Revista Bioética
Print version ISSN 1983-8042 | On-line version ISSN 1983-8034

Rev. Bioét. vol.33 Brasília 2025

Between the algorithm and the Hippocratic Oath: bioethics in the age of artificial intelligence

Juracy Barbosa dos Santos [1], *Guilhermina Rego* [1], *Rui Nunes* [1]

1. Faculdade de Medicina do Porto, Universidade do Porto, Porto, Portugal.

Abstract

The advancement of artificial intelligence has profoundly transformed medical practice. From clinical decision-support systems to screening and diagnostic algorithms, artificial intelligence has demonstrated potential for early diagnosis, personalized therapies, resource optimization, error reduction, and expanded access to specialized care. However, this technological revolution imposes significant ethical challenges on the classic principles of bioethics, such as beneficence, non-maleficence, confidentiality, and respect for patient autonomy, enshrined since the Hippocratic Oath. The incorporation of artificial intelligence into medicine raises tensions between automated efficiency and human values. This article reviews the literature on the primary ethical dilemmas of applying artificial intelligence in medical practice and evaluates to what extent Hippocratic principles can be preserved or adapted in the face of technological transformations, aiming to contribute to the debate on the ethical directions of contemporary medicine in a scenario of increasing automation.

Keywords: Bioethics. Artificial intelligence. Ethics, medical. Personal autonomy.

Resumo

Entre o algoritmo e o Juramento de Hipócrates: bioética na era da inteligência artificial

O avanço da inteligência artificial tem transformado profundamente a prática médica. De sistemas de apoio à decisão clínica a algoritmos de triagem e diagnóstico, a inteligência artificial tem demonstrado potencial para diagnósticos precoces, terapias personalizadas, otimização de recursos, redução de erros e ampliação do acesso a cuidados especializados. Contudo, essa revolução tecnológica impõe desafios éticos significativos aos princípios clássicos da bioética, como beneficência, não maleficência, confidencialidade e respeito à autonomia do doente, consagrados desde o Juramento de Hipócrates. A incorporação da inteligência artificial na medicina suscita tensões entre eficiência automatizada e valores humanos. Este artigo revisa a literatura sobre os principais dilemas éticos da aplicação da inteligência artificial na prática médica e avalia em que medida os princípios hipocráticos podem ser preservados ou adaptados diante das transformações tecnológicas, a fim de contribuir para o debate sobre os rumos éticos da medicina contemporânea em um cenário de crescente automação.

Palavras-chave: Bioética. Inteligência artificial. Ética médica. Autonomia pessoal.

Resumen

Entre el algoritmo y el Juramento Hipocrático: bioética en la era de la inteligencia artificial

El avance de la inteligencia artificial ha transformado profundamente la práctica médica. Desde sistemas de apoyo a la decisión clínica hasta algoritmos de triaje y diagnóstico, la inteligencia artificial ha demostrado potencial para diagnósticos precoces, terapias personalizadas, optimización de recursos, reducción de errores y ampliación del acceso a cuidados especializados. No obstante, esta revolución tecnológica impone desafíos éticos significativos a los principios clásicos de la bioética, tales como la beneficencia, la no maleficencia, la confidencialidad y el respeto a la autonomía del paciente, consagrados desde el Juramento Hipocrático. La incorporación de la inteligencia artificial en la medicina suscita tensiones entre la eficiencia automatizada y los valores humanos. Este artículo revisa la literatura sobre los principales dilemas éticos de la aplicación de la inteligencia artificial en la práctica médica y evalúa en qué medida los principios hipocráticos pueden preservarse o adaptarse ante las transformaciones tecnológicas, con el fin de contribuir al debate sobre los rumbos éticos de la medicina contemporánea en un escenario de creciente automatización.

Palabras clave: Bioética. Inteligencia artificial. Ética médica. Autonomía personal.

The authors declare no conflict of interest.

Research

The introduction of artificial intelligence (AI) in medicine represents one of the greatest technological leaps in recent health history. From clinical decision support systems to screening and diagnostic algorithms, AI has demonstrated potential for early diagnosis, personalized therapies, resource optimization, reduced medical errors, and expanded access to specialized care, especially in remote regions or areas with a shortage of qualified healthcare professionals [1,2]. In areas such as dermatology, radiology, and cardiology, deep learning-based algorithms already demonstrate performance comparable to, or even superior to, that of human experts in specific tasks [3].

However, the rapid expansion and incorporation of AI into clinical practice does not occur without challenges. The accelerated adoption of these technologies raises significant ethical concerns, particularly regarding the fundamental principles of bioethics. Since its beginnings, medicine has been guided by ethical principles aimed at protecting and promoting patients' well-being. The Hippocratic Oath, dating from the 5th century BC, enshrined fundamental values such as beneficence, non-maleficence, confidentiality, and respect for patient autonomy—pillars of bioethics even today [4]. These values were reinforced in the 20th century by documents such as the Nuremberg Code and the Declaration of Helsinki, which consolidated the foundations for ethics in research and medical practice [3].

