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

Rev. Bioét. vol.34 Brasília 2026

Ethical monitoring in intelligent health support systems

Ricardo Rheingantz Abuchaim [1,2], *Daniel Brito de Araujo* [2]

1. Advocacia-Geral da União, Pelotas/RS, Brasil. **2.** Universidade Federal de Pelotas, Pelotas/RS, Brasil.

Abstract

This article aims to propose ethical guidelines and an evaluation model for the use of artificial intelligence in medical diagnoses. A critical analysis of bioethical and normative literature, based on the principles of autonomy, beneficence, non-maleficence, and justice was conducted. From this analysis, a hierarchy of the main ethical challenges involved is developed, ranging from data quality and governance to the ethical sustainability of the technologies employed. An evaluation model based on objective metrics is also proposed, aimed at the continuous monitoring of the application of artificial intelligence in clinical practice. It concludes that the balance between technological innovation and humanization of care is essential, with medical associations playing a strategic regulatory role in preserving the fundamental values of medicine.

Keywords: Artificial intelligence. Ethics. Ethics, medical. Intelligent systems.

Resumo

Supervisão ética em sistemas inteligentes de apoio à saúde

O artigo tem como objetivo propor diretrizes éticas e um modelo de avaliação para o uso de inteligência artificial em diagnósticos médicos. Realiza-se análise crítica da literatura bioética e normativa, fundamentada nos princípios da autonomia, beneficência, não maleficência e justiça. Com base nessa análise, desenvolve-se uma hierarquia dos principais desafios éticos envolvidos, que abrange desde a qualidade e a governança dos dados até a sustentabilidade ética das tecnologias empregadas. Propõe-se, ainda, um modelo de avaliação baseado em métricas objetivas, voltado ao monitoramento contínuo da aplicação da inteligência artificial na prática clínica. Conclui-se que o equilíbrio entre inovação tecnológica e humanização do cuidado é essencial, cabendo às associações médicas um papel regulador estratégico na preservação dos valores fundamentais da medicina.

Palavras-chave: Inteligência artificial. Ética. Ética médica. Sistemas inteligentes.

Resumen

Supervisión ética en sistemas inteligentes de apoyo a la salud

El artículo tiene como objetivo proponer directrices éticas y un modelo de evaluación para el uso de la inteligencia artificial en diagnósticos médicos. Un análisis crítico de la literatura bioética y normativa, basado en los principios de autonomía, beneficencia, no maleficencia y justicia fue realizado. A partir de este análisis, se desarrolla una jerarquía de los principales retos éticos implicados, que abarca desde la calidad y la gobernanza de los datos hasta la sostenibilidad ética de las tecnologías empleadas. Además, se propone un modelo de evaluación basado en métricas objetivas, orientado al seguimiento continuo de la aplicación de la inteligencia artificial en la práctica clínica. Se concluye que el equilibrio entre la innovación tecnológica y la humanización de la atención es esencial, y que las asociaciones médicas tienen un papel regulador estratégico en la preservación de los valores fundamentales de la medicina.

Palabras clave: Inteligencia artificial. Ética. Ética médica. Sistemas inteligentes.

Update

The authors declare no conflict of interest.

Artificial intelligence (AI) is revolutionizing the healthcare sector with its disruptive capacity in medical diagnostics. Its implementation provides significant advances in diagnostic accuracy, cost optimization, and improvement of clinical outcomes [1]. Through advanced algorithms and techniques leveraging machine learning, the systems process significant volumes of medical data, including image analysis and laboratory result interpretation, establishing new paradigms in diagnostic medicine.

The technological landscape of AI in healthcare demands a robust governance framework to ensure the safety and effectiveness of the systems in a clinical setting. The algorithm architecture must include rigorous validation protocols to mitigate risks of inaccurate diagnoses and consequent adverse impacts on treatment [2]. Information security management (ISM) has a critical role, given the sensitivity of the health data processed, with emphasis on mitigating algorithmic biases that may perpetuate healthcare disparities between different demographic groups [3]. The introduction of continuous model auditing and validation methodologies is essential to ensure equity in AI-based diagnostics.

The informed consent paradigm requires adaptation to the digital context to ensure transparency in the functioning of algorithms and their role in the decision-making process [4]. The technical documentation must explicitly state the systems' capabilities and limitations, to provide adequate understanding by patients and healthcare professionals.

AI system governance demands a clear definition of tasks and mechanisms for liability [5]. The regulatory framework must be consistent with technological evolution, with guidelines for the development, validation, and continuous monitoring of solutions. The digital transformation of healthcare arising from AI requires consideration of equity in access to technological innovations. Specific policies must consider potential socioeconomic impacts, as well as aspects related to professional capacity-building and labor market restructuring in the sector [6].

