Open Access Article

Ann. Acad. Med. Siles. (online) 2025; 1(nr specj.): 1–11
eISSN 1734-025X
DOI: 10.18794/aams/207480
www.annales.sum.edu.pl

PRACA POGLĄDOWA
REVIEW

AI in medicine: An analysis of threats, risks, and solutions for the future of healthcare

Zastosowanie AI w medycynie: analiza zagrożeń, ryzyka i rozwiązań dla przyszłości opieki zdrowotnej

Helena Brawańska[1], Marta Jędrzejowska[1], Karolina Lau[2], Janusz Kasperczyk[2]

[1]Students' Scientific Club, Department of Environmental Medicine and Epidemiology, Faculty of Medical Sciences in Zabrze, Medical University of Silesia, Katowice, Poland

[2]Department of Environmental Medicine and Epidemiology, Faculty of Medical Sciences in Zabrze, Medical University of Silesia, Katowice, Poland

ABSTRACT

Artificial intelligence (AI) is playing an increasingly significant role in medicine, impacting diagnostics, treatment, and the organization of healthcare systems. This paper analyzes the potential benefits and risks associated with the use of AI in medicine. It focuses on technical, ethical, and regulatory aspects, as well as the impact of AI on patient safety and the effectiveness of clinical decision-making. A comprehensive literature review was conducted in the PubMed, Scopus, and Google Scholar databases, considering publications from 2019 to 2025. A qualitative synthesis of the 52 selected articles identified key challenges and recommendations for further research and implementation of AI in healthcare. The analysis indicates that AI significantly improves diagnostic precision and therapy efficiency, while also posing risks of algorithmic errors, model bias, and breaches of patient privacy. Effective implementation of AI requires legal regulations, clear application guidelines, and multidisciplinary collaboration among experts. Future research should focus on developing mechanisms to increase trust in AI systems and ensure their responsible use in medicine.

KEYWORDS

artificial intelligence in medicine, AI risks, AI ethics, data security, AI legal regulations, machine learning

Received: 18.06.2025 **Revised:** 26.06.2025 **Accepted:** 26.06.2025 **Published online:** 02.10.2025

Address for correspondence: Helena Brawańska, Studenckie Koło Naukowe, Katedra i Zakład Medycyny i Epidemiologii Środowiskowej, Wydział Nauk Medycznych w Zabrzu ŚUM, ul. Jordana 19, 41-808 Zabrze, tel. +48 507 960 796, e-mail: s85617@365.sum.edu.pl

Publisher: Medical University of Silesia, Katowice, Poland

STRESZCZENIE

Sztuczna inteligencja (*artificial intelligence* – AI) odgrywa coraz większą rolę w medycynie, wpływając na diagnostykę, leczenie oraz organizację systemów opieki zdrowotnej. W pracy przeanalizowano potencjalne korzyści oraz zagrożenia związane z wykorzystaniem AI w medycynie. W szczególności skupiono się na aspektach technicznych, etycznych i regulacyjnych, a także na wpływie AI na bezpieczeństwo pacjentów oraz skuteczność podejmowanych decyzji klinicznych. Przeprowadzono kompleksowy przegląd literatury w bazach danych PubMed, Scopus i Google Scholar, uwzględniając publikacje z lat 2019–2025. Na podstawie wyselekcjonowanych 52 artykułów dokonano jakościowej syntezy wyników, identyfikując kluczowe wyzwania oraz rekomendacje dotyczące dalszych badań i wdrażania AI w opiece zdrowotnej. Wyniki analizy wskazują, że AI znacząco zwiększa precyzję diagnostyczną i skuteczność terapii, jednocześnie niesie za sobą ryzyko błędów algorytmicznych, stronniczości modeli oraz naruszenia prywatności pacjentów. Skuteczne wdrożenie AI wymaga uregulowań prawnych, przejrzystych zasad stosowania oraz współpracy ekspertów z różnych dziedzin. Przyszłe badania powinny koncentrować się na opracowaniu mechanizmów zwiększających zaufanie do systemów AI i zapewniających ich odpowiedzialne wykorzystanie w medycynie.

SŁOWA KLUCZOWE

sztuczna inteligencja w medycynie, ryzyko związane z AI, etyka AI, bezpieczeństwo danych, regulacje prawne AI, uczenie maszynowe

INTRODUCTION

Artificial intelligence (AI) is revolutionizing medicine by supporting diagnostics, medical documentation, and therapy-related decision-making [1]. AI refers to the field of computer science dealing with systems capable of performing tasks that typically require human intelligence, such as data analysis, pattern recognition, and decision-making [2]. Among its applications, AI-based clinical decision support systems (AI-CDSSs) are gaining increasing importance by assisting medical staff in clinical decisions through algorithms that analyze patient data [3,4].

One of the key areas of AI is machine learning (ML), which allows computers to analyze data and improve their performance without manual programming. There are two main approaches: supervised learning, which uses labeled data for training, and unsupervised learning, which analyzes unlabeled data to find hidden patterns [5]. A more advanced form of ML is deep learning (DL), based on multi-layered artificial neural networks processing vast amounts of data. Each layer transforms information and successive analyses lead to increasingly abstract conclusions until the final result is obtained [6]. DL is widely used in imaging diagnostics, where AI systems analyze medical images to detect pathologies [7].

Generative models form a separate category capable of creating new data based on existing patterns. They work by modeling the probability distribution of input data and generating new samples with similar features. In medicine, they are used to synthetically generate diagnostic images (e.g., magnetic resonance imaging [MRI], computed tomography [CT], or electroencephalogram [EEG] scans), which can facilitate AI training and improve performance in tasks such as detecting pathological changes [2,8]. However, synthetic data – despite its usefulness – can lead to distortions in algorithm performance. Models trained solely on synthetic data often perform poorly at analyzing real-world data. This is because synthetic data may not reflect the full complexity, variability, and significant features of real medical records, leading to generalization issues and higher risk of diagnostic errors. This phenomenon, known as domain mismatch, is a major challenge for the safe, reliable development of AI systems in medicine [9,10].

