International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE IN MEDICINE: A SYSTEMATIC REVIEW
DOI	https://doi.org/10.31435/ijitss.2(50).2026.5652
RECEIVED	10 February 2026
ACCEPTED	29 May 2026
PUBLISHED	11 June 2026
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE IN MEDICINE: A SYSTEMATIC REVIEW

Martyna Musiał *(Corresponding Author, Email: martynamusial2001@gmail.com)*
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0009-0001-5324-0839

Anna Matusik
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0009-0003-4390-4528

Kamila Kasprzycka
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0009-0001-5915-2047

Monika Spaczyńska-Kwiatkowska
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0000-0003-2660-8674

Wiktoria Niedziela
University of Opole, Opole, Poland
ORCID ID: 0009-0007-9140-9328

Zuzanna Mielniczek
Medical University of Silesia, Katowice, Poland
ORCID ID: 0009-0006-3460-2976

Anna Szwed
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0009-0001-1689-9706

Katarzyna Słowik
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0009-0002-6951-9479

Gabriela Matoga
Medical University of Silesia in Katowice, Katowice, Poland
ORCID ID: 0009-0006-4824-0101

ABSTRACT

Artificial intelligence (AI) is increasingly being used in modern medicine, influencing decision-making processes, diagnostics, treatment, education, and the organization of the healthcare system. Despite its significant potential for improving the quality and efficiency of medical care, the rapid development and implementation of new systems are associated with numerous and complex ethical challenges. The aim of this review is to analyze the current literature on the ethical aspects of the use of artificial intelligence in medicine, with particular emphasis on autonomy, transparency, privacy, and accountability. The review was conducted using the PubMed and SpringerLink databases and other reliable sources, covering publications from 2015 to 2025. The analysis included open-access peer-reviewed articles in English focusing on the ethical, clinical, and legal aspects of AI use in healthcare, and the selected publications were subjected to a detailed analysis. The results indicate that artificial intelligence can significantly improve clinical practice by increasing diagnostic accuracy and optimizing therapeutic processes. At the same time, significant ethical concerns remain, including the impact on patient and physician autonomy, limited transparency of systems, threats to patient privacy resulting from the processing of large datasets, and the lack of clear regulations regarding liability for decisions based on AI recommendations. The analysis also revealed interconnections between these areas. In summary, although artificial intelligence has enormous potential, its ethical implementation requires a human-centered approach, a clear regulatory framework, and interdisciplinary collaboration.

KEYWORDS

Artificial Intelligence, Ethics, Autonomy, Privacy, Transparency, Responsibility

CITATION

Martyna Musiał, Anna Matusik, Kamila Kasprzycka, Monika Spaczyńska-Kwiatkowska, Wiktoria Niedziela, Zuzanna Mielniczek, Anna Szwed, Katarzyna Słowik, Gabriela Matoga. (2026) Ethical Aspects of Artificial Intelligence in Medicine: A Systematic Review. *International Journal of Innovative Technologies in Social Science*. 2(50). doi: 10.31435/ijitss.2(50).2026.5652

1. Introduction

Artificial intelligence (AI) is being used more and more widely in virtually all areas of life, and its rapid development is particularly evident in healthcare, where AI is utilized in screening, diagnosis, treatment, and education. There are many examples of AI applications that have significantly contributed to medical progress. One such example is the "Surgical Collective Consciousness" channel, which uses AI to analyze recordings of surgical procedures, enabling the identification of errors and enhancing the effectiveness of the learning process (Nasir, Siddiqui & Ahmed, 2025). Another example is BelongAI Dave, an artificial intelligence that serves as a mentor and assistant for cancer patients and their families (Hou et al., 2025). In psychiatry, virtual therapists used in the treatment of people with autism spectrum disorders and dementia are gaining increasing attention. Intelligent animal-like robots, such as the seal Paro, serve as home assistants, offering companionship and reducing loneliness among the elderly, those living alone, or those suffering from dementia. Furthermore, research is underway on the use of robots in the treatment of sexual dysfunctions, including pedophilia (Fiske, Henningsen & Buyx, 2019). Neural networks are used in screening tests, for example, in the classification of skin lesions or the localization of metastases in lymph nodes (Jackson et al., 2024). Beyond clinical medicine, artificial intelligence is widely used for maintaining medical records and storing personal data; it also plays a significant role in educational processes. For example, AI models are capable of passing medical exams at expert levels (Nasir et al., 2025), and mobile internet devices provide doctors and students with easy and quick access to numerous scientific databases and medical applications (Al. Kuwaiti et al., 2023). Additionally, virtual medical simulations facilitate student learning and enable case analysis under controlled conditions. The use of AI significantly contributes to cost optimization, increased efficiency, and the personalization of medical care (Busch, Adams & Bressem, 2023). At the same time, the growing use of artificial intelligence in various areas of life raises significant ethical concerns, primarily regarding false results, liability, data privacy, and fairness. Concerns also arise regarding the potential for

discrimination stemming from errors in the data used during training, which may lead to incorrect diagnoses and clinical actions; as well as the risk of violating patient dignity and autonomy (Weidener & Fischer, 2024). Concerns also extend to the potential use of artificial intelligence by students during exams, which could lead to issues of unfairness and the erosion of medical knowledge (Busch, Adams & Bressem, 2023). Consequently, ethical concerns regarding the use of artificial intelligence in medicine have become the subject of numerous studies. Among other developments, Bartenschlager's AI ethics model was created, which is based on the concept of Trustworthy AI and encompasses: autonomy, safety, data protection, fairness, well-being, and accountability (Hou et al., 2025). This article aims to provide a comprehensive analysis of ethical issues related to the use of artificial intelligence in medicine, focusing primarily on issues such as autonomy, transparency, privacy, and accountability.