This study examines the role of medical professionals and Brazil's Federal Council of Medicine (CFM) in the governance of AI in diagnostics and proposes guidelines (hierarchy and metrics) for ethical and safe integration of the technology. The analysis considers technical and regulatory aspects, with a view to establishing a model that maximizes the benefits of technological innovation while preserving fundamental principles of medical ethics and patient autonomy. The synergy between AI and human expertise represents a new paradigm in medicine, which demands a robust governance framework to ensure responsible and patient-centered development [7]. The success of this transformation depends on the alignment between technological innovation, ethical principles, and adequate regulation.

This work aims to: 1) propose ethical guidelines for the implementation of AI systems in medical diagnostics, based on bioethical principles; 2) develop a hierarchy of ethical challenges to guide the prioritization of initiatives; and 3) establish objective metrics for assessing the ethical effectiveness of these systems. The analysis aims to define the role of medical professionals and the CFM in the governance of these systems and, thus, support the safe and responsible integration of the technology into clinical practice.

It is a conceptual analysis and critical review of the scientific and normative literature on AI applied to healthcare. The study is based on the classic bioethical principles (autonomy, beneficence, non-maleficence, and justice), current regulatory frameworks, and recent bibliographic references on AI ethics. The analysis provides the basis of a proposal for a hierarchy of ethical challenges and metrics for systematic assessment of the responsible implementation of intelligent systems in medical diagnostics.

Algorithmic biases and algorithmic transparency

It is essential to clarify the concepts that underpin the ethical concerns associated with AI systems. Algorithmic biases manifest when AI systems produce unfair judgments or decisions due to discriminatory characteristics present in the training data or in the algorithm structure [4].

Algorithmic discrimination occurs when the system is influenced by characteristics that are effectively irrelevant to the issue under analysis, typically related to biases about members of certain groups. A critical aspect of algorithmic biases is that they may not only reflect but also amplify and automate historical discrimination present in the data used to train the systems, thus perpetuating cycles of inequality through computational processing [3]. This issue is especially relevant in decision-making support systems that impact individual rights and opportunities.

In turn, algorithmic transparency refers to the possibility of understanding how an AI system reaches its conclusions or outputs. This concept is intrinsically related to the ability to explain the system's internal workings and its decisions [4]. A fundamental challenge posed by algorithmic transparency lies in the fact that, especially in systems that use machine learning and neural networks, it is often impossible for experts to identify exactly the patterns extracted from data and how they are used by the system to produce results [8]. "Algorithmic opacity" becomes even more problematic when combined with the presence of biases, as it hinders the identification and correction of potential discriminations present in the system, precluding the physician from effectively participating in the decision-making process.

Ethical principles in health diagnostics by intelligent systems

It is undeniable that AI alters medical diagnosis activity quantitatively and qualitatively; however, the use of intelligent tools must always respect the patient's fundamental right to make decisions about their own health [4]. To guarantee this right, it is essential that the patient understands the process and actively participates in it. When an AI system is used to assist in diagnosis, the patient needs to be informed clearly and simply: this means explaining not only how the technology works, but also its limitations and potential failures and biases [9]. It is important to emphasize that AI does not replace the physician; it is a support tool, so the final decision must always be made by the medical science professional in dialogue with the patient. The system only provides additional information to assist in this process [9].

Regarding personal information, the patient must maintain full control over their data, which means they can authorize or not the use of this information, check which data is being used, and, if necessary, request corrections or even complete deletion of the records [4].

The ethical principle of beneficence, which prioritizes maximizing therapeutic benefit, constitutes one of the pillars in the application of AI in medical diagnostics. The architecture of AI systems enhances diagnostic accuracy thanks to the processing of expressive volumes of clinical data with minimization of errors and optimization of healthcare outcomes [1]. In turn, the principle of non-maleficence, often stated as "do no harm," demands a robust framework of technical and procedural safeguards; therefore, it requires a rigorous system validation methodology, including extensive testing across diverse populations for mitigation of algorithmic biases and prevention of adverse outcomes [2].

The implementation of AI in diagnostic medicine establishes an innovative paradigm founded on the principle of maximization of therapeutic benefits for the patient. The most significant contributions of this technology notably include its ability for early detection of pathological conditions through advanced clinical data analysis systems—these systems enable the identification of subtle physiopathological changes in the initial stages, prior to the evident manifestation of clinical symptoms [1]. The predictive capability provides healthcare professionals with the opportunity to institute therapeutic interventions in the preliminary stages of pathological development, which results in significantly higher rates of therapeutic success and substantial improvements in patient quality of life. It is also relevant to note that the immediate availability of information based on scientific evidence and support for clinical decision-making contribute to the improvement of professional skills and, thus, to the improvement of the qualitative standards of the healthcare provided.