AI brings numerous benefits to medicine, especially in diagnostics, pathology, and telemedicine. In radiology, AI algorithms assist in analyzing x-rays, CT scans, and MRIs, enhancing the efficiency of detecting diseases such as lung disease, cancer, and bone fractures [11]. In oncology, AI is used in screening tests, such as mammography, improving breast cancer detection rates and reducing false positives. In pathology, AI analyzes histopathological images, identifying tumors (e.g., gliomas or lymphomas) with up to 96% accuracy [6,12]. In cardiology, AI supports analysis of electrocardiograms (ECGs), echocardiograms, and cardiac MRIs, allowing for earlier detection of arrhythmias and coronary syndromes [11]. Telemedicine applications enable remote patient monitoring, analysis of test results, and diagnostic support in areas with limited access to medical professionals [6].

However, the rapid development of AI poses challenges regarding regulation, transparency, and data security. Legal frameworks like the European Union’s (EU) Artificial Intelligence Act (AI Act) classify medical AI systems as high-risk technologies and establish requirements for human oversight, transparency, and performance monitoring. In the USA, the Food and Drug Administration’s (FDA) Artificial Intelligence/Machine Learning-based Software as a Medical Device Action Plan (AI/ML-based SaMD Action Plan) mandates model auditing and clinical robustness. Data protection is governed by law (the Health Insurance Portability and Accountability Act [HIPAA] in the USA and the General Data Protection Regulation [GDPR] in the EU) and best practices are outlined in the Good Machine Learning Practices (GMLP) [13,14].

Tools such as GPT-4 can enhance healthcare efficiency, but also introduce risks that demand regulation and informed governance. The aim of this paper is to analyze major threats associated with AI in medicine, including AI hallucinations, diagnostic errors, algorithmic biases, lack of transparency, and cybersecurity issues. As the technology advances, understanding its limitations and implementing safety strategies becomes crucial [9,15,16].

MATERIAL AND METHODS

Research scope and research area

This narrative review explores the opportunities, challenges, and regulatory concerns related to the use of AI in clinical practice. The guiding research question was, "How does the use of artificial intelligence in medicine impact patient safety, diagnostic and therapeutic effectiveness, and ethical/legal standards in healthcare systems?" The study addressed various dimensions, including technical reliability, clinical outcomes, legal compliance, ethical risks, and data privacy, focusing on the systems used in diagnostics, decision support, and therapy planning.

Literature search strategy

We conducted a structured literature review using the databases PubMed, Scopus, and Google Scholar, covering publications from 2019 to 2025. Both Medical Subject Headings (MeSH) and free-text keywords were used to ensure comprehensiveness. The search strategy focused on interdisciplinary terms connecting artificial intelligence with medical practice, legal regulations, and bioethics. The primary MeSH terms are presented in Table I.

Inclusion and exclusion criteria

We applied specific inclusion and exclusion criteria, outlined in Table II, to ensure the relevance and quality of the studies we included.

Table I. MeSH keywords used in this review

AI fundamentals	"artificial intelligence", "machine learning", "deep learning", "neural networks (computer)", "generative models"
Clinical application areas	"diagnosis, computer-assisted", "radiology", "oncology", "cardiology", "telemedicine", "clinical decision support systems"
Risk and ethics	"ethics, medical", "data privacy", "confidentiality", "informed consent", "bias, algorithmic", "trust"
Legal and regulatory context	"legislation, medical", "regulatory compliance", "medical device regulation", "liability, legal"
Safety and system performance	"patient safety", "reproducibility of results", "software validation", "adverse events reporting"

Table II. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Peer-reviewed studies on real-world applications of AI in medicine	Articles unrelated to medical AI or not relevant to healthcare practice
Studies analyzing AI's impact on patient safety, clinical decision-making, or diagnostic accuracy	Opinion pieces, commentaries, or non-empirical papers
Papers discussing legal, ethical, and regulatory aspects of AI in healthcare	Conference abstracts, letters to the editor, or inaccessible full texts
Reviews or meta-analyses with methodological transparency	Articles lacking outcome-based data or specific discussion of AI mechanisms
Documents referring to AI legislation (e.g., the AI Act, HIPAA, the GDPR, or the MDR)	Publications not in English or Polish without a verified translation

AI – artificial intelligence; HIPAA – Health Insurance Portability and Accountability Act; GDPR – General Data Protection Regulation; MDR – Medical Devices Regulation.

Data synthesis and categorization

Articles were classified into thematic domains reflecting the main challenges and areas of AI implementation in medicine. The categorization was aligned with the analytical framework presented in the following section and it included the following four categories:

1. AI in clinical diagnostics and decision support systems (e.g., imaging interpretation, risk stratification tools, or early disease detection)

2. Ethical, legal, and regulatory challenges (e.g., liability in algorithm-based decisions, informed consent, or transparency standards)
3. Data privacy, cybersecurity, and model robustness (e.g., adversarial attacks, bias mitigation, or reproducibility problems)
4. Patient experience and system integration (e.g., hybrid models in practice, overreliance on automation, or equity in access to AI tools)

Legal acts, strategic documents, selected foundational reviews, and key regulatory guidelines (e.g., the AI Act, GMLP, HIPAA, or the GDPR) were analyzed to support the conceptual framework and provide context for thematic classification and regulatory interpretation.