2. Methodology

This review was conducted based on a broad and systematic analysis of the scientific literature, the aim of which was to collect, organize, and critically evaluate current discussions regarding the benefits, challenges, and limitations associated with the use of artificial intelligence in medicine. This process was multi-stage and included the identification, selection, and qualitative analysis of available scientific publications, in order to ensure the most comprehensive coverage of the topic under study. Relevant publications were identified using reputable scientific databases such as PubMed and Springer Link, as well as other reliable academic sources and publication repositories. The time frame of the review covered the years 2015-2025, which allowed for the inclusion of the latest developments in the rapidly evolving field of artificial intelligence in medicine, while maintaining a perspective on the evolution of the concept over the past decade. This approach allowed us to focus on the most up-to-date and, at the same time, methodologically significant studies. The search strategy was precisely targeted at identifying publications addressing key ethical challenges related to the use of artificial intelligence in medicine. Particular attention was paid to issues such as patient and healthcare professional autonomy, transparency and explainability of AI systems, privacy and the protection of medical data, and accountability for decisions made with the involvement of algorithms. Additionally, publications concerning the mechanisms of artificial intelligence, including machine learning and neural networks, as well as existing and developing legal regulations and ethical frameworks regarding its applications in healthcare, were included. The search process utilized a set of carefully selected keywords, including: "artificial intelligence," "ethics," "medicine," "privacy," "transparency," "autonomy," and "accountability." This allowed us to narrow down the results to the most relevant scientific works, directly related to the research problems under analysis. The analysis included only peer-reviewed scientific articles published in English and available in open access. This criterion was applied to ensure high scientific reliability of the source material, its verifiability, and easy accessibility for further analysis and verification of the results. At the same time, sources that did not meet scientific standards were excluded, including non-peer-reviewed materials, publications older than 2015, as well as works published in languages other than English, with the aim of limiting potential interpretive differences and ensuring analytical consistency. Additionally, duplicate articles and publications were excluded from the analysis, as well as works not directly related to the adopted research context, i.e., the clinical and ethical context of artificial intelligence applications in medicine. This approach allowed us to avoid data duplication and focus exclusively on literature of high substantive and research value. The final selection of studies was based on their methodological quality, the transparency of the presented results, and their clinical and ethical significance for the analyzed issue. Particular attention was paid to the robustness of the research methods used, the clarity of the conclusions, and their potential application in medical practice and medical education. The results derived from the selected literature were then subjected to a comparative analysis aimed at assessing the benefits, challenges, and limitations associated with the use of artificial intelligence in medicine and medical education. This analysis enabled the identification of major research trends, discrepancies in results, and gaps in existing knowledge, providing a basis for further discussion and conclusions.