AI also provides a new dimension to personalized medicine. Through systematic

and meticulous analysis of individualized data, advanced computational systems enable the development of specific therapeutic protocols, based on the genetic characteristics, clinical history, and behavioral determinants of each patient, establishing a model of individualized therapeutic precision [1].

A distinctive characteristic of AI systems lies in their capacity for progressive learning. The continuous processing of clinical data enables systematic refinement of algorithms and results in increasing enhancement of operational accuracy and efficiency. The iterative process establishes a positive feedback mechanism, in which the progressive use of the system results in successive increments in its patient support capacity.

Optimized healthcare resource management constitutes another significant benefit resulting from the installation of AI in healthcare systems. Efficient distribution of equipment and professionals, mediated by advanced algorithms, provides a substantial expansion of access to qualified healthcare services, with criteria-based allocation of available resources, according to traced healthcare demands [10].

In all applications, technological development maintains the promotion of patient well-being as its guiding principle. Every implemented innovation should be oriented toward the pursuit of optimized clinical outcomes, minimized suffering, and comprehensive promotion of health. Compliance with the principle of beneficence ensures that the development of AI in medicine preserves its primary objective: to promote health and improve the quality of life of the population, being established as a technological tool in the service of progress in healthcare.

Regarding the principle of justice in the integration of AI into medical diagnostic systems, it establishes fundamental guidelines for the equitable distribution of technological resources and benefits [6]. The effective implementation of these systems requires a comprehensive methodological framework that transcends traditional socioeconomic and geographical barriers, in order to ensure the effective democratization of access to advanced diagnostic resources in all spheres of society.

The development and enhancement of AI systems demand meticulous consideration of the mitigation of algorithmic biases. The architecture of computational models must be based on judiciously selected databases, which adequately provide for population diversity in its multiple demographic, ethnic, and socioeconomic dimensions [4]. Representativeness constitutes an essential element for ensuring uniform diagnostic accuracy across diverse population segments. The data collection and processing methodology represents a critical component in the development process. Ethical implementation requires systematic incorporation of minority groups into algorithm training databases, with rigorous procedures that prevent systematic discrimination and ensure equity in diagnostic results for all population segments [3].

Economic viability arises as a fundamental aspect in democratizing access to AI-based diagnostic systems. The structuring of the implementation model demands balanced pricing policies and subsidy mechanisms that ensure universal access to diagnostic resources, regardless of the socioeconomic conditions of system users [6].

In the international context, these technologies present significant potential to reduce disparities in healthcare, particularly through the effective implementation of remote diagnostic resources. The execution process requires careful consideration of regional specificities and particular demands of areas traditionally underserved by conventional healthcare systems [1].

Equitable distribution of technological benefits demands a methodologically structured and comprehensive implementation strategy. The strategic planning must prioritize areas with high potential for positive social impact, rather than focusing exclusively on centers with high offer of healthcare services, ensuring that optimized diagnostic accuracy and response times benefit the entire population [4].

The development process needs to incorporate effective community participation mechanisms, integrating diverse cultural and social perspectives. Participatory engagement ensures alignment between the implemented technological functionalities and the specific needs of the communities served, promoting a

culturally appropriate and socially responsible implementation [6].

The normative framework must establish fundamental guarantees of non-discriminatory healthcare, preservation of privacy, and universal access to healthcare services. Strict compliance with the principle of justice enables AI systems to act as effective instruments in the promotion of healthcare equity and to significantly contribute toward the mitigation of current disparities in the global public healthcare conjuncture.

Data governance in healthcare

Data governance in healthcare within the context of diagnostic AI systems represents a critical domain that demands a comprehensive and rigorous approach to ensure both clinical effectiveness and the protection of individual rights [4]. The issue can be analyzed according to different fundamental dimensions that are interrelated. Healthcare data privacy is an essential pillar of governance, given the highly sensitive nature of medical information. Diagnostic AI systems process significant volumes of clinical data, which contain medical histories, test results, diagnostic images, and genetic information. The protection of this information requires the use of robust anonymization and pseudonymization protocols, in order to ensure that data used to train and operate AI systems do not allow individual patient identification [6].

Information security in diagnostic AI systems demands a technological architecture that includes multiple layers of protection, including advanced encryption for data in transit and at rest, granular role-based access controls, multifactor authentication for system users, and detailed logging of all operations performed. Security must be considered in system design and adopt privacy by design and security by design principles [4].

