test AI-based technologies for their ethical function and implementation is very clear. In addition, several groups of stakeholders, such as vulnerable populations, healthcare providers and managers of healthcare organizations are clearly underrepresented in the current discussion about ethical issues of AI application in healthcare.

9 Suggestions for further research

Given the lack of practical methods or frameworks to test for adherence to upholding ethical principles during the whole life cycle of AI-based technology in healthcare, further research to develop such practical tools are needed in the future.

In addition, further research is needed to identify different stakeholders, users and beneficiaries of AI-based technology in healthcare and to engage with them in a discussion about AI ethics and practical solutions to ensure ethical AI. There is a need for interdisciplinary collaboration between different stakeholders, regulatory and legislative authorities to ensure ethical AI, transparency in implementation and timely reporting of both best practice and lessons learned, and to inform how to best built sound governance models to support the implementation of harmonized ethical frameworks. Due consideration for public health and global health implications is also needed in terms of research priorities, to address equity and diversity challenges.