Elaboration of results and synthesis

Each category was discussed in terms of benefits, limitations, barriers to implementation, and identified research gaps. Due to the methodological heterogeneity of the included studies, no quantitative meta-analysis was performed. A visual outline of the review process and analytical stages is presented in Figure 1.

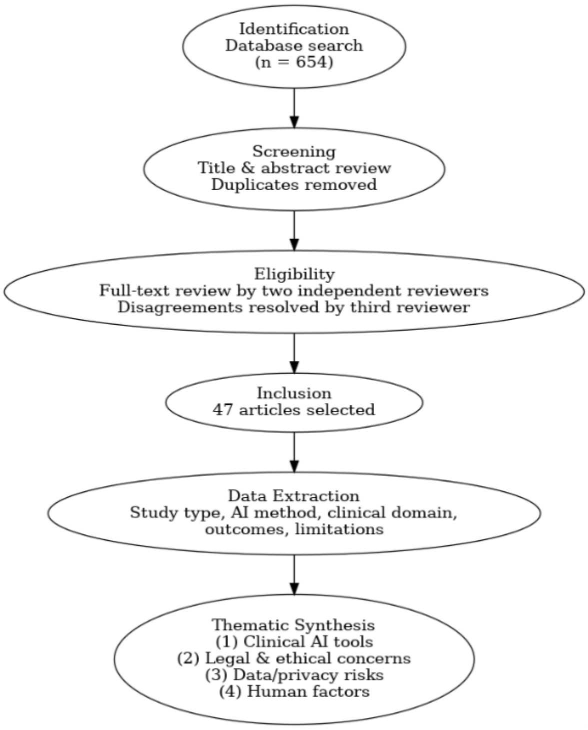

Fig. 1. Research methodology.

LEGAL AND REGULATORY ASPECTS OF AI USE IN MEDICINE

The application of AI in medicine requires the consideration of numerous legal conditions and regulatory standards aimed at protecting patients, ensuring transparency about the algorithms, and assigning responsibility to the entities involved. Due to the potential health and safety risks, AI systems used in medicine are classified as high-risk technologies and are subject to strict oversight [14,17].

In the EU, the key legal framework is the AI Act, which defines rules for AI usage based on risk levels. Systems used in healthcare, including diagnostics and treatment, are classified as high-risk, requiring compliance with stringent requirements. The AI Act mandates human oversight, algorithm transparency, auditability of decision-making processes, and operational risk management [17].

These requirements are further elaborated in the GMLP, a set of guidelines developed by organizations such as the FDA and international bodies assessing ML systems. GMLP emphasizes data quality, version control of models, algorithmic transparency, and the ability to monitor and audit systems in clinical environments [13]. Additionally, a Predetermined Change Control Plan is required, which defines the rules for modifying algorithm behavior without compromising user safety.

Data protection obligations are regulated by the GDPR in the EU and HIPAA in the USA. These regulations impose specific duties on entities processing medical data, which is recognized as sensitive information. The use of AI for processing such data requires explicit patient consent, anonymization mechanisms, and compliance with principles such as data minimization, transparency, and limitation of purposes [18,19].

AI systems performing diagnostic or therapeutic functions are also covered by the EU Medical Devices Regulation (MDR), which classifies medical software (including AI) as medical devices and imposes obligations regarding certification, clinical evaluation, and safety [20]. Likewise, the FDA has implemented the AI/ML-based SaMD Action Plan, which includes risk assessment, clinical validation, continuous monitoring, and verification of algorithm performance in real-world settings [21].

Mental health support applications using AI, while not always classified as medical devices, can still pose risks to user safety. According to Articles 6 and 14 of the AI Act, if these systems threaten fundamental rights or user health, they must be treated as high-risk and comply with evaluation and transparency standards. If such an application has diagnostic or therapeutic functionality, it may also fall under the MDR, requiring conformity and clinical safety assessment procedures. When the tool is not considered a medical device, GDPR provisions still apply [18]. In the USA, such apps are often beyond FDA oversight unless they meet the SaMD definition, which would require a clinical impact assessment and compliance with FDA guidelines [21].

Another important issue is the attribution of responsibility for decisions made by AI. According to the draft EU directive, liability may be shared among the front-end operator (e.g., physician), back--end operator (system provider), and software manufacturer. Each party is required to oversee the system, report incidents, and ensure data input quality. In practice, this means that in case of a diagnostic error caused by AI, responsibility may lie with the user, the system developer, or the implementing institution [14].

AI HALLUCINATIONS AND THE GENERATION OF FALSE INFORMATION

One of the most serious threats related to the implementation of artificial intelligence in healthcare is the generation of incorrect or fabricated information by AI models, particularly generative ones. For instance, AI can incorrectly diagnose cancer based on an x-ray image, leading to unnecessary medical procedures [1,22].

As discussed in the legal section, AI systems classified as high-risk are subject to human oversight, event logging, and strict transparency requirements, including compliance with GMLP [13]. To reduce the occurrence of AI hallucinations, it is crucial to use diverse, representative data sets and to conduct independent clinical validations. Standards such as CONSORT-AI and SPIRIT-AI help ensure transparency and reliability in reporting study results involving AI models, thereby reducing the risk of inaccurate diagnostic recommendations [23].

LACK OF TRANSPARENCY REGARDING ALGORITHMS (BLACK BOX PROBLEM)

AI models often function as a “black box,” meaning that their decision-making processes are not transparent. This can lead to distrust and difficulties in assessing the risk of error [24]. In the event of an incorrect diagnosis, it becomes unclear who is responsible: the physician, the software developer, or the medical institution. The opacity of AI models also makes it difficult for doctors to trust their recommendations [25].

As discussed above, the EU’s AI Act, the FDA’s AI/ML-based SaMD Action Plan, and GMLP require that high-risk systems be audited – subject to human oversight – and provide interpretability wherever possible. These regulations aim to increase accountability and allow clinicians to make decisions based on understandable input and algorithm outputs [13,14].