However, the advent of AI in medicine introduces new and complex ethical challenges that strain these principles. Among the dilemmas that emerge are the opacity and lack of explainability of algorithms—the so-called black boxes—which make it difficult for physicians and patients to understand and validate automated decisions. This lack of algorithmic transparency compromises the principle of autonomy by preventing fully informed consent. In addition, questions arise about accountability for clinical errors resulting from AI-based decisions, with doubts as to whether the blame falls on the physician, the programmer, the health institution, or the algorithm itself [2,5,6].

The privacy and protection of sensitive patient data are other critical points. Since AI relies on large volumes of data for training and operation, it is essential to ensure that the handling of information complies with laws and regulations such as Law 14,510/2022, which provides for the provision of telehealth services [7], the General Data Protection Law (LGPD, Law 13,709/2018) [8] and Resolution 2,314/2022 of the Federal Council of Medicine (CFM), which defines and regulates telemedicine as a form of providing medical services [9], in Brazil; the General Data Protection Regulation (GDPR) [10] and the Network and Information Systems (NIS Directive) [11] in the European Union; the Health Insurance Portability and Accountability Act (HIPAA) [12] and the Health Information Technology for Economic and Clinical Health Act (HITECH) [13] in the United States. Violation of privacy or misuse of medical data can have devastating consequences for patients, including discrimination, stigmatization, and moral and financial harm.

In addition, there is a risk of widening inequalities in healthcare access. The deployment of AI-based solutions, often developed in high-income contexts, may not adequately consider the needs, languages, and sociocultural realities of vulnerable populations or developing countries. This can accentuate existing disparities in the healthcare system and compromise the principle of distributive justice, which is fundamental to equity in medical care [14].

Also worthy of concern is the potential dehumanization in medicine. The physician-patient relationship, historically centered on listening, care, and empathy, can be negatively affected if trust is shifted to automated systems. Excessive technological dependence can make care more impersonal, compromising the comprehensiveness of care and the subjective dimension of the illness experience [15]. Medical training, in this scenario, must also be rethought, since skills such as critical thinking, clinical judgment, and ethical sensitivity need to be preserved and strengthened, even in an increasingly data-driven environment [16].

The mismatch between the speed of technological innovation and the capacity for ethical and legal regulation, therefore, imposes the urgent need to review bioethical foundations from a new perspective. Such issues require not only technical analysis but a rigorous bioethical approach that considers the impact of AI on the fundamental values that govern medical practice. Recent literature reinforces that isolated ethical

principles are insufficient to guarantee the moral integrity of AI applications: a robust, transparent, and participatory governance system involving multiple stakeholders, including physicians, engineers, patients, legal professionals, and representatives of civil society, is necessary [17,18].

This article aims to review the existing literature on the main ethical dilemmas related to the application of artificial intelligence in medical practice, to recover the Hippocratic principles, and to analyze how they can (or cannot) be preserved amid ongoing technological transformations. Through critical, multidisciplinary analysis, it aims to contribute to the debate on the ethical directions of contemporary medicine in a context of increasing automation.

Method

This article is characterized as a qualitative, exploratory study, with a methodological approach based on a narrative literature review. The objective is to gather, analyze, and critically interpret the scientific literature on emerging ethical dilemmas arising from the application of AI in medical practice, in light of the classical principles of bioethics as established in the Hippocratic Oath.

According to Gil [19], bibliographic research allows the examination of already established theoretical contributions. It is especially useful in studies that aim to understand and discuss topics that are still being consolidated in academia, such as the interface between bioethics and AI. An exploratory approach, in turn, is suitable for the initial deepening of a problem that is still poorly defined, as it allows the identification of categories and gaps relevant to future empirical investigation [20].

The methodological path followed these steps:

1) Definition of the study topic and problem by delimiting the bioethical scope of AI's role in clinical and hospital contexts.

2) Systematic bibliographic survey, carried out in the PubMed, Scopus, SciELO and Web of Science databases, with the following descriptors and combinations: "bioethics," "artificial intelligence," "machine learning," "medical ethics," "Hippocratic Oath," "clinical decision support," "algorithmics bias," among others, with the aid of the Boolean operator "and". Articles published between 2015 and 2024, in Portuguese and English, that directly addressed the ethical aspects of AI in healthcare were included.

3) Selection and screening of sources based on reading titles, abstracts, and subsequently, full texts. The inclusion criteria were: relevance to the topic, publication in peer-reviewed journals, and theoretical or practical contribution to the ethical debate. Duplicate publications, conference abstracts without full text, and articles with a strictly technical or computational focus on AI, without ethical discussion, were excluded.

4) Analysis and synthesis of results through critical reading and thematic categorization, with emphasis on recurring topics such as medical autonomy, responsibility for automated decisions, algorithmic explainability, and equity in access to AI-based technologies.