3. Results

3.1. Autonomy

Patient autonomy is one of the fundamental rights to self-determination. It encompasses a person's right to form their own judgments, make decisions in accordance with their own moral principles, and exercise control over their own life and health. In the context of medical practice, the principle of autonomy obligates healthcare personnel to provide the patient with accurate information about their health status and available treatment options, to respect their choices, to be truthful, and to maintain medical confidentiality. It should be emphasized that the patient should play a key role in the decision-making process regarding their health, and the approach to patients should be individualized, not group-based (Busch et al., 2023). An important aspect of autonomy in the context of ethical challenges related to the use of artificial intelligence is ensuring that the patient fully understands both the purpose and the mode of operation of a given technology. It is therefore necessary to thoroughly assess whether the patient is aware that no human, whether they understand the mechanism and functions that a specific technology is intended to perform in their environment, and whether they are aware of the potential consequences of its use, including, for example, the risk of privacy violations (Fiske, Henningsen & Buyx, 2019). One of the fundamental elements guaranteeing human autonomy is informed consent; however, this principle does not apply to individuals who, due to their mental or physical health condition or other limitations, are unable to express it. This group includes, among others, elderly patients, children, people with disabilities, and patients in acute psychotic states. Limited capacity to make informed decisions exacerbates the moral dilemmas associated with the use of AI in diagnostic and therapeutic processes—these patients may not be able to fully understand how the technology works, assess its advantages and disadvantages, and thus make an informed choice, which increases the risk of manipulation and coercion (Fiske, Henningsen & Buyx, 2019; Varkey, 2020). The principle of autonomy applies not only to patients, but also to healthcare professionals and medical students. Inaccurate information generated by technologies used in the educational process, as well as a lack of transparency regarding how algorithms work, can lead to a restriction of students' autonomy, making it difficult for them to make informed, rational, and ethical decisions based on factual data. It is crucial to ensure that technology users have an adequate level of knowledge regarding how these technologies function, their potential applications, limitations, and potential risks. It is also essential to foster critical thinking skills and sound judgment among students when using AI tools (Busch et al., 2023). The reduction in autonomy caused by the increasingly widespread use of AI also poses a threat to healthcare professionals. Despite the high accuracy and utility of AI systems, final diagnostic and therapeutic decisions should be made by specialists- based on their current knowledge, professional judgment, and experience. AI models may lead to manipulation in this regard, which will limit physicians' autonomy and negatively impact the patient's therapeutic process (Hou et al., 2025). On the one hand, AI-based tools can enhance patient autonomy by increasing access to information, enabling health monitoring, or facilitating participation in decision-making. Examples include mobile apps for self-monitoring of health parameters, which can strengthen the patient's position in the doctor-patient relationship and promote a shared decision-making model (Sauerbrei et al., 2023). However, there is concern that the technologies used may negatively impact the doctor- patient relationship (Sauerbrei et al., 2023; Varkey, 2020). An equal role in the decision-making process is very important; however, excessive reliance on AI systems may lead to its depersonalization, or even its transformation into a purely transactional relationship, which undermines the patient's trust in the doctor. It should also be remembered that artificial intelligence does not possess the ability to account for the pluralism of values—biological, psychological, moral, and spiritual—which constitutes a significant advantage of humans in the decision-making process (Hou et al., 2025). Additional concerns arise from the phenomenon of paternalism, understood as undesirable interference in the patient's decisions, justified by concern for their well-being. There is a legitimate fear that this mechanism, hitherto associated mainly with medical practice and stemming from the principle of beneficence, will be transferred to artificial intelligence systems. As a result, they could make decisions on behalf of the patient and the physician, violating the principle of shared decision-making (Bullock, Gergel & Kingma, 2015). If algorithm recommendations begin to be regarded as more objective or precise than a doctor's clinical judgment, there is a risk of excessive reliance on the system's decisions. In such a situation, patient autonomy may be limited not by the doctor's authority, but by the authority of the technology itself (Hou et al., 2025).

3.2. Transparency

The transparency of algorithms used in the diagnostic and treatment process forms the basis of the principles of explainability and autonomy for both the patient and the physician. It also plays a key role in building trust within the physician-patient relationship and between the physician, the patient, and artificial intelligence systems. Decisions should be based on systematic explanations that enable accurate and informed decisions (Karimian, Petelos, & Evers, 2022). Transparency can be defined as the ability to obtain answers to questions regarding the mechanism and mode of operation of a given artificial intelligence system based on information gathered during its operation. Consequently, the concepts of "white box" and "black box" have emerged. A white box can be defined as the transparency of the source code, which is important from the perspective of programmers, or the availability of adequate explanations regarding the system's operation for end users, such as healthcare workers and patients (Bogina et al., 2021). In contrast, the black box refers to a situation in which it is impossible to interpret and understand the model's operation (Gallée et al., 2023). Transparency in artificial intelligence applications remains a significant challenge due to the high level of sophistication and complexity of the algorithms used. It is also of key importance in the context of legal liability, as the use of AI in medical practice can lead to consequences resulting from potential errors and incorrect decisions. For this reason, the decision-making process of artificial intelligence systems should be as transparent as possible to enable proper attribution of liability in the event of legal disputes. The transparency of AI systems used in medicine should be based primarily on the availability of information regarding their characteristics and the assumptions on which final results and recommendations—this makes the processes carried out by AI explainable, which facilitates the understanding of cause-and-effect relationships by both doctors and patients, and consequently enables decision-making in accordance with the principle of beneficence, which is the foundation of healthcare (Hou et al., 2025). Another important aspect is trust—for a given technology to be considered trustworthy, its principles and mechanisms of operation must be transparent. In practice, however, due to confidentiality issues and commercial interests, achieving this goal can be difficult, which raises legitimate ethical concerns regarding potential consequences for patients. An additional factor limiting transparency is the enormous volume of data processed and the complexity of the computational processes, which make it difficult to understand how the models work. A lack of transparency may stem both from the ambiguity of the decision-making process itself and from the limited transparency of the input data on which the process is based, leading to problems in assessing the system's reliability (von Eschenbach, W. J., 2021). Insufficient transparency can lead to decisions being made without adequate human involvement, which carries serious ethical implications; therefore, it is crucial to provide users with easy access to information about how algorithms work and to present this content in a clear and accessible manner (Naik et al., 2022; Ouchchy et al., 2020). It is essential that both patients and healthcare professionals understand the fundamentals of the correctness of decisions made by AI and have confidence in the proposed solutions. To this end, appropriate training for healthcare professionals may prove useful, as it will contribute to a better understanding of how the AI algorithms used in a given facility work (Quazi, 2024). One solution to the problem of limited transparency is the development of XAI, or explainable artificial intelligence, which, through the use of appropriate techniques, enables the interpretation of results generated by these models (Mabillard et al., 2021). XAI encompasses methods that allow for an understanding of both how algorithms work and the results they produce. In practice, systems utilizing XAI can significantly facilitate the interpretation by doctors and patients of decisions made by AI and the processes that lead to them. This significantly contributes to increasing the transparency of the techniques used, and consequently, the explainability of the diagnostic process and trust in the proposed solutions will also increase. (Al. Kuwaiti et al., 2023). It is also extremely important that AI models appropriately signal their limitations and potential errors. This helps prevent the dissemination of incorrect information, which could lead to the erosion of medical knowledge, disruptions in the decision-making process, and actions detrimental to the patient, which would contradict the principle of beneficence (Busch et al., 2023). Ultimately, it should be emphasized that trust in decisions generated by artificial intelligence should stem from trust in the physician's judgment. AI can be a valuable tool supporting the diagnostic and therapeutic process, but it should not be the sole decision-maker. It is crucial for the physician to verify the results and base final actions on their knowledge and experience (von Eschenbach, W. J., 2021).