One proposed solution to the black box problem is Explainable Artificial Intelligence (XAI), which aims to enable users to understand why an algorithm made a specific decision. However, the development of such solutions remains limited – especially in the case of complex DL models, where interpreting the functioning of neural networks remains challenging [26].

In clinical practice, hybrid models are increasingly being used; they combine AI capabilities with traditional medical knowledge and physician experience. This approach fosters collaboration between the algorithm and the specialist – AI analyzes the data (e.g., imaging or lab results), but the final interpretation and therapeutic decision remain in human hands. Hybrid models increase physicians’ trust in technology because they support rather than replace the physician’s role and they improve interpretability and legal compliance [27,28]. Such solutions promote the responsible, flexible use of AI in everyday clinical practice [29].

REPRODUCIBILITY ISSUES IN AI-GENERATED RESULTS

Reproducibility is a cornerstone of medical practice, yet AI systems often generate different outcomes even when analyzing identical data. This can lead to ambiguous diagnoses and inconsistent therapeutic recommendations [30]. In a study on cancer diagnostics, it was found that AI algorithms analyzing the same MRI scans produced contradictory results due to minor differences in input parameters [16].

As discussed in the legal section, EU legislation (the AI Act and the MDR) obliges AI system manufacturers to ensure the stability of their models and to document the decision-making process. The draft directive on AI liability also requires access to documentation from which the algorithm’s decision can be reconstructed [14]. To enhance consistency, it is necessary to implement mechanisms for “freezing” model versions and maintaining version control of algorithms. This can help reduce the risk of non-reproducible outcomes [16].

DATA QUALITY AND DIAGNOSTIC ERRORS

AI systems learn from historical data, which are often incomplete, biased, or unrepresentative. These limitations can lead to serious diagnostic errors [23,31,32]. One example is reduced accuracy in detecting skin cancer in patients with darker skin tones, because training datasets primarily included images of lighter-skinned individuals [24,30]. Another case involved an AI model used to assess pneumonia risk, which incorrectly assumed that asthma patients had a lower risk of complications because they typically received more intensive care [33]. In oncology, AI systems have shown lower detection rates for tumors in women from ethnic minorities, as the training data came mostly from white patients [34]. Additionally, some algorithms may misinterpret textual data, resulting in diagnostic errors. For instance, a system analyzing electronic health records misclassified a patient's occasional alcohol consumption as alcohol addiction, which led to exclusion from a liver transplant list [25].

Another issue is the limited portability of models across regions. Algorithms trained on data from the USA may underperform in other countries, and those trained in English may struggle to interpret medical records in other languages [14].

As outlined in the legal and regulatory section, regulations such as Article 10 of the AI Act and GMLP require AI developers to monitor data quality, manage data sources, and eliminate algorithmic bias [13]. The use of a Predetermined Change Control Plan is crucial to avoid uncontrolled changes in model performance. The use of systematic algorithm audits and diverse training datasets are also recommended [34], significantly improving diagnostic accuracy and patient safety.

OVERRELIANCE ON AI AND THE DEHUMANIZATION OF MEDICINE

While AI supports diagnostic and therapeutic decisions, excessive reliance on it can lead to negative consequences. For example, an AI system used for diagnosing heart disease overlooked critical clinical factors, resulting in incorrect therapeutic recommendations [1]. Studies show that physicians using AI-assisted radiological image analysis are more likely to miss significant abnormalities when relying blindly on algorithmic suggestions [35]. This phenomenon, known as "cognitive automation," may limit critical thinking in clinical practice. Physicians, convinced of AI's accuracy, might lower their diagnostic vigilance, thus compromising the quality and safety of care [1]. Furthermore, over-automation of the diagnostic process can adversely affect the doctor–patient relationship. Human contact becomes limited and decisions are based on data analysis without considering the patient's individual experiences and needs. Patients may feel reduced to a data set, which can diminish empathy and trust in the treatment process [26]. This issue is also discussed in the legal context, where Article 14 of the AI Act requires that the final medical decision be made by a human – not by an algorithm [27].

THREATS IN AI APPLICATIONS THAT SUPPORT MENTAL HEALTH

Generative AI is increasingly being used in wellness applications that support mental health. Although these tools can complement traditional therapy, studies indicate that improper functioning may lead to serious consequences. For example, AI chatbots have been reported to respond inappropriately to suicidal thoughts. In one study involving five popular AI-based mental health apps, 50% of them failed to provide any assistance and, in extreme cases, responses exacerbated the crisis – such as replying "Don't u coward" to suicidal ideation [36]. AI may also misjudge a patient's condition and fail to direct them to professional help.

The legal requirements for such tools are detailed in the section on legal and regulatory aspects above. These include the classification of such applications as high-risk systems under the AI Act, the possibility of categorization as medical devices under the MDR, and the obligations stemming from the GDPR and FDA regulations in the USA. The lack of a unified regulatory approach to AI-powered wellness apps creates risks for users, especially in mental health support. As a result, both in the EU and the USA, there is growing advocacy for stricter regulation and mandatory referral to professional help in crisis situations [36,37].

CYBERSECURITY AND PROTECTION OF PATIENT DATA

AI systems in medicine operate on large datasets containing sensitive information, such as personal data, imaging results, genetic data, or medical histories. Due to their confidential nature, these data are an attractive target for cybercriminals, making AI systems particularly vulnerable. Although modern ML models are highly effective at data analysis, they are also susceptible to various types of attacks that can result in serious privacy breaches and diagnostic errors [38,39]. One of the most dangerous threats is adversarial attacks, which involve introducing subtle, often imperceptible changes to medical images. These manipulations can completely mislead the algorithm, resulting in misclassifications or incorrect diagnoses – even with a confidence level of 99% [38,40]. Another risk includes membership inference attacks, which aim to determine whether a specific patient's data was used in model training. Such intrusions can lead to the disclosure of personal or health-related information [41]. Model inversion attacks also pose a threat, as they can partially reconstruct input data – such as an MRI image – compromising patient privacy and violating data protection regulations [42].