The methodology adopted aims not only to map the state of the art of bioethical dilemmas in medical AI, but also to contextualize them in light of classic Hippocratic principles—such as beneficence, non-maleficence, and confidentiality—and, thereby, identify points of convergence and tension with our technological paradigms. This approach contributes to the critical maturation of the debate. It provides conceptual support for the formulation of public policies and clinical guidelines, as well as for the development of technologies more sensitive to the ethical principles of medicine.

Discussion

Medical ethics between Hippocrates and the digital age

Since antiquity, medicine has been guided by ethical principles that reflect a moral commitment to the well-being of the patient. The Hippocratic Oath, dating from the 5th century BC, synthesizes this legacy and constitutes a symbolic and normative landmark of Western medical ethics, insofar as it enunciates fundamental commitments such as promoting good, abstaining from causing harm, respecting confidentiality, and the moral conduct of the healthcare professional [3]. These values, which gave rise to the bioethical pillars of beneficence, non-maleficence, autonomy, and

justice, have remained a moral compass throughout the centuries, even in the face of scientific and sociocultural transformations. However, the rise of artificial intelligence represents a new inflection point, as it challenges the applicability and sufficiency of these traditional guidelines to guide medical practices mediated by technologies that make decisions, learn autonomously, and operate on a large scale [2].

This transformation creates ethical challenges, as the algorithmic logic that governs AI systems is not always transparent or compatible with the humanistic values of traditional medicine [4]. By incorporating machine learning algorithms, clinical decision support systems, and predictive technologies based on big data, digital medicine expands diagnostic and therapeutic possibilities while also shifting the centrality of human clinical judgment. In other words, AI can exponentially expand the diagnostic and predictive capacity of medical practice. However, it can also shift decision-making authority from professionals to complex, often opaque, statistical models. In this context, tension arises between the ethical tradition—anchored in interpersonal relationships and prudent judgment—and the new algorithmic rationality, whose logic often becomes incomprehensible, non-auditable, or uncontrollable [2].

This technological transition imposes on traditional bioethics the challenge of reconfiguring its foundations so that it can guide practices in technologized environments, i.e., in which decisions are no longer made exclusively by human subjects, but by hybrid systems in which the physician delegates or shares responsibilities with machines [5,21]. The medicine of the future will thus be, in addition to human practice supported by technology, a co-production between humans and algorithms [6]. This scenario requires a profound ethical reconfiguration that addresses issues of shared responsibility, algorithmic justice, transparency, and data governance. At the same time, Hippocratic ethics continues to offer a fundamental reference, provided it is reinterpreted in the light of new technological mediations. As Floridi and colleagues point out [17], the ethical development of AI must integrate humanistic principles from its conception, aiming to construct a just, transparent, and dignified digital society.

Autonomy and algorithmic opacity

The principle of autonomy is central to modern medical ethics and represents the patient's right to make informed decisions about their own care. The full realization of this principle, grounded in transparency in the physician-patient relationship, depends on clear communication, an understanding of risks, and active patient participation in the therapeutic process. However, the rise of AI, especially in forms of deep learning, threatens to undermine this essential transparency. Complex algorithms, trained with large volumes of data, often operate as "black boxes," i.e., they are unable to explain, in a comprehensible way, how they arrived at a particular recommendation [1,2,4,22].

Such opacity undermines the basis of informed consent, since not even healthcare professionals fully understand the criteria algorithms use. As Morley and collaborators [21] point out, when a physician is unable to explain to a patient how a decision was made—whether it is a diagnosis, risk classification, or therapeutic proposal—the very ethical pact that underlies autonomy is shaken. Instead of making their own decisions, the patient becomes dependent on a system whose logic is inaccessible to them.

Furthermore, algorithmic opacity introduces new layers of informational asymmetry. Patients with lower levels of education, limited digital access, or low health literacy are even more vulnerable to a loss of autonomy, as they face additional barriers to understanding decision-making processes. As argued by the authors, autonomy is not only an individual right but a relational phenomenon that requires appropriate contexts for its exercise—among them, access to clear and intelligible information [6].

The issue is exacerbated when considering the ambiguity surrounding the ethical and legal responsibility for AI-aided decisions. When a clinical decision is made based on algorithmic recommendations, who should be held accountable for its results? The physician? The algorithm developer? The institution [5]? The lack of clarity regarding moral accountability—what some authors call an "ethical vacuum" [23]—compromises not only autonomy, but the very legitimacy of care.

In this scenario, it becomes urgent to move toward more explainable and auditable systems [24].

Algorithmic transparency, coupled with the development of participatory consent and mediation mechanisms, is essential to preserving patient autonomy in an era dominated by automated decisions.

Beneficence and risk of algorithmic bias

As a moral obligation to act in the patient's best interest, beneficence takes on new dimensions in medicine mediated by artificial intelligence. AI tools have demonstrated, in several studies, accuracy comparable to or superior to that of humans in tasks such as interpreting radiological images for skin cancer classification [1] and predicting cardiovascular events [22]. However, this technical superiority does not imply ethical neutrality. As some authors warn, the performance of algorithms depends heavily on the quality, diversity, and representativeness of the data used in their training [21].