A brief explanation: granularity refers to detailed and specific role-based access controls, which enable precise definition of which users can access certain information or functionalities according to their role, thus providing refined access management with precise permission settings for each user type.

Privacy by design is a principle establishing that privacy must be considered in the initial system development, not as an element added later, meaning that data protection measures are incorporated into the very architecture of the diagnostic AI system to ensure that personal data protection is fundamental. Security by design, similarly to the previous one, means that security must be a foundational element integrated in the design phase, that is, an integral part of the architecture, processes, and functionalities, which makes the system intrinsically more secure and less vulnerable to threats.

Informed consent becomes newly complex in the context of AI in healthcare. Patients need to understand not only that their data will be used for their own diagnosis, but also that they can be employed to train and improve AI systems. The referred informed consent must be granular, to enable individuals to exercise control over different uses of their medical information [4].

Secure interoperability between systems represents another crucial aspect. Diagnostic AI systems often need to integrate data from multiple healthcare sources and institutions, and this integration must preserve data confidentiality and integrity through standardized medical information exchange protocols that incorporate robust security mechanisms [1].

Data storage and retention require specific policies that balance clinical needs, regulatory requirements, and individual rights; thus, unavoidably, this requires the definition of appropriate retention periods, secure data disposal procedures, and mechanisms for patients to exercise rights such as access, correction, and deletion of their information [4].

Data governance must also consider data quality and integrity aspects. Diagnostic AI systems critically depend on information accuracy and completeness; therefore, they require rigorous data validation processes, inconsistency detection, and audit trails that document all data handling activities [2].

Data sharing for research and development represents an area that demands specific governance. Although it is fundamental for scientific advancement and system improvement, it must adopt protocols that guarantee effective

anonymization and ethical use of information, with special consideration of the risks of re-identification through database cross-referencing [6].

Regulatory compliance is a cross-sectional element of governance that considers multiple regulatory frameworks, such as personal data protection legislation, specific healthcare regulations, and ethical guidelines for AI use, whose compliance must be continuously monitored and documented [4].

Finally, governance must incorporate accountability and transparency mechanisms such as: clear documentation of policies and procedures, regular audits, channels for questions and complaints, and proactive communication with stakeholders about data management practices. Data governance needs to be dynamic and adaptive so as to continuously evolve in order to assimilate new technological challenges, regulatory requirements, and social expectations as to healthcare data protection. Successful implementation of data governance is fundamental for building and maintaining the trust necessary for the successful adoption of diagnostic AI systems [6].

Liability framework

The integration of AI into medical diagnostics establishes a new paradigm of professional liability that comprises multiple critical dimensions, and three dimensions should be considered: the medical professional, the CFM, and the AI system requirements. The training framework arises as a fundamental requirement, demanding a comprehensive program for training on algorithmic functionalities, systemic limitations, and adequate integration of automated recommendations into the decision-making process [5].

Regarding physicians, the decision-making process maintains the primacy of medical judgment, and AI systems are positioned as complementary advisory tools. The critical assessment of automated recommendations should be based on professional expertise and the particularities of each clinical case, in order to avoid exclusive dependence on automation [7].

Incident governance establishes responsibility for the systematic notification of errors and biases identified in systems. This is a true feedback mechanism that constitutes an essential element for algorithmic refinement and continuous optimization of operational safety and effectiveness [2].

The intelligent system use validation process requires judicious assessment of its applicability in specific populations, based on an analysis that addresses fundamental performance metrics such as false positive/negative rates and systematic monitoring of potential degradation in diagnostic accuracy [2].

Transparent communication with patients is an ethical imperative and demands adequate clarification of AI's participation in the diagnostic/therapeutic process, in order to preserve patient autonomy through informed consent and understanding of technological implications [9].

Responsible use presupposes strict adherence to the intended scope of the systems, to avoid inappropriate applications or extrapolations that may result in damage. Clinical judgment should guide the mitigation of potential risks through the understanding of systemic limitations [8].

Continuous professional development arises as a requirement for consistency with technological evolution: it is a process that includes obtaining up-to-date knowledge on new systems, algorithmic modifications, and best operational practices [2].

Ethical practice demands the adoption of rigorous standards in AI use to preserve patient rights, confidentiality, and healthcare equity. Compliance with these principles ensures that intelligent systems contribute positively toward quality and safe diagnostic process and, thus, toward preserving the integrity of medical liability [4].

Regarding the CFM, the organization plays a strategic role in guiding and regulating the ethical implementation of AI in healthcare, with multiple fundamental attributions. First, the development of guidelines constitutes a primary responsibility, demanding the systematic creation and updating of ethical standards for AI use. This framework must include essential principles of safety, transparency, liability, and fairness, in order to establish a model for ethical implementation [8].