With the growing role of health data in training AI models, there is an increasing risk of commercial misuse. Technology companies are increasingly leveraging data collected from mobile apps, wearables, or diagnostic systems – not only for medical purposes, but also for marketing and often without the patient's full awareness of this use of their data. As Kanter and Packel [43] note, privacy is often infringed within vague legal boundaries and patients do not always have real control over how their data are utilized. Rajpurkar et al. [44] highlight the need to develop clear oversight and control mechanisms for AI models processing health data. This is especially important in the case of wellness apps or AI chatbots that do not formally fall under medical device regulations. Consequently, patients may unknowingly consent to the use of their data for non-medical purposes, undermining trust in the patient–technology relationship.

Access to medical data is regulated by HIPAA (USA), the GDPR (EU), and the AI Act, which specify who may process patient data and under what conditions. The AI Act introduces additional safeguards to prevent unauthorized access to sensitive data, thereby enhancing protection [13]. Ensuring cybersecurity must be a top priority, particularly in the context of regulations such as GDPR and HIPAA, which mandate the protection of patient data. Risk minimization requires the use of advanced encryption technologies, multi-factor authentication, and real-time user activity monitoring. Some hospitals implement early warning systems that detect unusual access patterns and automatically block suspicious activity [43,44].

INFORMED PATIENT CONSENT FOR THE USE OF AI

Medical law requires that informed consent be obtained from patients for all medical interventions, which also includes the use of AI in diagnosis and treatment. It is essential that the patient is properly informed about the role of AI in the clinical decision-making process, its potential limitations, associated risks, and available therapeutic alternatives. This approach is reflected in both legal regulations and policy documents, such as the AI Act and the "White Paper on AI in Clinical Practice" [14].

As technology develops, patients are increasingly becoming active users of AI systems – using health monitoring applications, self-diagnosis tools, and remote therapeutic support systems. However, this role requires appropriate education and support. A low level of digital literacy and a lack of understanding of algorithmic operations can lead to misinterpreted results and improper use of technology, especially among older individuals or those who are digitally excluded. Lichosik [45] emphasizes that medical personnel should act as guides through technology: supporting patients in making informed decisions and overseeing the functioning and use of AI systems in clinical practice. Many currently available solutions are not user-friendly – they are too complex and poorly adapted to the realities of underdeveloped healthcare facilities [4].

In this context, increasing attention is being paid to the need to distinguish between general consent, related to traditional medical procedures, and so-called technological consent, which pertains specifically to the use of AI tools such as AI-CDSSs. This consent should inform patients of how the algorithm works, its limitations, its level of autonomy, and who is responsible for the decisions it makes [14,30,39]. Transparency regarding the functioning of AI is a necessary condition for ensuring truly informed consent: the patient must know how their data is processed and how it affects clinical decisions [39]. As noted by Amann et al. [29], the explainability of algorithms is the foundation of trust and ethical AI use in healthcare.

COSTS AND INEQUALITIES IN ACCESS TO AI

One of the challenges of implementing AI in medicine is the high cost, which can exacerbate inequalities in access to modern diagnostics. The EU foresees subsidizing the implementation of AI in public healthcare, as well as programs for standardizing access [14]. European regulations emphasize the need to ensure equal access to AI and to eliminate financial barriers, for example, through state subsidies [27]. Studies published in *Health Technology Assessment* indicate that integrating artificial intelligence can involve a significant investment of time and money. This process requires thorough testing and validation to ensure proper system performance in real clinical conditions [46]. Including AI in reimbursement strategies that promote value-based care could provide significant incentives for the development and implementation of validated algorithms. In developed countries, the private sector is showing interest in investing in AI and other innovative healthcare technologies. Nevertheless, the high costs associated with purchasing scanners for digitizing histopathological specimens and specialized computer equipment pose a significant barrier to the adoption of AI tools in everyday clinical practice, especially in smaller and underfunded facilities. Additionally, the development of large, centralized image archives and databases, along with the purchase of hardware and software, requires substantial capital investment – which may be unattainable for smaller institutions [46,47].

The failure to consider local conditions and the limited resources of smaller medical facilities may lead to increasing disparities in access to advanced technologies. Research shows that AI-CDSS systems implemented in clinics and hospitals with limited infrastructure often turn out to be misaligned with the local context: they fail to account for real-world resources and are overly complex or poorly integrated with existing systems [4]. As a result, these technologies are used less frequently or completely abandoned, which can lead to the digital exclusion of smaller centers, further deepening existing disparities in the quality of healthcare services [3].

Although the implementation of AI in daily diagnostics involves considerable costs, it also brings about numerous benefits. The use of this technology can ultimately reduce healthcare expenses by limiting the need to repeat costly tests and minimizing the risk of errors. AI also increases access to specialized medical care, particularly in regions with shortages of healthcare personnel. Advanced analytical technologies can automate many routine diagnostic tasks, such as analyzing medical images, which reduces professionals' workloads and lowers the costs of conducting and interpreting tests. Medical errors, including incorrect interpretations of results, can generate additional expenses and delay treatment. Through precise data analysis, AI helps to reduce such occurrences. Furthermore, patient monitoring systems can also decrease the number of doctor's visits and hospitalizations, directly translating into reduced overall healthcare costs. Studies published in *Health Affairs* and the *Journal of the American College of Radiology* show that using AI in radiology can lower diagnostic costs by up to 20% and can reduce the error rate by up to 15%, significantly contributing to reduced treatment expenses [46].