Several studies indicate that systems trained with historically biased data tend to replicate and even amplify pre-existing inequalities. An emblematic example is the population health management algorithm that underestimated the severity of the condition of black patients, as it used the cost of health care as a proxy [14], an indicator influenced by structural inequality. The uncritical application of these systems can therefore violate the principle of beneficence by producing discriminatory and ineffective outcomes for certain groups.

Furthermore, the risk of algorithmic bias is not limited to racial or socioeconomic issues. Even systems developed for automated recruitment have shown gender bias, penalizing female candidates [25], and have also demonstrated inferior performance in Black populations, older people, or patients with multiple comorbidities, reflecting biases in the representativeness of training data. When this type of logic is transferred to the field of health, the potential for ethical harm increases, as clinical decisions can determine life-or-death outcomes. Instead of promoting good, AI can inadvertently reinforce discriminatory patterns and hinder equitable access to quality care—thereby violating the principle of beneficence and undermining the Hippocratic ideal of treating all patients equally.

In this context, beneficence requires not only technical effectiveness but also constant vigilance over the equity of AI systems. It is necessary to develop algorithmic audit methodologies capable of identifying discriminatory patterns, correcting biases, and ensuring that the benefits of AI are distributed fairly [6]. Ethical evaluation should be part of the system's life cycle itself, from design to clinical implementation [25].

Non-maleficence and invisible harm

Non-maleficence, an ethical principle expressed by the maxim "*primum non nocere*"—that is, "first, do no harm"—requires professionals to avoid or minimize harm to the patient's health and well-being during medical care. In the context of AI, however, this understanding should be broadened to encompass unintended risks, since harm is not always visible, immediate, or direct, and this does not minimize the potential for significant harm to patients' health, integrity, or dignity. The use of automated systems can lead to unintended clinical, psychological, and social consequences, often imperceptible at the beginning, but with profound effects in the medium and long term [21,23].

Among the most discussed risks are diagnostic errors caused by system malfunction, inadequate interpretation, or out-of-scope use. Even though AI can offer superior accuracy in certain contexts, it is as fallible as the data and parameters that inform it. Automated clinical decisions can ignore patient singularities and lead to inappropriate conduct in complex or atypical situations [26]. This can be even more problematic when professionals overconfidently rely on algorithmic recommendations, a phenomenon known as automation bias.

Another relevant ethical risk is the dehumanization of care, since, with the advancement of technology, there is a tendency to value what is quantifiable and predictable to the detriment of subjectivity [27]. The replacement of clinical listening with interaction with digital interfaces can reduce the space for the expression of the patient's suffering, fear, and vulnerability, resulting in the dehumanization of care and neglect of subjective and affective aspects essential to the experience of illness [27]. This affective emptying, although difficult to measure, constitutes a type of invisible damage—emotional

and existential—that contradicts the integrality of Hippocratic care.

In addition, the use of AI in health can generate adverse collective effects. Security failures, data leaks, or biased systemic decisions undermine society's trust in health systems. Non-maleficence, here, must go beyond the individual dimension and consider the systemic and institutional ethical impacts of automation.

Incorporating robustness, traceability, and continuous auditing measures to mitigate these risks is essential [23]. Transparency, reviewability, and the possibility of "human override" (the ability to intervene or replace AI decisions with a human being) are indispensable ethical components for medical AI to operate under the principle of non-maleficence [24]. The technology must be tested not only for its efficiency, but also for its ethical resilience.

Justice, equity, and digital access to AI in health

The principle of justice, understood as the obligation to treat everyone equitably and to promote fair access to health resources, acquires new layers of complexity in the age of artificial intelligence. If, on the one hand, AI promises to expand the reach of medical care through telemedicine, remote screening, and automated diagnoses, on the other hand, these technological advances often encounter structural barriers that deepen existing inequalities [21,28,29].

Literature shows that low- and middle-income countries, as well as peripheral regions and vulnerable communities, often lack the minimum infrastructure needed to implement AI-based solutions: connectivity, system interoperability, professional training, equipment, and public policies geared toward digital inclusion [29,30]. The absence of these conditions compromises not only the effectiveness of the technologies but also the equity in the distribution of their benefits.

Social justice should be a guiding principle in the development of AI technologies, to avoid the creation of a "two-speed" medicine: one of high performance and precision, for those who can pay; another, limited and conventional, for those who depend on the public system [17]. The commodification of AI in healthcare can widen the gap between patients from different social classes and consolidate what has been called "automated injustice" [31].

This risk requires a regulatory approach focused on distributive justice. Global ethical guidelines on AI converge on the need for policies that promote inclusive access, representative data, and the participation of underrepresented populations in the design and validation of systems [5]. In Brazil, for example, the ethical implementation of AI depends on investments in public digital health, incentives for interoperability, and professional training [28].