3.3. Privacy and Data Protection

Artificial intelligence-based technologies require access to large datasets, which form the basis for making accurate and reliable clinical decisions. This data is collected and stored, creating a risk of data breaches, which in turn raises a significant issue regarding the confidentiality of patients' personal data and may severely compromise their privacy and security. Analysis enabling decisions that benefit the patient relies on the integration of data from multiple sources, such as laboratory and imaging test results, medical records, demographic data, genomic information, and health insurance data (Nasir et al., 2025). The scale and sensitive nature of this information increase the risk of its unauthorized disclosure or misuse. Data protection is one of the most serious challenges associated with the implementation of artificial intelligence in medicine. Ensuring security requires the use of strict protective mechanisms at every stage of the data lifecycle—from collection and sharing, through analysis, to archiving. Sensitive data is collected, analyzed, and stored on a massive scale. For example, AI/ML (Artificial Intelligence and Machine Learning) applications are designed to detect specific patterns based on the analysis of massive databases and then generate predictions regarding diseases—their progression, likelihood of occurrence, and treatment. (Gerke & Rezaeikhonakdar, 2022). Another example of intensive data use is the development of the IoMT—the Internet of Medical Things—which includes devices such as smartwatches, rings, and wristbands that monitor specific vital signs, as well as remotely monitored devices, including blood pressure monitors and glucose meters, which transmit data directly to the patient or physician. Such solutions significantly revolutionize and streamline clinical practice; however, the mass collection and transmission of data, without adequate oversight, can pose a serious threat to patient privacy (Shakeel et al., 2022). There is a real risk that data used to train artificial intelligence systems will be intercepted by unauthorized parties and used for commercial or criminal purposes. Privacy protection remains closely linked to the level of public trust in the healthcare system. Patients must be confident that their health information is properly secured and used only for clearly defined purposes. A lack of such trust can negatively impact the relationship between the patient and the healthcare provider. It is therefore crucial to obtain the patient's informed consent for the processing of sensitive data, which includes full information about potential risks, the security measures in place, and the purpose of data use (Mohanasundari et al., 2023; Murphy et al., 2021). Consequently, patients should be informed that they may choose not to share their data in the absence of a clearly defined oversight body that ensures the data will not be used for commercial or discriminatory purposes, e.g., by insurance entities (Mabillard et al., 2021). One solution that reconciles the need for large datasets with the necessity of protecting privacy is the use of generative data. This consists of realistic yet synthetic patient data, created by artificial intelligence models in such a way that it replicates the statistical properties of real-world datasets but is not linked to any specific individuals. This makes it possible to train machine learning models on rich and diverse datasets without the risk of exposing sensitive medical data. Generative data can reflect various clinical scenarios, rare disease cases, or atypical combinations of symptoms that occur sporadically in real-world datasets. This, in turn, enables the creation of more robust and accurate predictive models. An additional advantage of this approach is the ability to more easily share data among research institutions and scientific teams, as synthetic datasets are not subject to the same stringent legal restrictions as personal data. This could significantly accelerate the development of innovations in medicine, including AI-assisted diagnostics and personalized medicine. (Murdoch, 2021). Over the years, there has been a significant improvement in data protection, primarily due to the introduction of comprehensive legal regulations in response to the growing importance of information in the digital age. Examples of such measures include the European Union's General Data Protection Regulation (GDPR), the guidelines "Ethics and Governance of Artificial Intelligence in Health Care" developed by the World Health Organization (WHO), and the U.S. Health Insurance Portability and Accountability Act (HIPAA). These regulations have established standards for the collection, storage, and processing of data, particularly sensitive data such as medical information. However, the rapid development of artificial intelligence technologies and the growing scale of cyber threats mean that existing legal frameworks often cannot keep pace with the speed of technological change. It is therefore necessary to systematically update them and create new regulatory guidelines. At the same time, technological solutions are being developed to enhance data security, including advanced encryption techniques, secure cryptographic key management, multi-level access authorization mechanisms, and data anonymization and pseudonymization procedures. Their application makes it possible to reduce the risk of unauthorized access and minimize potential harm resulting from security breaches, which is particularly important in the context of artificial intelligence systems operating on large medical datasets (Alowais et al., 2023; Chan & Zary, 2019; Jackson et al., 2024). Data privacy breaches can have real social and individual consequences for patients—ranging from the risk of discrimination in education and the workplace, through

injustices in the insurance system, to a sense of injustice and lack of control over the management of information regarding one's own health (Yadav et al., 2023). At the same time, excessive restrictions on data access can negatively impact the quality of analyses conducted by artificial intelligence systems, and thus the final results that influence further treatment plans and patient prognosis. Responsible data management is therefore the foundation of the principle of patient autonomy and respect for patient rights (Karimian et al., 2022); thus, it is important to strike a balance between the availability of data necessary for the development and effectiveness of AI systems and ensuring a high level of privacy protection and medical information security (Hou et al., 2025; Quazi, 2024).