CONCLUSIONS AND PROSPECTS

The application of AI in medicine entails both hope and serious controversy. Modern technology offers solutions with much diagnostic, analytical, and organizational potential. However, its responsible integration into clinical practice must be based on a critical perspective – one that considers both its limitations and the ethical boundaries of implementation.

Limitations of the approach and sources

Although this work adopted a narrative approach, a wide range of literature was utilized, including sources from both highly developed countries and those with limited access to advanced technologies (e.g., Poland, Croatia, Germany, and the United Kingdom). Nevertheless, there is a risk of selectivity and a lack of empirical data regarding local implementations of AI. A more systematic approach is needed to assess the quality of data [44].

Recommendations for practice and future research

1. Adapting AI to clinical realities – AI solutions should take into consideration the conditions of smaller facilities: their infrastructure, staffing levels, and degree of digitalization [4].
2. Increasing focus on education and digital literacy – Misunderstanding how AI systems work, among both physicians and patients, may lead to incorrect clinical decisions and a lack of trust [45].
3. Developing responsibility and oversight frameworks – Clear regulations are needed to define accountability for decisions made with the use of AI. There is also a need to standardize auditing tools for algorithms.
4. Interdisciplinary implementation teams – Only collaboration between physicians, computer scientists, ethicists, and legal experts can ensure that AI is implemented with respect for clinical and humanistic values [30].
5. Explainable AI as a necessary condition – Technology must be not only effective, but also understandable. A physician cannot make decisions based on a "black box" without the ability to verify the algorithm's reasoning [25].
6. Technological consent – Patients should be informed about the fact that decisions regarding their therapy are partially supported by an AI system, in addition to any medical risks [39].

7. Support for participatory research (co-design) – It is recommended to conduct research involving doctors, nurses, and patients even at the design stage of AI tools. This approach enhances the acceptability, effectiveness, and alignment of solutions with clinical realities. It better adapts the technology to the actual needs of users and increases their engagement and trust in new systems [48,49].
8. Strengthening education in ethics and AI in medicine – Introducing educational modules that address not only the technical aspects, but also the ethical, legal, and communication issues related to the use of AI can help prepare future doctors for the responsible, informed use of these tools [50,51].

Limits to AI development in medicine

Technological advancement must not become an end in itself. Not everything that is technically possible should be uncritically implemented in healthcare. Examples such as incorrect recommendations from mental health chatbots [36] or racial and gender biases in algorithms [35] show that without proper oversight, AI can cause real harm. The limits are defined not only by computational capabilities, but primarily by issues of trust, patient dignity, and physician autonomy. The dehumanization of the doctor–patient relationship, the decline of clinical experience, or the risk of digital exclusion are costs that must not be accepted in the name of efficiency.

Critical conclusions

The development of AI in medicine should be conscious, gradual, and guided by caution. Technology may support the physician, but it should not replace clinical judgment. The usefulness of algorithms must be balanced with their explainability and social, ethical acceptance. AI should not serve to justify cost-cutting measures that worsen the quality of care; it must be a tool that supports responsible and integrated medical practice.

A key direction for future research should therefore not only be the development of algorithms, but also an understanding of how they are perceived and interpreted in real clinical settings. The future of AI depends on whether it can be rooted in the values defined by medicine: respect, safety, and patient dignity.

SUMMARY

The conclusions lead to a clear picture: artificial intelligence offers numerous benefits in terms of improving the effectiveness and accessibility of medical services, but it also entails real risks. Algorithms can streamline diagnosis and support clinical decisions, but they are not immune to errors, biases, or interpretation issues. Uncontrolled automation may diminish the physician's role, while unclear legal responsibility increases the risks for both patients and institutions.

Responsible AI implementation requires both advanced technology and well-thought-out legal, ethical, and organizational frameworks. It is necessary to build public trust, develop transparent operational mechanisms, and involve interdisciplinary teams in the design and testing of tools. In the coming years, special attention must be paid to researching how AI systems are understood and accepted by users in clinical practice. It will be equally important to adapt technology to local conditions and to foster open dialogue between the developers, users, and beneficiaries of AI systems. Only such an approach will allow for the real, safe, and equitable utilization of AI's potential in healthcare.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author's contribution

Study design – J. Kasperczyk, K. Lau, H. Brawańska
Data collection – H. Brawańska, K. Lau
Manuscript preparation – H. Brawańska, M. Jędrzejowska, J. Kasperczyk
Literature research – H. Brawańska, M. Jędrzejowska
Final approval of the version to be published – J. Kasperczyk, H. Brawańska, M. Jędrzejowska

REFERENCES

1. Lee P., Bubeck S., Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N. Engl. J. Med. 2023; 388(13): 1233–1239, doi: 10.1056/NEJMsr2214184.

2. He J., Baxter S.L., Xu J., Xu J., Zhou X., Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat. Med. 2019; 25(1): 30–36, doi: 10.1038/s41591-018-0307-0.

3. Gomez-Cabello C.A., Borna S., Pressman S., Haider S.A., Haider C.R., Forte A.J. Artificial-Intelligence-based Clinical Decision Support Systems in Primary Care: A scoping review of current clinical implementations. Eur. J. Investig. Health Psychol. Educ. 2024; 14(3): 685–698, doi: 10.3390/ejihpe14030045.

4. Wang L., Zhang Z., Wang D., Cao W., Zhou X., Zhang P. et al. Human-centered design and evaluation of AI-empowered clinical decision support systems: a systematic review. Front. Comput. Sci. 2023; 5: 1187299, doi: 10.3389/fcomp.2023.1187299.