Therefore, justice in the age of AI is not just a matter of resource distribution, but also of epistemological inclusion. Historically marginalized groups must be incorporated into the technological development process, from data production to the formulation of validation criteria. Only in this way will it be possible to build a truly plural artificial intelligence committed to the Hippocratic values of universal medicine.

Confidentiality, LGPD, and the new ethics of privacy

Medical confidentiality is one of the oldest pillars of professional ethics in healthcare. Since the Hippocratic Oath, preserving the secrecy of information obtained during care has been considered a moral and legal duty. However, AI, by operating with large volumes of sensitive data, poses unprecedented challenges to this principle [6,21].

AI systems depend on extensive clinical data for training, validation, and real-time operation. Often, there is highly identifiable information, such as medical images, genomes, medical records, and personal metadata. Even anonymization techniques can be ineffective, as algorithms can re-identify individuals by cross-referencing data [15,30]. The massive collection of this data, often in the background (via apps, wearables, online portals), challenges the ability to obtain informed consent and the ethical governance of these information banks.

In Brazil, the LGPD [8], which establishes guidelines for the processing of personal data, requires a clear legal basis, data minimization, revocation mechanisms, the right to portability, deletion, and

informed consent, especially in relation to so-called "sensitive data," such as health data.

The regulation of telehealth in Brazil, through Law 14,510/2022 [7], reinforces the obligation to preserve the integrity, confidentiality, and availability of digital information in remote care. CFM Resolution 2,314/2022 [9] establishes that the physician must guarantee, through appropriate technological means, the protection of patient data and inform them about the risks of virtual care. Thus, these regulations converge to build a secure, ethical, and patient-centered digital ecosystem when medical services are provided in Brazilian territory.

In the international context, HIPAA [12] and HITECH [13] stand out in the United States, which define standards for the security, integrity, and auditing of digital clinical data; and the GDPR [10] and NIS Directive [11] in Europe, which establish the principle of privacy by design (a data protection principle according to which privacy should be incorporated from the initial stages of the development of systems, products, services or processes) and require accountability from data controllers and cybersecurity in critical sectors such as health.

These laws indicate that contemporary medical ethics must incorporate a bioethics of data: a set of ethical and technical responsibilities aimed at protecting the privacy and informational self-determination of patients. We live in an era of "surveillance capitalism," in which personal data is continuously collected and traded, so that the protection of confidentiality is not only a technical issue, but also a political and cultural one [32].

Moral responsibility and accountability in hybrid systems

With the introduction of artificial intelligence into clinical practice, the traditional notion of moral responsibility—centered on the figure of the physician as an autonomous decision-maker—is undergoing profound transformation. In environments where clinical decisions are shared between humans and algorithms, the line between causality, responsibility, and guilt becomes blurred [4,6].

This ambiguity is known in the literature as the "responsibility vacuum" [23] and is characterized by the difficulty of assigning responsibility in cases of AI-assisted medical error. For example, if an algorithm recommends a course of action based on statistical evidence and the physician follows it, causing harm to the patient, who is responsible?

The literature proposes the concept of "shared responsibility," which recognizes the multiplicity of agents involved, from programmers and data scientists to hospital and medical factors [17]. However, simply recognizing this complexity does not resolve the ethical dilemma, as it can dilute the guilt effect, weakening institutional accountability mechanisms [33].

In response to this challenge, authors advocate developing ethical and legal norms that hold all parties accountable, with an emphasis on the auditability of systems and the traceability of decisions. This implies, for example, that every AI system used in healthcare must maintain accessible records of the data used, the models applied, the recommendations issued, the contexts in which they were issued, and by whom [5].

In clinical practice, physicians must maintain critical autonomy in the face of automated systems and use them as support tools, not as substitutes for clinical judgment. AI should be an "augmented intelligence" [2], not an autonomous authority. Resuming the centrality of the physician as the guardian of clinical judgment, even in advanced technological contexts, can be a way to rebalance the Hippocratic principle of prudence with innovation. In this sense, preserving the ethical centrality of the physician in the decision-making process is a means of protecting both patient safety and the integrity of the medical profession.

Crisis of the physician-patient relationship in the algorithmic age

Traditional medicine has always been built on a complex interpersonal relationship, marked by active listening, empathy, trust, and humanized clinical judgment. However, the increasing mediation of medical practice by digital systems—from automated screening by chatbots (programs developed to simulate conversation with human beings through natural language) to clinical decision-making algorithms—threatens to reduce the subjective and ethical dimensions of the interaction between physician and patient [15,21].

Some authors point out that, by prioritizing efficiency, standardization, and objectivity, AI-based systems tend to marginalize humanistic aspects of care, such as understanding the patient's suffering, pain, or expectations [27]. The Hippocratic model presupposes a physician who knows the patient in their entirety—biological, emotional, and social – and who, therefore, can adapt conduct ethically and prudently. However, the excessive use of technology can lead to the so-called "algorithmic depersonalization" of care, in which the patient's subjectivity is reduced to numerical or probabilistic variables, with no room for personal narratives or singularities [28]. In addition, trust, which is an essential component of the therapeutic alliance, can be transferred to technology—a phenomenon known as trust displacement—which weakens the bond with the professional, impairs treatment adherence, and creates unfounded expectations of technical infallibility [34].