3.4. Responsibility

The issue of liability is another important ethical concern related to the use of artificial intelligence systems in medicine. In the traditional diagnostic and therapeutic process, liability for any errors or oversights rests with the healthcare professionals involved in the patient's care. However, the situation becomes significantly more complicated when AI systems are incorporated into the decision-making process. This stems primarily from the lack of a transparent legal framework that would unambiguously define the rules for assigning liability for errors arising from the use of technology and clearly regulate the division of liability among the entities involved in the design, implementation, and use of AI systems. Such ambiguity poses a threat to both medical staff and patients. Healthcare workers are exposed to increased legal risk, which may lead to higher levels of occupational stress, while patients may encounter difficulties in seeking justice and obtaining potential compensation (Hou et al., 2025). As a result, determining who bears responsibility for decisions supported by artificial intelligence becomes a significant challenge from both a legal and ethical perspective. The problem of assigning responsibility remains closely linked to the issue of the explainability of AI systems. If the mechanisms of an algorithm's operation remain opaque to the user, it is difficult to expect a physician to take full responsibility for a decision based on the recommendations of a technology whose functioning they are unable to understand. A lack of understanding of the algorithm's mechanisms limits the ability to critically evaluate its results, which in practice can lead to unthinking reliance on the technology or, conversely, to its complete rejection. Currently, there is still a lack of clear legal regulations that would precisely define the division of responsibility in such situations. It is unclear whether responsibility for any errors should rest with the physician using the system, the algorithm's developer, the technology provider, or the management of the medical facility that decided to implement it. As a result, the process of assigning liability becomes opaque and fosters legal disputes and a sense of uncertainty among medical staff (Sauerbrei et al., 2023). This issue is further complicated by the proprietary nature of algorithms and AI technologies, which limits access to the information necessary for determining the causes of errors. On the one hand, it seems reasonable to expect that a healthcare professional using a given technology in their practice bears responsibility for any oversights, as AI systems should serve as support tools in the diagnostic and therapeutic process, and final decisions should be made based on the physician's knowledge and experience. Maintaining a critical approach takes on particular importance in this context, especially in non-standard situations. Algorithms are trained on specific datasets, so the emergence of a more complex case, a rare complication, or treatment resistance may result in erroneous recommendations proposed by AI due to limitations in the training data. Awareness of these limitations and the ability to critically interpret the system's results remain crucial for the safety of the diagnostic and therapeutic process (Naik et al., 2022). On the other hand, responsibility also lies with technology manufacturers and providers, who are expected to ensure the safety, effectiveness, and adequate quality of AI systems (Murphy et al., 2021). Potential errors can arise at any stage of the technology's operation—from data selection and collection, through processing and analysis, to the interpretation of results. Due to the large number of people involved in individual processes, assigning responsibility for potential errors becomes particularly difficult (Shafqat, 2024). An additional challenge is the difficulty of predicting all possible consequences of AI systems' operation, which hinders the creation of precise regulatory frameworks. In addition to legal responsibility, moral responsibility also plays a significant role. In situations where formal regulations are unclear, it is ethical norms, social pressure, concern for professional reputation, and an individual's value system that can significantly influence clinical decisions. However, due to the lack of established authorities and standards in this area, the issue of moral responsibility still requires in-depth reflection. The global scale of artificial intelligence system implementation further complicates the issue of accountability. It is difficult to predict who will be affected by the technology and to what extent, which hinders the enforcement of justice in the event of harm. Existing regulations, such as the AIA—the Artificial Intelligence Act—despite containing specific requirements and guidelines regarding

human oversight and ensuring a thorough understanding of system operations by end-users, do not specify concreto verification mechanisms or sanctions for non-compliance, which may limit the actual accountability of entities implementing AI technologies (Winkel, 2024). Potential solutions to the problem of assigning liability include, above all: developing a detailed legal framework, establishing transparent accountability structures, continuously monitoring system operations along with feedback mechanisms, and implementing independent reliability audits. Clearly defining the scope of responsibility during the design, implementation, and use of technology will enable more effective identification of sources of error and the implementation of corrective measures (Quazi, 2024). Regulation should be developed by interdisciplinary teams comprising technology developers, researchers, ethics experts, and public policymakers, which will allow for the multidimensional nature of the problem to be addressed and for transparent accountability principles to be established (Floridi and Cowls, 2019). In summary, the issue of liability in the context of using artificial intelligence in medicine remains a complex and still insufficiently regulated problem. The combination of limited explainability of systems, the diffusion of liability among various entities, and the dynamic development of technology means that current approaches prove insufficient. Therefore, it is necessary to develop a coherent, transparent legal framework and ethical standards that will allow not only for the effective enforcement of liability, but also for the safe and informed use of the potential of artificial intelligence in healthcare.