5. Davis A., Billick K., Horton K., Jankowski M., Knoll P., Marshall J.E. et al. Artificial intelligence and echocardiography: A primer for cardiac sonographers. J. Am. Soc. Echocardiogr. 2020; 33(9): 1061–1066, doi: 10.1016/j.echo.2020.04.025.

6. Mintz Y., Brodie R. Introduction to artificial intelligence in medicine. Minim. Invasive Ther. Allied Technol. 2019; 28(2): 73–81, doi: 10.1080/13645706.2019.1575882.

7. Wang J., Chen Y., Yu S.X., Cheung B., LeCun Y. Compact and optimal deep learning with recurrent parameter generators. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE 2023, doi: 10.1109/wacv56688.2023.00389.

8. Khosravi B., Li F., Dapamede T., Rouzrokh P., Gamble C.U., Trivedi H.M. et al. Synthetically enhanced: unveiling synthetic data's potential in medical imaging research. EBioMedicine 2024; 104: 105174, doi: 10.1016/j.ebiom.2024.105174.

9. Tangsrivimol J.A., Darzidehkalani E., Virk H.U.H., Wang Z., Egger J., Wang M. et al. Benefits, limits, and risks of ChatGPT in medicine. Front. Artif. Intell. 2025; 8: 1518049, doi: 10.3389/frai.2025.1518049.

10. Rajotte J.F., Bergen R., Buckeridge D.L., El Emam K., Ng R., Strome E. Synthetic data as an enabler for machine learning applications in medicine. iScience 2022; 25(11): 105331, doi: 10.1016/j.isci.2022.105331.

11. Larentzakis A., Lygeros N. Artificial intelligence (AI) in medicine as a strategic valuable tool. Pan Afr. Med. J. 2021; 38: 184, doi: 10.11604/pamj.2021.38.184.28197.

12. Niewęgłowski K., Wilczek N., Madoń B., Palmi J., Wasyluk M. Zastosowania sztucznej inteligencji (AI) w medycynie. Med. Og. Nauk Zdr. 2021; 27(3): 213–219, doi: 10.26444/monz/142085.

13. Palaniappan K., Lin E.Y.T., Vogel S. Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector. Healthcare 2024; 12(5): 562, doi: 10.3390/healthcare12050562.

14. Wałdoch K. Odpowiedzialność cywilna za szkody wyrządzone w związku z zastosowaniem sztucznej inteligencji w medycynie. [Doctoral thesis]. Uniwersytet Gdański. Gdańsk 2024.

15. Haug C.J., Drazen J.M. Artificial intelligence and machine learning in clinical medicine, 2023. N. Engl. J. Med. 2023; 388(13): 1201–1208, doi: 10.1056/NEJMra2302038.

16. Kataoka M., Uematsu T. AI systems for mammography with digital breast tomosynthesis: expectations and challenges. Radiol. Imaging Cancer 2024; 6(4): e240171, doi: 10.1148/rycan.240171.

17. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). EUR-Lex [online] http://data.europa.eu/eli/reg/2024/1689/oj [accessed on May 2025].

18. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). EUR-Lex [online] http://data.europa.eu/eli/reg/2016/679/oj [accessed on May 2025].

19. H.R.6216 – National Artificial Intelligence Initiative Act of 2020 (116th Congress [2019–2020]). Congress.gov, 03/12/2020 [online] https://www.congress.gov/bill/116th-congress/house-bill/6216 [accessed on May 2025].

20. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance). EUR-Lex [online] http://data.europa.eu/eli/reg/2017/745/oj [accessed on May 2025].

21. U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan [pdf], January 2021, https://www.fda.gov/media/145022/download.

22. Vodanović M., Subašić M., Milošević D., Savić Pavičin I. Artificial intelligence in medicine and dentistry. Acta Stomatol. Croat. 2023; 57(1): 70–84, doi: 10.15644/asc57/1/8.

23. Bandyopadhyay A., Oks M., Sun H., Prasad B., Rusk S., Jefferson F. et al. Strengths, weaknesses, opportunities, and threats of using AI-enabled technology in sleep medicine: a commentary. J. Clin. Sleep Med. 2024; 20(7): 1183–1191, doi: 10.5664/jcsm.11132.

24. Cestonaro C., Delicati A., Marcante B., Caenazzo L., Tozzo P. Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. Front. Med. 2023; 10: 1305756, doi: 10.3389/fmed.2023.1305756.

25. Czochra M., Bar D. Śmierć pacjenta wywołana zastosowaniem sztucznej inteligencji w technologiach medycznych – analiza prawnokarna. Stud. Law Res. Paper. 2019; 2(25): 67–81, doi: 10.34697/2451-0807-sp-2019-2-006.

26. Bączyk-Rozwadowska K. Odpowiedzialność cywilna za szkody wyrządzone w związku z zastosowaniem sztucznej inteligencji w medycynie. PPM 2021; 3(3–4): 5–35, doi: 10.70537/z7xnk378.

27. Ferry J., Laberge G., Aïvodji U. Learning hybrid interpretable models: Theory, taxonomy, and methods. arXiv. 2023; arXiv: 2303.04437, doi: 10.48550/arXiv.2303.04437.

28. Wang T., Lin Q. Hybrid predictive models: When an interpretable model collaborates with a black-box model. J. Mach. Learn. Res. 2021; 22(137): 1–38.

29. Amann J., Blasimme A., Vayena E., Frey D., Madai V.I. et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med. Inform. Decis. Mak. 2020; 20(1): 310, doi: 10.1186/s12911-020-01332-6.

30. Habuza T., Navaz A.N., Hashim F., Alnajjar F., Zaki N., Serhani M.A. et al. AI applications in robotics, diagnostic image analysis and precision medicine: Current limitations, future trends, guidelines on CAD systems for medicine. Inform. Med. Unlocked 2021; 24: 100596, doi: 10.1016/j.imu.2021.100596.