Additionally, the use of AI in early stages of care, such as automated screening by chatbots, can reduce human contact at crucial moments for the patient, generating insecurity, loneliness, or a feeling of abandonment. Technology tends to devalue listening and clinical intuition—abilities that, although not quantifiable, are essential to recognizing human suffering in its complexity [27].

This scenario challenges the physician's very identity. If before he was seen as a caregiver and advisor, today he risks being reduced to a technical systems operator, losing ground to clinical reasoning and the ethics of the encounter. It is important to reclaim the notion that medicine must be, above all, a commitment to the well-being of the other [3]—something that cannot be delegated to machines, however efficient they may be.

To face this crisis, it becomes urgent to promote medical training focused on digital ethics, critical thinking about AI, and the development of communication skills for technology-mediated interactions [16]. Restoring the centrality of the human relationship in care is an ethical imperative that no innovation can replace.

Reinterpretation of Hippocratic principles in the 21st century

The analysis of the reviewed studies reveals that the ethical principles enunciated by Hippocrates remain a valuable moral basis for contemporary medicine but require reinterpretation in light of the challenges posed by artificial intelligence. This reinterpretation can be summarized in four movements:

1) Autonomy 2.0: patient autonomy, in the context of AI, is not limited to the choice of treatments and includes the right to explanation, algorithmic non-discrimination, and protection of sensitive data. This gives rise to the need for expanded informed consent, capable of addressing technological risks and the functioning of automated systems [24].

2) Inclusive beneficence: doing good requires ensuring that the benefits of technologies are distributed equitably. This implies monitoring data biases, prioritizing the ethical development of algorithms, and preventing innovations from increasing social inequalities [14].

3) Systemic non-maleficence: preventing harm means not only avoiding clinical errors, but also anticipating systemic consequences of automation, such as digital exclusion, loss of medical skills, and the distortion of care based on Hippocratic principles [23].

4) Digital justice: justice in the age of AI requires that all those involved have access to safe, effective, and ethically designed technologies. Algorithmic governance must consider cultural pluralism, diversity of contexts, and respect for patients' fundamental rights [17].

Thus, the Hippocratic legacy is not discarded, but updated, functioning as a living ethical matrix that can guide moral decisions in complex techno-scientific environments.

Bioethical proposals for patient-centered AI

Based on the reviewed literature, some bioethical guidelines for the development and application of AI in medicine can be outlined:

1) Transparency and algorithmic explainability: systems must be auditable and understandable, to allow the physician and patient to understand the criteria for the suggested decisions [4,24].

2) Dynamic informed consent: mechanisms for continuous and contextualized consent should be developed in a way that adapts to the interactive nature of digital technologies [35].

3) Inclusive participation: AI development should include multiple actors, with special attention to the voices of end users—physicians and patients—to ensure alignment with clinical values [25,35].

4) Ethical governance: AI implementation should be accompanied by multidisciplinary ethics committees that review its impacts in real time, to promote accountability adjustments [18,36].

5) Medical education for the digital age: medical curricula should include training in technology ethics, data literacy, and critical thinking about AI, to avoid uncritical dependence or irresponsible delegation [16].

These proposals, although still under construction, point the way to a more ethical, transparent, and human dignity-centered algorithmic medicine.

Limitations of the current literature and gaps for future studies

Despite the growing scientific literature on AI ethics in medicine, the field presents significant gaps. First, there is a predominance of theoretical and normative studies, with a scarcity of empirical data on physicians' and patients' experiences in real-world contexts of AI use. The absence of qualitative studies limits the understanding of how ethical dilemmas manifest themselves in everyday practice [21].

Second, many studies have a Euro-American focus and therefore disregard digital inequalities and the cultural, legal, and social specificities of other contexts, such as those in Latin America and Africa. Studies on algorithmic justice need to incorporate post-colonial and global health perspectives, at the risk of repeating technocentric and exclusionary patterns [37] and reproducing historical exclusions with a technological veneer.

Furthermore, there is a lack of integration between legal, clinical, and ethical frameworks. Few models offer practical guidance that articulates the requirements of the LGPD, medical ethics, and clinical efficiency in a single framework. In this sense, there is an urgent demand for interdisciplinary research that engages with law, sociology, computer science, and moral physiology, and engagement with patients and the expansion of ethical deliberation forums can be promising paths, without losing innovation.

The ethical advancement of AI in medicine requires not only technical innovation but also public and institutional commitment to human dignity. Preserving the spirit of the Hippocratic Oath while adapting it to digital times is the greatest challenge—and the greatest opportunity—of contemporary bioethics.