4. Discussion

The rapid development of artificial intelligence is impacting every aspect of both society and the individual. This review focuses on the ethical dilemmas associated with the use of AI in medicine, with particular emphasis on issues of patient and healthcare professional autonomy, the transparency of AI systems, privacy and the protection of sensitive data, and accountability for decisions made with the involvement of these systems. The analyzed publications indicate that the growing use of artificial intelligence in diagnostic and therapeutic processes creates significant opportunities to improve the quality of healthcare, but at the same time raises a number of complex ethical and social challenges. One of the most important issues is the impact of AI on patient and physician autonomy. On the one hand, these technologies increase access to medical information, which can strengthen the patient's role in the decision-making process and promote a model of shared decision-making. At the same time, AI systems can support physicians by providing additional analyses and recommendations, potentially improving the accuracy of clinical decisions. On the other hand, the limited transparency of these systems' operations may undermine informed consent, particularly among vulnerable groups, and increase the risk of uncritical reliance on algorithmic recommendations. Consequently, this may lead to a reduction in the autonomy of both patients and medical staff, as well as a shift in decision-making responsibility from humans to technology. Another area requiring special attention is the issue of system opacity. As transparency increases, so does the explainability of the technology's operation, which enhances the autonomy of both the physician and the patient in the decision-making process and improves the quality of final decisions. This also fosters trust and enables proper attribution of responsibility in the event of legal disputes. In this context, the development of the XAI concept and appropriate education of medical personnel are of significant importance. However, the complexity of systems, the so-called "black box" problem, and limitations arising from commercial interests and data confidentiality remain challenges. Data privacy also remains a complex issue. Access to large datasets improves the accuracy of diagnosis and treatment due to the broad range of cases used in machine learning. It enables the development of personalized medicine, continuous monitoring of a patient's health, and ensures ongoing scientific progress and the emergence of new innovations. At the same time, sharing, analyzing, and storing vast amounts of data increases the risk of data leaks and misuse, particularly in the context of rising cybercrime. This can negatively impact patients' trust in technology and the healthcare system. Conversely, restricting access to data reduces the quality of AI models, hindering their effective training. Numerous legal regulations and methods for encryption, anonymization, and access control have already been established, which enhance the protection of medical data; however, due to current technological advancements, their constant modernization is necessary. The issue of assigning liability for erroneous decisions made based on AI recommendations remains unclear. There are no clear regulations specifying who is responsible for errors: the physician, the manufacturer, the technology provider, or the medical facility, which can lead to occupational stress and fear of using AI. This may also make it difficult for patients to seek justice and obtain compensation.

However, the rapid pace of technological development makes it difficult to create effective legal regulations, given the difficulty of predicting the effects of AI. It is therefore crucial to develop accountability

systems, oversight mechanisms, and audits, as well as to promote the overriding role of humans in making final decisions. An analysis of the collected research also shows that the ethical challenges associated with artificial intelligence cannot be considered in isolation. Autonomy, transparency, privacy, and accountability remain closely intertwined and mutually dependent. Limited transparency hinders the assignment of responsibility, the lack of data protection undermines patient autonomy, and an unclear framework of responsibility erodes public trust in the technology. This points to the need for a holistic approach to AI ethics, based on interdisciplinary collaboration involving physicians, computer scientists, lawyers, ethicists, and public policymakers. Attention should also be paid to the limitations of this study—the analysis covered specific English-language publications from specific years that were available via open access, which may have significantly limited the scope of research perspectives. Additionally, the dynamic development of technology may render some conclusions obsolete as new technological and regulatory solutions emerge. Despite these limitations, the review provides a synthetic overview of the current ethical implications associated with the use of AI systems in medicine and identifies the most important directions for further research. Future research should also focus on developing harmonized international guidelines to ensure the safe and ethical integration of artificial intelligence into clinical practice. Furthermore, ongoing interdisciplinary collaboration will be essential to adapt ethical frameworks to the rapidly evolving capabilities of AI systems in healthcare. It is also crucial to advance education among both medical staff and patients regarding the informed and responsible use of AI-based tools. Furthermore, it will be necessary to monitor the long-term effects of implementing these technologies in order to identify new ethical and social risks on an ongoing basis.