31. Briganti G., Le Moine O. Artificial intelligence in medicine: today and tomorrow. Front. Med. 2020; 7: 27, doi: 10.3389/fmed.2020.00027.

32. Kimmerle J., Timm J., Festl-Wietek T., Cress U., Herrmann-Werner A. Medical students' attitudes toward AI in medicine and their expectations for medical education. J. Med. Educ. Curric. Dev. 2023; 10: 23821205231219346, doi: 10.1177/23821205231219346.

33. Tomašev N., Glorot X., Rae J.W., Zielinski M., Askham H., Saraiva A. et al. Developing deep learning continuous risk models for early adverse event prediction in electronic health records: an AKI case study. Protocol Exchange 2019, doi: 10.21203/rs.2.10083/v1.

34. Parikh R.B., Teeple S., Navathe A.S. Addressing bias in artificial intelligence in health care. JAMA 2019; 322(24): 2377–2378, doi: 10.1001/jama.2019.18058.

35. Homolak J. Opportunities and risks of ChatGPT in medicine, science, and academic publishing: a modern Promethean dilemma. Croat. Med. J. 2023; 64(1): 1–3, doi: 10.3325/cmj.2023.64.1.

36. De Freitas J., Cohen I.G. The health risks of generative AI-based wellness apps. Nat. Med. 2024; 30(5): 1269–1275, doi: 10.1038/s41591-024-02943-6.

37. Milne-Ives M., Selby E., Inkster B., Lam C., Meinert E. Artificial intelligence and machine learning in mobile apps for mental health: A scoping review. PLOS Digit. Health 2022; 1(8): e0000079, doi: 10.1371/journal.pdig.0000079.

38. Ma X., Niu Y., Gu L., Wang Y., Zhao Y., Bailey J. et al. Understanding adversarial attacks on deep learning based medical image analysis systems. Pattern Recognit. 2021; 110: 107332, doi: 10.1016/j.patcog.2020.107332.

39. Kiener M. Artificial intelligence in medicine and the disclosure of risks. AI Soc. 2020; 36(3): 705–713, doi: 10.1007/s00146-020-01085-w.

40. Finlayson S.G., Bowers J.D., Ito J., Zittrain J.L., Beam A.L., Kohane I.S. Adversarial attacks on medical machine learning. Science 2019; 363(6433): 1287–1289, doi: 10.1126/science.aaw4399.

41. Hu H., Salcic Z., Sun L., Dobbie G., Yu P.S., Zhang X. Membership inference attacks on machine learning: A survey. ACM Comput. Surv. 2022; 54(11s): 1–37, doi: 10.1145/3523273.

42. Newaz A.I., Haque N.I., Sikder A.K., Rahman M.A., Uluagac A.S. Adversarial attacks to machine learning-based smart healthcare systems. GLOBECOM 2020 – 2020 IEEE Global Communications Conference, p. 1–6, doi: 10.1109/GLOBECOM42002.2020.9322472.

43. Kanter G.P., Packel E.A. Health care privacy risks of AI chatbots. JAMA 2023; 330(4): 311–312, doi: 10.1001/jama.2023.9618.

44. Rajpurkar P., Chen E., Banerjee O., Topol E.J. AI in health and medicine. Nat. Med. 2022; 28(1): 31–38, doi: 10.1038/s41591-021-01614-0.

45. Lichosik D. Opieka i leczenie onkologiczne pacjenta w dobie sztucznej inteligencji (AI). [Film / Speech]. Baza Wiedzy Akademii Tarnowskiej, 2024 [online] https://rpt.atar.edu.pl/info/media/UAST9dd433f2f93e40bbb3a84ed9497b246d?ps=20&lang=pl&title=&pn=1&cid=23299 2023.

46. Barański J. Intelligent revolution in medicine – the application of artificial intelligence (ai) in medicine: overview, benefits, and challenges. Przegl. Epidemiol. 2024; 78(3): 287–302, doi: 10.32394/pe/194484.

47. Ahmad Z., Rahim S., Zubair M., Abdul-Ghafar J. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. Diagn. Pathol. 2021; 16(1): 24, doi: 10.1186/s13000-021-01085-4.

48. Donia J., Shaw J.A. Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. Big Data Soc. 2021; 8(2): 20539517211065248, doi: 10.1177/20539517211065248.

49. Benz C., Scott-Jeffs W., McKercher K.A., Welsh M., Norman R., Hendrie D. et al. Community-based participatory-research through co-design: supporting collaboration from all sides of disability. Res. Involv. Engagem. 2024; 10(1): 47, doi: 10.1186/s40900-024-00573-3.

50. Weidener L., Fischer M. Teaching AI ethics in medical education: A scoping review of current literature and practices. Perspect. Med. Educ. 2023; 12(1): 399–410, doi: 10.5334/pme.954.

51. Alam F., Lim M.A., Zulkipli I.N. Integrating AI in medical education: embracing ethical usage and critical understanding. Front. Med. 2023; 10: 1279707, doi: 10.3389/fmed.2023.1279707.

52. Biała Księga AI w praktyce klinicznej: stosowanie sztucznej inteligencji przy udzielaniu świadczeń zdrowotnych [pdf]. [Wersja 1.0]. Koalicja AI w Zdrowiu / Grupa Robocza ds. Sztucznej Inteligencji / Polska Federacja Szpitali. wZdrowiu. Warszawa, czerwiec 2022, https://aiwzdrowiu.pl/wp-content/uploads/2022/06/BIA_A-KSIE_GA_AI-W-ZDROWIU_2022.pdf [accessed on May 2025].