The role of professional supervision requires active advocacy of policies protecting patient rights, ensuring the ethical use of systems and

resolving critical issues such as informational privacy, algorithmic biases, and democratization of access, all while operating in coordination with regulatory agencies for the development of an appropriate legal framework [4].

Promoting algorithmic transparency arises as an imperative, as it should encourage developers to implement interpretable systems. In parallel, sponsoring professional capacity-building in critical assessment of automated recommendations is also a measure that is the responsibility of the CFM, so continuing professional development provides resources and opportunities for training on emerging technologies, ethical implications, and responsible integration into clinical practice [2].

The promotion of training in new technologies is intimately connected to the establishment of standards and certifications aimed at ensuring the exclusive implementation of secure, effective systems that comply with ethical principles. This is a process that provides for the definition of objective parameters for performance, data governance, and integration in healthcare services [5].

The ethical review process constitutes a fundamental safeguard, through which new technologies are submitted to systematic scrutiny prior to widespread adoption. It is a preventive assessment that traces and addresses potential ethical implications [6].

Fostering collaboration between developers, researchers, and clinicians enables the development of systems aligned with medical ethics. This integration aims at supporting research in ethical applications and sharing of best practices [1].

Social engagement aims at promoting public trust and system alignment with community values. This process incorporates patient perspectives into technological development and responds to social concerns [6].

The CFM's ultimate measure is the systematic monitoring of healthcare impact, with adjustments based on guidelines and policies that respond to arising ethical issues through continuous feedback from professionals and patients [2]. The CFM's comprehensive operation aims at ensuring that AI use in healthcare complies with rigorous ethical standards, safeguard patient interests, and support medical practice and at promoting the responsible advancement of technological medicine.

Considering the requirements of the intelligent system, a practical diagnostic AI assistant ethical assessment framework must include multiple analytical dimensions that allow for systematic ethical compliance verification at all stages of the system life cycle. This assessment begins with an analysis of the quality and representativeness of data used to train the system, checking for potential biases in datasets, ensuring adequate demographic diversity, and assuring that the data was obtained ethically and with appropriate patient consent [3].

Algorithmic transparency is the second pillar in the assessment, demanding verification of the system's ability to explain its decision-making processes in a way that is understandable by healthcare professionals and patients. The system must allow for traceability of decisions and identification of factors that influenced each suggested diagnosis. Transparency should extend to known system limitations, which need to be clearly documented and communicated [4].

The assessment must consider aspects of equity and non-discrimination; thus, it must verify if the system maintains consistent diagnostic accuracy levels for different population groups. The measure requires rigorous statistical system performance analysis stratified by relevant demographic characteristics, such as age, gender, ethnicity, and socioeconomic conditions. Any traced disparities must be properly investigated and resolved [3].

Data privacy and security must be evaluated according to objective criteria that verify the provision of appropriate technical and organizational controls, including analysis of anonymization mechanisms, encryption protocols, access controls, backup and recovery procedures, and data retention and disposal policies. The assessment must also verify the existence of clear procedures for responding to security incidents [4].

The impact on the medical autonomy and decision-making process represents a crucial aspect of ethical assessment. The analysis must address how the system integrates into the clinical workflow and whether it appropriately preserves the autonomy of healthcare professionals and their patients. The system must be clearly positioned as a decision-making support tool, without replacing

medical judgment or compromising the doctor-patient relationship [7].

There must be clear mechanisms for determining liability in case of errors or failures, and the assessment must examine the determination of roles and responsibilities, human supervision processes, audit procedures, and mechanisms for tracing issues and correcting traced issues [5].

System ethical sustainability and maintenance over time must be considered and achieved by verifying the effective implementation of processes for continuous monitoring of performance, model updates, validation of new incorporated data, and adaptation to changes in ethical standards or regulatory requirements. The assessment must also examine contingency plans for situations of system degradation or obsolescence [1].

Alignment with established ethical regulatory frameworks and guidelines needs to be systematically verified and include compliance with data protection legislation, specific healthcare regulations, and professional ethical standards. The assessment must also consider compliance with relevant international guidelines for AI use in healthcare [6].

Finally, the assessment process must produce a detailed report documenting findings, specific recommendations for problem correction, and action plans for implementing improvements. This report must be periodically updated, in a continuous cycle of ethical assessment and improvement of the system. The assessment framework must be applied not only during the initial development of the system but also periodically throughout its operation, in order to ensure the continuous maintenance of high ethical standards. The frequency and depth of reassessments can be adjusted based on factors such as system criticality, the pace of technological evolution, and changes in the regulatory context [2].