Final considerations

Artificial intelligence represents one of the most significant transformations in medical practice in the 21st century. While it offers powerful resources for personalized diagnoses, prognoses, and therapies, its large-scale adoption poses profound ethical challenges that reconfigure the traditional foundations of medicine. These transformations demand not only technical updates for healthcare professionals but also critical reflection on the values that should guide their use.

A critical analysis of the literature reveals that the Hippocratic principles of beneficence, non-maleficence, autonomy, and justice remain fundamental but need to be reinterpreted in light of the risks and potential of the digital age. Autonomy, for example, can no longer be guaranteed solely by traditional informed consent: it requires algorithmic explainability and transparency of automated decisions. Beneficence demands constant vigilance against biases in data and machine learning systems, which can compromise the quality and equity of care. Non-maleficence requires a systemic, interdisciplinary approach that accounts for the unintended effects of algorithms in clinical practice. Furthermore, justice implies a commitment to democratizing access to technologies to avoid the concentration of resources in privileged contexts.

In addition, national and international legal frameworks provide an important normative basis but remain insufficient to address the ethical complexities introduced by AI systems in healthcare. While they provide relevant guidelines for data protection and information security, many of these regulations still do not fully address the challenges of

algorithmic autonomy, responsibility in hybrid (human-machine) systems, and transparency in deep learning models, which are often considered "black boxes."

In this context, contemporary bioethics needs to evolve to incorporate, in a truly interdisciplinary way, the concepts of algorithmic governance, protection of sensitive data, distributive justice, and moral accountability. This involves the active participation of healthcare professionals, jurists, engineers, philosophers, and representatives of civil society in the formulation of public policies and clinical guidelines. The ethical development of AI in health cannot be the sole responsibility of programmers or technology companies: it must be a joint effort, guided by human well-being.

Preserving the spirit of the Hippocratic Oath does not mean resisting technological progress, but rather ensuring that innovation serves human dignity, equity, and compassionate care. It is up to medicine, ethics, and law to jointly build a model of artificial intelligence that, rather than merely learning from data, respects the fundamental values of life and care. Ultimately, the true advancement of digital medicine will depend not only on the ability to predict diseases but also on promoting humanity in every automated decision.

References

1. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM *et al*. Dermatologist-level classification of skin cancer with deep neural networks. Nature [Internet]. 2017 [acesso 5 jun 2025];542(7639):115-8. DOI: 10.1038/nature21056
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med [Internet]. 2019 [acesso 5 jun 2025];25(1):44-56. DOI: 10.1038/s41591-018-0300-7
3. Pellegrino ED, Thomasma DC. For the patient's good: the restoration in health care. New York: Oxford University Press; 1998.
4. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep [Internet]. 2019 [acesso 5 jun 2025];49(1):15-21. DOI: 10.1002/hast.973
5. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell [Internet]. 2019 [acesso 5 jun 2025];1(9):389-99. DOI: 10.1038/s42256-019-0088-2
6. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial Intelligence in Healthcare. 2020:295-336. DOI: 10.1016/B978-0-12-818438-7.00012-5
7. Brasil. Lei nº 14.510, de 27 de dezembro de 2022. Dispõe sobre a prestação de serviços de telessaúde. Diário Oficial da União [Internet]. Brasília, 28 dez 2022 [acesso 6 jun 2025]. Disponível: https://tinyurl.com/mr39srpb
8. Brasil. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União [Internet]. Brasília, 15 ago 2018 [acesso 6 jun 2025]. Disponível: https://bit.ly/4b4cssd
9. Conselho Federal de Medicina. Resolução nº 2.314, de 20 de abril de 2022. Define e regulamenta a telemedicina como forma de prestação de serviços médicos. Diário Oficial da União [Internet]. Brasília, p. 227, 5 maio 2022 [acesso 6 jun 2025]. Seção 1. Disponível: https://tinyurl.com/hyrzwr3f
10. European Parliament and Council. Regulation (EU) 2016/679 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). Official Journal of the European Union [Internet]. Geneva, 2016 [acesso 5 jun 2025]. Disponível: https://bit.ly/409B7qj
11. European Commission. Directive (EU) 2016/1148 on Security of Network and INformation Systems (NIS Directive). Official Journal of the European Union [Internet]. 2016 [acesso 6 jun 2025]. Disponível: https://bit.ly/4aQwtUa