5. Conclusions

Artificial intelligence is becoming one of the most important and fastest-growing tools in modern medicine, significantly influencing the transformation of both diagnostic and therapeutic processes. Its applications include, among others, assisting in the interpretation of test results, analyzing large clinical datasets, predicting disease progression, personalizing treatment, and optimizing the management of the healthcare system. As a result, this leads to improved quality of medical care, increased staff efficiency, and better utilization of available resources. However, the analysis clearly indicates that the dynamic development and implementation of artificial intelligence systems are associated with numerous and complex ethical challenges that require a systemic, comprehensive, and interdisciplinary approach. These issues cannot be addressed solely within a single discipline but require collaboration among specialists in medicine, computer science, law, ethics, and the social sciences to develop safe and responsible solutions. A key conclusion from the analysis is the necessity of preserving the primary role of humans in the decision-making process. Regardless of the degree of technological advancement of AI systems, they remain susceptible to limitations arising from data quality, training methods, and the computational models employed. The risk of error, limited transparency of algorithm operation, and the inability to fully account for the pluralism of moral, social, and cultural values constitute significant risks in the context of their application in medicine. Therefore, ultimate decision-making responsibility should always rest with the physician, who bases their decisions on clinical knowledge, professional experience, and an individual assessment of the patient's situation. Artificial intelligence systems should therefore be treated as tools that support the decision-making process, functioning as advanced clinical assistants rather than autonomous decision-makers. Their role should be to provide additional information, analyses, and recommendations that can assist the physician, but not replace their judgment. At the same time, it is particularly important to ensure the patient's active and informed participation in the medical decision-making process, which includes respecting their autonomy, values, and individual preferences, as well as the necessity of obtaining informed consent for proposed diagnostic and therapeutic measures. Transparency of artificial intelligence systems is of fundamental importance in this context. A lack of transparency directly reduces the explainability of algorithms, which in turn lowers the level of trust among users, including both physicians and patients. A lack of transparency hinders fully informed clinical decision-making, increases the risk of errors, and significantly complicates the process of assigning responsibility in disputed situations or in the event of adverse clinical outcomes. It is worth emphasizing that potential errors in AI systems cannot be treated solely as statistical data or technical deviations, because in clinical practice, every incorrect decision can lead to serious health consequences, and in extreme cases, even to permanent damage to the patient's health or loss of life. Another key area requiring special attention is the protection of privacy and the security of sensitive patient data. These should be a priority at every stage of the lifecycle of artificial intelligence systems - from data collection, through processing and analysis, to storage and use. It is essential

to strike a balance between access to large, diverse datasets, which are a prerequisite for the effective training of AI models and ensuring an adequate level of protection for this information against unauthorized access, leakage, or misuse. Another significant, still unresolved issue is the question of legal and ethical responsibility for decisions made with the support of artificial intelligence systems. The lack of clear and consistent legal regulations in this area constitutes a significant barrier to the safe implementation of these technologies in clinical practice, while also making it difficult for patients to effectively assert their rights in the event of errors or harm. Therefore, it is necessary to both continuously improve technological solutions and systematically update legal regulations to adapt to the rapidly changing digital environment. These regulations should clearly define the scope of responsibility of individual entities and strengthen mechanisms for the protection of patient data. An extremely important element of implementing artificial intelligence in medicine is also ensuring an adequate level of education and training for medical personnel regarding the operation of these systems. Increasing the digital competencies of physicians and other healthcare workers contributes to a better understanding of the capabilities and limitations of AI technology, which consequently leads to improved quality of clinical decisions, reduced risk of errors, and increased trust in new technological tools. The analysis also showed that key concepts such as autonomy, transparency, privacy, and accountability are closely interrelated and cannot be considered in isolation. Changes in one of these areas directly affect the others, meaning that effectively addressing ethical issues related to AI requires a holistic approach. This also necessitates close interdisciplinary collaboration among specialists from various fields, which enables a more balanced and responsible implementation of the technology. In summary, artificial intelligence has enormous transformative potential in medicine; however, its ethical implementation requires maintaining the primary role of humans in the decision-making process and establishing a clear, consistent, and enforceable framework of accountability. Further development of AI in the healthcare sector should proceed in parallel on three levels: technological, regulatory, and ethical, to ensure its safe, transparent, and responsible application, with the patient's best interests as the overriding priority.

REFERENCES

1. Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education, 23*, Article 689. https://doi.org/10.1186/s12909-023-04698-z
2. Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-Muhanna, F. A. (2023). A review of the role of artificial intelligence in healthcare. *Journal of Personalized Medicine, 13*(6). https://pmc.ncbi.nlm.nih.gov/articles/PMC10301994/
3. Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2021). Educating software and AI stakeholders about algorithmic fairness, accountability, transparency and ethics. *International Journal of Artificial Intelligence in Education, 32*, 808–833. https://doi.org/10.1007/s40593-021-00248-0
4. Bullock, E., Gergel, T., & Kingma, E. (2015). Conference report: Interdisciplinary workshop in the philosophy of medicine: Parentalism and trust. *Journal of Evaluation in Clinical Practice, 21*(3), 542–548. https://doi.org/10.1111/jep.12365
5. Busch, F., Adams, L. C., & Bressem, K. K. (2023). Biomedical ethical aspects towards the implementation of artificial intelligence in medical education. *Medical Science Educator, 33*(4), 1007–1012. https://pmc.ncbi.nlm.nih.gov/articles/PMC10403458/
6. Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education, 5*(1). https://mededu.jmir.org/2019/1/e13930/
7. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1), 1–14. https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8
8. Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5). https://pmc.ncbi.nlm.nih.gov/articles/PMC6532335/
9. Gallée, L., Kniesel, H., Ropinski, T., & Götz, M. (2023). Artificial intelligence in radiology—Beyond the black box. *RöFo: Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren, 195*(9), 797–803. https://www.thieme-connect.com/products/ejournals/html/10.1055/a-2076-6736?articleLanguage=en
10. Gerke, S., & Rezaeikhonakdar, D. (2022). Privacy aspects of direct-to-consumer artificial intelligence/machine learning health apps. *Intelligence-Based Medicine, 6*, Article 100061. https://www.sciencedirect.com/science/article/pii/S266652122200014X