Results

The analysis provided the identification of two main results: 1) a hierarchy of ethical challenges structured in six levels (data quality and integrity; algorithmic transparency and security; human-machine interface; equity and justice; liability and governance; and ethical sustainability); 2) an assessment model with specific metrics for each ethical dimension: equity (accuracy disparity between groups), transparency (time for understandable explanations), privacy (re-identification test success rate), autonomy (physician-system agreement), governance (incident resolution time), and sustainability (frequency of ethical updates).

Hierarchy proposal

The ethical implementation of AI systems in medical diagnostics presents a complex hierarchy of interrelated challenges, which can be structured in increasing levels of complexity and interdependence. At the most fundamental level are the challenges related to data quality and integrity, as they constitute the foundation upon which the entire diagnostic system will be built. Data quality directly impacts diagnostic accuracy and potential system biases, therefore establishing a fundamental dependency relationship with all higher levels of the hierarchy [3].

At a second hierarchical level are the technical execution challenges, which include algorithmic transparency and system security. The first depends directly on data and system architecture quality, while security needs to be considered both at the data level and in algorithmic processes—these are technical aspects that constitute the framework that is necessary for addressing more complex ethical issues at higher levels [4].

The third level includes challenges related to the human-machine interface and the system integration into the clinical setting, encompassing aspects such as the preservation of medical autonomy, adequate communication of the system's capabilities and limitations, and the maintenance of the doctor-patient relationship. This level fundamentally depends on the solidity of the previous levels, as an effective interface requires reliable data and technically robust systems [9].

At the fourth level are the challenges posed by equity and justice in the distribution of technology benefits. This involves ensuring equitable access to diagnostic AI systems and preventing algorithmic discrimination. Equity can only be achieved with representative data, transparent systems, and adequate interfaces [6].

The fifth level addresses liability and governance challenges, with the establishment of clear frameworks for determining liability and managing risks. Effective governance requires understanding and control of all aspects of the system, from data to social impacts [5].

At the highest level of the hierarchy are challenges related to system ethical sustainability and evolution over time. They encompass the capacities for adaptation to new ethical standards, incorporation of technological advancements, and response to changes in social needs. This level depends on the harmonious functioning of all previous levels and requires robust mechanisms for continuous monitoring and adjustment [1].

The hierarchical structure shows that failures at the most fundamental levels can compromise the entire ethical implementation of the system. For example, data quality issues can propagate through all levels and affect everything from technical accuracy to social equity. Similarly, deficient algorithmic transparency can negatively impact user trust and governance effectiveness [3].

The prioritization of initiatives within this hierarchy must adopt a bottom-up approach, that is, it must first ensure the solidity of the fundamental levels and only then advance to more complex challenges. However, the planning must adopt a top-down view and consider how decisions at each level will impact the system's broader ethical goals [6].

The hierarchical framework must be used as a dynamic guide for ethical implementation, based on the recognition that the different levels, although distinct, operate in an integrated manner and require continuous attention for maintenance of the ethical integrity of the system as a whole. Understanding the dependency and priority relations is fundamental for effectively allocating resources and developing successful strategies [2].

Metrics

The assessment of the ethical effectiveness of AI systems in medical diagnostics requires a comprehensive set of metrics that enable systematic monitoring of compliance with ethical principles. The metrics can be organized into complementary dimensions that capture different aspects of the system's ethical performance. Within the scope of equity and non-discrimination, essential metrics include the disparity in diagnostic accuracy between different demographic groups, calculated by statistically comparing true positive and negative rates stratified by characteristics such as gender, age, ethnicity, and socioeconomic status. The standard deviation of these rates between groups provides a quantitative indicator of the uniformity of system performance. Additionally, training data representativeness must be measured by the population similarity index, which compares the demographic distribution of the data with the target population [3].

To assess algorithmic transparency, relevant metrics include the average time required for generating understandable explanations of diagnostic decisions, the explanation completeness rate measured by the ratio of identifiable decision-making factors, and the healthcare professional satisfaction index for provided explanations, collected through structured surveys. Decision traceability can be quantified by the percentage of diagnoses for which it is possible to completely reconstruct the decision-making chain [4].

As for privacy and security, critical metrics involve the success rate of re-identification tests on anonymized data, the average time to detect and respond to unauthorized access attempts, and the compliance rate for established security protocols. The effectiveness of access controls can be measured by the ratio of properly authorized accesses in periodic audits [6].

Medical and patient autonomy can be assessed through metrics such as the rate of agreement between system recommendations and final physician decisions, the average time dedicated to discussing recommendations with patients, and the patient satisfaction index for involvement in decision-making processes. The frequency of system recommendation replacement provides an important indicator for the preservation of clinical judgment [9].