12. United States. Public Law 104-191. Health Insurance Portability and Accountability Act of 1996 (HIPAA). Federal Register [Internet]. Washington, 20 ago 1996 [acesso 6 jun 2025]. Disponível: https://tinyurl.com/vvkr9vm8
13. United States. Health Information Technology for Economic and Clinical Health Act (HITECH). Federal Register [Internet]. Washington, n° 209, Rules and Regulations, 30 out 2009 [acesso 10 set 2024]. Disponível: https://bit.ly/4d4yRYR
14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science [Internet]. 2019 [acesso 6 jun 2025];366(6464):447-53. DOI: 10.1126/science.aax2342
15. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health [Internet]. 2013 [acesso 8 jun 2025];11(3):256-70. DOI: 10.1057/sth.2013.10
16. Masters K. Artificial intelligence in medical education. Med Teach [Internet]. 2019 [acesso 14 jun 2025];41(9):976-80. DOI: 10.1080/0142159X.2019.1595557
17. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V *et al*. AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach [Internet]. 2018 [acesso 6 jun 2025];28(4):689-707. DOI: 10.1007/s11023-018-9482-5
18. Mittelstadt BD. Principles alone cannot guarantee ethical AI. Nat Mach Intell [Internet]. 2019 [acesso 6 jun 2025];1(11):501-7. DOI: 10.1038/s42256-019-0114-4
19. Gil AC. Métodos e técnicas de pesquisa social. 6ª ed. São Paulo: Atlas; 2008.
20. Severino AJ. Metodologia do trabalho científico. 24ª ed. São Paulo: Cortez; 2016.
21. Morley J, Machado C, Burr C, Cowls J, Taddeo M, Floridi L. The ethics od AI in health care: a mapping review. Soc Sci Med [Internet]. 2020 [acesso 5 jun 2025];260:113172. DOI: 10.1016/j.socscimed.2020.113172
22. Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Narayan SM. Deep learning for cardiovascular medicine: a practical primer. Eur Heart J [Internet]. 2019 [acesso 6 jun 2025];40(25):2058-73. DOI: 10.1093/eurheartj/ehz056
23. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics [Internet]. 2016 [acesso 6 jun 2025];22(2):303-41. DOI: 10.1007/s11948-015-9652-2
24. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. Sci Robot [Internet]. 2017 [acesso 8 jun 2025];2(6):eaan6080. DOI: 10.1126/scirobotics.aan6080
25. Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters [Internet]. 10 out 2018 [acesso 2 jul 2025]; Disponível: https://bit.ly/4rhewTv
26. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA [Internet]. 2018 [acesso 6 jun 2025];320(21):2199-200. DOI: 10.1001/jama.2018.17163
27. Harari YN. Homo Deus: a brief history of tomorrow. New York: Harper; 2017.
28. Silva AG, Oliveira TC, Lima MADS. Artificial intelligence in Brazilian health care: a viewpoint. JMIR Med Inform [Internet]. 2020 [acesso 7 jun 2025];8(10):e23557. DOI: 10.2196/23557
29. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health [Internet]. 2020 [acesso 7 jun 2025];2(8):e435-40. DOI: 10.1016/S2589-7500(20)30142-4
30. Goodman B. The algorithmic governance of health: a case study of the United States. Health Care Anal [Internet]. 2021 [acesso 8 jun 2025];29(1):39-57. DOI: 10.1007/s10728-019-00385-0
31. Eubanks V. Automating inequality: how high-tech tools profile, police, and punish the poor. New York: Picador; 2019.
32. Zuboff S. The age of surveillance capitalism. New York: PublicAffairs; 2019.
33. Nemitz P. Constitutional democracy and technology in the age of artificial intelligence. Philos Trans A Math Phys Eng Sci [Internet]. 2018 [acesso 7 jun 2025];376(2133):20180089. DOI: 10.1098/rsta.2018.0089

34. Zhang B, Dafoe A. Artificial intelligence: American attitudes and trends [Internet]. Oxford: University of Oxford; 2019 [acesso 8 jun 2025]. DOI: 10.2139/ssrn.3312874

35. Sadan T. Toward participatory design of ethical AI. Philos Technol [Internet]. 2020 [acesso 8 jun 2025];33(2):167-88. DOI: 10.1007/s13347-019-00371-1

36. Dove ES, Townend D, Meslin EM, Bobrow M, Littler K, Nicol D *et al*. Ethics review for international data-intensive research. Science [Internet]. 2016 [acesso 8 jun 2025];351(6280):1399-400. DOI: 10.1126/science.aad5269

37. Benjamin R. Race after technology: abolitionist tools for the New Jim Code [Internet]. Cambridge: Polity Press; 2019 [acesso 8 jun 2025]. Disponível: https://tinyurl.com/272s7b8z

Juracy Barbosa dos Santos – PhD Student – drjuracybarbosa@gmail.com
0009-0009-1729-9778

Guilhermina Rego – PhD – guilherminarego@med.up.pt
0000-0002-8590-9832

Rui Nunes – PhD – ruinunes@med.up.pt
0000-0002-1377-9899

Correspondence
Juracy Barbosa dos Santos – Faculdade de Medicina/Universidade do Porto. Al. Professor Hernani Monteiro, CP 4200-319. Porto, Portugal.

Contribution of the authors
Juracy Barbosa dos Santos participated in the study conception, data analysis and manuscript writing. Guilhermina Rego participated in the manuscript review. Rui Nunes participated in the critical analysis and final manuscript review.

Data availability: All data used or generated in the research are fully described and presented in the body of the article.

Editor in charge: Dilza Teresinha Ambrós Ribeiro

Received: 8.24.2025
Revised: 12.18.2025
Approved: 1.27.2026