11. Hou, J., Cheng, X., Liao, J., Zhang, Z., & Wang, W. (2025). Ethical concerns of AI in healthcare: A systematic review of qualitative studies. *Nursing Ethics*. https://journals.sagepub.com/doi/10.1177/09697330251385024
12. Jackson, P., Ponath Sukumaran, G., Babu, C., Tony, M. C., Jack, D. S., Reshma, V. R., Davis, D., Kurian, N., & John, A. (2024). Artificial intelligence in medical education: Perception among medical students. *BMC Medical Education, 24*(1), Article 804. https://pmc.ncbi.nlm.nih.gov/articles/PMC11283685/
13. Karimian, G., Petelos, E., & Evers, S. M. A. A. (2022). The ethical issues of the application of artificial intelligence in healthcare: A systematic scoping review. *AI and Ethics, 2*, 539–551. https://doi.org/10.1007/s43681-021-00131-7
14. Mabillard, V., Demartines, N., & Joliat, G. R. (2021). How can reasoned transparency enhance co-creation in healthcare and remedy the pitfalls of digitization in doctor-patient relationships? *International Journal of Health Policy and Management, 11*(10), 1986–1990. https://pmc.ncbi.nlm.nih.gov/articles/PMC9808292/
15. Mohanasundari, S. K., Kalpana, M., Madhusudhan, U., Vasanthkumar, K., Rani, B., Singh, R., Vashishtha, N., & Bhatia, V. (2023). Can artificial intelligence replace the unique nursing role? *Cureus, 15*(12). https://pmc.ncbi.nlm.nih.gov/articles/PMC10811613/
16. Murdoch, B. (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Medical Ethics, 22*(1), Article 122. https://pmc.ncbi.nlm.nih.gov/articles/PMC8442400/
17. Murphy, K., Di Ruggiero, E., Upshur, R., Willison, D. J., Malhotra, N., Cai, J. C., Malhotra, N., Lui, V., & Gibson, J. (2021). Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Medical Ethics, 22*, Article 14. https://doi.org/10.1186/s12910-021-00577-8
18. Nasir, M., Siddiqui, K., & Ahmed, S. (2025). Ethical-legal implications of AI-powered healthcare in critical perspective. *Frontiers in Artificial Intelligence, 8*, Article 1619463. https://pmc.ncbi.nlm.nih.gov/articles/PMC12263600/
19. Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Frontiers in Surgery, 9*, Article 862322. https://pmc.ncbi.nlm.nih.gov/articles/PMC8963864/
20. Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & Society, 35*(4), 927–936. https://doi.org/10.1007/s00146-020-00965-5
21. Quazi, F. (2024). Ethics & responsible AI in healthcare. *International Journal of Geo-Information Systems*. https://ijgis.pubpub.org/pub/wbmyd4xu/release/1
22. Ratti, E., Morrison, M., & Jakab, I. (2025). Ethical and social considerations of applying artificial intelligence in healthcare: A two-pronged scoping review. *BMC Medical Ethics, 26*(1), 1–18. https://pmc.ncbi.nlm.nih.gov/articles/PMC12107984/
23. Sauerbrei, A., Kerasidou, A., Lucivero, F., & Hallowell, N. (2023). The impact of artificial intelligence on the person-centred, doctor-patient relationship: Some problems and solutions. *BMC Medical Informatics and Decision Making, 23*(1), Article 73. https://pmc.ncbi.nlm.nih.gov/articles/PMC10116477/
24. Shakeel, T., Habib, S., Boulila, W., Koubaa, A., Javed, A. R., Rizwan, M., Gadekallu, T. R., & Sufiyan, M. (2022). A survey on COVID-19 impact in the healthcare domain: Worldwide market implementation, applications, security and privacy issues, challenges and future prospects. *Complex & Intelligent Systems, 9*(1), 1027–1058. https://pmc.ncbi.nlm.nih.gov/articles/PMC9151356/
25. Shafqat, M. (2024). Ethical implications of AI in clinical practice: Balancing innovation and responsibility. *AI Medical Nexus, 1*(2). https://researchcorridor.org/index.php/aimn/article/view/429/409
26. Varkey, B. (2020). Principles of clinical ethics and their application to practice. *Medical Principles and Practice, 30*(1), 17–28. https://pmc.ncbi.nlm.nih.gov/articles/PMC7923912/
27. von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology, 34*(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0
28. Weidener, L., & Fischer, M. (2024). Artificial intelligence in medicine: Cross-sectional study among medical students on application, education, and ethical aspects. *JMIR Medical Education, 10*, Article e51247. https://pmc.ncbi.nlm.nih.gov/articles/PMC10799276/
29. Winkel, M. (2024). Society in charge: The connection of artificial intelligence, responsibility, and ethics in German media discourse. *AI and Ethics, 5*(1), 2839–2866. https://doi.org/10.1007/s43681-024-00604-5
30. Yadav, N., Pandey, S., Gupta, A., Dudani, P., Gupta, S., & Rangarajan, K. (2023). Data privacy in healthcare: In the era of artificial intelligence. *Indian Dermatology Online Journal, 14*(6), 788–792. https://pmc.ncbi.nlm.nih.gov/articles/PMC10718098/