For governance and liability, relevant metrics include the average time for investigating and resolving reported incidents, the rate of implementation of ethical audit recommendations, and the completeness index for critical decision documentation. Supervision mechanism

effectiveness can be measured by the ratio of proactively traced issues *versus* issues reported externally [5].

System ethical sustainability can be assessed by metrics such as the frequency of model updates to incorporate new ethical standards, the average time for adaptation to regulatory changes, and the obsolescence rate of critical system components. Continuous investment in ethical improvement can be quantified by the budget ratio dedicated to ethics and compliance initiatives [1].

Metrics effectiveness must be periodically re-assessed through meta-analyses that examine their ability to predict and prevent ethical issues, including an analysis of the correlation between indicators and the occurrence of ethical incidents, as well as an assessment of the completeness of the coverage of ethical principles by the established metrics [2].

The metrics framework must be implemented through an integrated dashboard that allows for continuous monitoring and early identification of concerning trends. Alert thresholds must be established for each metric, which, when exceeded, trigger in-depth analyses. Weighted aggregation of metrics can provide composite indicators of the system's overall ethical status, which facilitates communication with stakeholders and strategic decision-making [6].

The metrics must be documented in an ethical scorecard that includes a precise definition of each indicator, calculation method, measurement frequency, those responsible for monitoring and corrective actions associated with identified deviations. The scorecard must be regularly updated to reflect the evolution of the ethical concerns and the lessons learned from the operation of the system [2].

Delimitation of liability

The intersection between professional liability and technological limitations in diagnostic AI systems constitutes a complex domain that demands clear structuring of liability limits. The delimitation needs to consider both the capabilities and restrictions inherent to AI systems and the ethical and legal obligations of healthcare professionals. The physician's professional responsibility remains a central element in the diagnostic process, even with the introduction of AI systems. The healthcare professional has a fundamental obligation to exercise independent clinical judgment and make final decisions considering the specific context of each patient. The responsibility encompasses the critical assessment of recommendations provided by AI systems, considering their known limitations and potential biases [5].

The technological limitations of AI systems must be clearly documented and communicated to physicians, with precise specification of the system's scope of application, description of the populations represented in the training data, identification of situations in which the system may exhibit reduced performance, and explicit statement of expected error margins. The liability for failures resulting from undocumented or inadequately communicated limitations falls on the system developers and suppliers [8].

The liability for diagnostic errors must consider the chain of events and decisions that led to the adverse outcome. When the healthcare professional properly follows the established protocols for system use, including appropriate consideration of its known limitations, their liability should be primarily assessed in relation to the adequacy of the clinical judgment and not pertinently to the intrinsic limitations of the technological system [5].

Diagnostic AI system developers and suppliers have specific liability related to the technical quality of the system, which includes the accuracy of predictions within the specified scope, the robustness of security mechanisms, and the effectiveness of monitoring and maintenance protocols. Deviations resulting from deficiencies in these areas shall be attributed primarily to said entities, provided that the system was used as specified [5].

Healthcare institutions that implement diagnostic AI systems assume responsibility for the adequacy of the support infrastructure, professional training, and the establishment of clear use protocols. Errors resulting from deficiencies in organizational aspects shall be

attributed primarily to the institution, even when they result in specific diagnostic errors [8].

An indispensable element in the delimitation of liability is the establishment of clear protocols for incident investigation. These protocols must allow for objective identification of the origin of failures, differentiating between known technological limitations, execution errors, procedural missteps, and clinical judgment errors—the differentiation is fundamental for appropriately determining liability and developing effective corrective measures [5].

Risk management in diagnostic AI systems must incorporate mechanisms for shared liability where appropriate, in order to achieve specific insurance frameworks to cover different types of inaccuracies, contractual agreements that specify liability limits, and protocols for cooperation between different stakeholders for the resolution of identified issues [5].

Continuous system performance and usage pattern monitoring is a responsibility shared among developers, healthcare institutions and professionals. The early identification of deviations or degradation of performance allows for preventive intervention and helps establish clear limits for liability when failures occur [2].

The technological evolution and the lessons obtained through system operation should serve as the basis for periodic updates in the scope of liability, which encompass refinement of usage protocols, adjustment of liability limits, and development of new safeguards as new limitations or risks are identified [1].

Final considerations

The central results of this work, the hierarchy for ethical challenges and the assessment metrics model, provide practical tools for the responsible implementation of AI systems in medical diagnostics. These results directly address the proposed objectives by establishing concrete guidelines based on bioethical principles and providing an organizational framework for prioritizing objective continuous monitoring initiatives and mechanisms [6].

The proposed framework—based on the pillars of autonomy, beneficence, non-maleficence, and justice—provides a robust framework to guide the responsible use of intelligent diagnostic support systems. The approach recognizes both the disruptive potential of AI for enhancing diagnostic accuracy and democratizing access to healthcare and the critical need for ethical safeguards that preserve patient centrality and medical practice integrity [4].

The CFM's strategic role proves indispensable for achieving this goal, through the development of guidelines, regulatory supervision, and the promotion of professional training. Its operation must ensure that technological evolution is aligned with fundamental values of medicine and promote harmonious integration between human *expertise* and computational capabilities [8].

The clear delimitation of liability and the establishment of objective metrics for assessment of ethical effectiveness constitute essential practical contributions of this work. These tools provide concrete mechanisms for implementing the principles discussed and thus facilitate the effective implementation and continuous monitoring of diagnostic AI systems [5].

Data governance in healthcare and the hierarchy for ethical challenges provide an organizational framework that enables a systematic and prioritized approach to the multiple aspects involved. The organization facilitates the efficient allocation of resources and the development of effective strategies to address the identified issues [6].

Finally, it is emphasized that the success of the technological transformation will fundamentally depend on the ability to maintain a dynamic balance between innovation and humanization in healthcare. The unwavering commitment to ethical principles, associated with a deep understanding of the potential and limitations of AI systems, will enable the construction of a more precise and accessible medicine, without compromising the humanistic essence that characterizes the doctor-patient relationship. The new medical practice paradigm, enhanced by AI but firmly grounded in ethical values, represents a promising evolution for the future of healthcare.

References

1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med [Internet]. 2022 [acesso 7 jan 2025];28(1):31-8. DOI: 10.1038/s41591-021-01614-0
2. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim DA *et al*. A Clinician's guide to artificial intelligence: how to critically appraise machine learning studies. Transl Vis Sci Technol [Internet]. 2020 [acesso 7 jan 2025];9(2):7. DOI: 10.1167/tvst.9.2.7
3. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA [Internet]. 2019 [acesso 7 jan 2025];322(24):2377-8. DOI: 10.1001/jama.2019.18058
4. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. International Journal of Law and Information Technology [Internet]. 2019 [acesso 7 jan 2025];27(2):171-203. DOI: 10.1093/ijlit/eaz004
5. Smith H, Fotheringham K. Artificial intelligence in clinical decision-making: Rethinking liability. Medical Law International [Internet]. 2020 [acesso 7 jan 2025];20(2):131-54. DOI: 10.1177/0968533220945766
6. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC *et al*. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics [Internet]. 2021 [acesso 3 abr 2020];22(1):14. DOI: 10.1186/s12910-021-00577-8
7. Desai AN. Artificial intelligence: promise, pitfalls, and perspective. JAMA [Internet]. 2020 [acesso 3 abr 2020];323(24):2448-9. DOI: 10.1001/jama.2020.8737
8. Matsuzaki T. Ethical issues of artificial intelligence in medicine. California Western Law Review [Internet]. 2018 [acesso 3 abr 2020]; 55(1):19. Disponível: https://bit.ly/4saPVkk
9. Soellner M, Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. BMC Med Inform Decis Mak [Internet]. 2021 [acesso 3 abr 2020];21(1):236. DOI: 10.1186/s12911-021-01596-6
10. Marchiori C, Dykeman D, Girardi I, Ivankay A, Thandiackal K, Zusag M *et al*. Artificial intelligence decision support for medical triage. AMIA Annu Symp Proc [Internet]. 2020 [acesso 3 abr 2020];793-802. Disponível: https://bit.ly/4aQmME1

Ricardo Rheingantz Abuchaim – PhD – ricardoabuchaim@hotmail.com
0009-0006-9802-6648

Daniel Brito de Araujo – PhD – araujodb@gmail.com
0000-0002-4840-945X

Correspondence
Ricardo Rheingantz Abuchaim – Rua Félix da Cunha, 722, Centro. 96010-000. Pelotas/RS, Brasil

Contribution of the authors
Ricardo Rheingantz Abuchaim participated in the study conception and design, manuscript writing, critical review of intellectual content and approval of the final version for publication. Daniel Brito de Araujo participated in the study design, manuscript writing, critical review of intellectual content and approval of the final version for publication.

Data availability: All data used or generated in this study are described and presented in full in the body of the article.

Editor in charge: Dilza Teresinha Ambrós Ribeiro

Received: 4.16.2025
Revised: 9.1.2025
Approved: 1.14.2026