

ARTIFICIAL INTELLIGENCE IN MEDICINE: A SYSTEMATIC LITERATURE REVIEW OF EMERGING RISKS AND CHALLENGES

Vanessa Schmidt Bortolini *

Wilson Engelmann **

Alexandre de Souza Garcia ***

Abstract: This study examined the key risks associated with the use of artificial intelligence (AI) in medical practice, emphasizing the profound transformations that technology is driving in the healthcare sector. Based on a systematic literature review and consultation of other bibliographic sources, ten major risks were identified: biases, discrimination, social implications, bias denial, black-box problems, reinforcement of prejudices, explainability, transparency, intelligibility, and privacy. The development of AI in healthcare has led to systems that are far more autonomous and complex than initially expected, posing significant challenges due to their direct impact on human health. This situation highlights the need for regulation and oversight by the relevant public authorities. Although the regulation of new technologies requires careful consideration regarding when and how to regulate, the risks associated with AI in medicine are already well-recognized. Failing to intervene could be seen as governmental inaction in fulfilling the responsibility to ensure health and equity. AI should function as a support tool, not a replacement for physicians, ensuring that specialists validate the accuracy and effectiveness of algorithmic recommendations.

Keywords: Artificial Intelligence; Medicine; risks; regulation; systematic literature review.

INTRODUCTION

It is surprising how, in the current era, certain technologies do not follow a linear progression, advancing exponentially. According to Paolo Benanti, this suggests that the next two decades will bring such significant technological changes that they will make everything that has happened so far practically insignificant¹.

* Master's student in Law, University of Vale do Rio dos Sinos, Rio Grande do Sul, Brazil. Email: vsbortolini@gmail.com / ORCID iD: <https://orcid.org/0000-0002-3200-4845>

** PhD and Master's in Public Law, University of Vale do Rio dos Sinos, Rio Grande do Sul, Brazil. Email: wengelmann@unisinis.br / ORCID iD: <https://orcid.org/0000-0002-0012-3559>

*** PhD in Administration, University of Vale do Rio dos Sinos, Rio Grande do Sul, Brazil. Email: garcia@resultare.com.br / ORCID iD: <https://orcid.org/0000-0002-4177-7612>

¹ Benanti, Paolo. *Oráculos: Entre Ética e Governança dos Algoritmos*. São Leopoldo:

One of these disruptive technologies is artificial intelligence (AI), consisting of a combination of data, algorithms and computational capacity that imitates human intelligence² and plays an increasingly important role in the health area, offering a set of technologies and techniques that can be applied to improve diagnosis, treatment, monitoring and management of health care. Due to the current growing generation of health data, an unprecedented transformation is taking place towards a new paradigm of medical sciences³.

Services that have always been provided by humans are now starting to be influenced or even fully executed by a system, challenging the basic foundations and assumptions of healthcare as we know them. Between 2019 and 2023 alone, global spending on AI grew from 37.5 to 97.9 billion dollars, largely due to the increasing availability of electronic health records and other patient-related data, which has enormous potential to improve people's health and well-being⁴.

Although some of the risks of AI in the most diverse applications have already been documented, they are difficult to easily access in one place. In view of this circumstance, MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) launched the AI Risk Repository, a database with more than 700 documented AI risks. The project was motivated by concerns that the global adoption of AI is outpacing the way people and organizations understand all the risks of the technology⁵. MIT found that the most frequently addressed risks include safety, flaws and limitations of AI systems (76% of documents), privacy and security (68%), and misuse (68%). In addition, human-computer interaction and misinformation were identified as the least addressed concerns in risk frameworks. Fifty-one percent of the risks analyzed were attributed to AI systems rather than humans, who accounted for 34%, and 65% of the risks arose after AI was deployed rather

Unisinos, 2020.

² Facchini Neto, Eugenio, and Roberta Scalzilli. "Pode a Ética Controlar o Desenvolvimento Tecnológico? O Caso da Inteligência Artificial, à Luz do Direito Comparado." In *Tutela Jurídica do Corpo Eletrônico*, edited by Cristiano Colombo, Wilson Engelmann, and José Luiz de Moura Faleiros Junior. Indaiatuba: Foco, 2022.

³ Moreno-Sanchez, Pedro A. "An Automated Feature Selection and Classification Pipeline to Improve Explainability of Clinical Prediction Models." In *IEEE 9th International Conference on Healthcare Informatics (ICHI)*, Finland, 2021.

⁴ Markus, Aniek F., Jan A. Kors, and Peter R. Riknbeek. "The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies." *Journal of Biomedical Informatics*, 2020.

⁵ Rajkumar, Radhika. "AI Risks Are Everywhere – and Now MIT Is Adding Them All to One Database." *ZDNet*, August 14, 2024.

than during development⁶.

The research question is: what are the main risks of AI specifically in the medical field? The research was carried out through a systematic literature review on the Web of Science platform, from 2020 to 2023, using the following keywords: “ethical use of artificial intelligence in healthcare”, in addition to consulting other bibliographic and legislative sources on the subject.

The general objective of the article is to identify the main risks that may be brought about by the use of AI in medical activity. The specific objectives include: a) to study some of the possibilities of applying AI in medical activity; b) to identify some of the main risks of AI in the health and technology market, through a systematic literature review (SLR).

I. ARTIFICIAL INTELLIGENCE IN MEDICAL ACTIVITIES

Technological advances are radically transforming the healthcare sector and medical activity. Telemedicine has become widespread, allowing remote consultations and real-time monitoring of patients by remote devices⁷. Robotics is revolutionizing surgery, with millions of robotic procedures performed globally⁸. Nanotechnology is also emerging as a powerful tool in medicine, with the creation of nanorobots for precise drug delivery and targeted treatment, such as in chemotherapy, reducing side effects and increasing therapeutic benefits. AI, in turn, plays an increasingly important role in this area, offering a set of technologies and techniques that can be applied to improve diagnosis, providing a set of techniques that align with the principles of a new medicine, focused on prevention, personalized, predictive and proactive treatment⁹.

In addition, the generation of personal and health data is facilitated by the widespread use of always-connected devices, such as smartphones, the proliferation of wearables and sensors and electronic medical records. There is already technology for smartphones that alerts the doctor, using AI software, when patients are on the verge of health problems. Data mining has revealed that smartphone usage and movement patterns can indicate the onset

⁶ Rajkumar, Radhika. "AI Risks Are Everywhere – and Now MIT Is Adding Them All to One Database." *ZDNet*, August 14, 2024.

⁷ Sánchez-Caro, Javier, and Fernando Abellán. *Telemedicina y Protección de Datos Sanitarios*. Granada: Comares, 2002.

⁸ Skinovsky, James, Mauricio Chibata, and Daniel Emílio Dalledone Siqueira. "Realidade Virtual e Robótica em Cirurgia – Aonde Chegamos e para Onde Vamos?" *Revista do Colégio Brasileiro de Cirurgiões* 35, no. 5 (2008): 334–37.

⁹ Moreno-Sanchez, Pedro A. "An Automated Feature Selection and Classification Pipeline to Improve Explainability of Clinical Prediction Models." In *IEEE 9th International Conference on Healthcare Informatics (ICHI)*, Finland, 2021.

of colds, as well as the presence of anxiety or stress¹⁰.

Thus, AI is already incorporated into clinical practice in medical care, being an instrument that assists, for example, in “deciding which medication to prescribe, whether or not a certain high-risk surgery is indicated, what the probability is of a given patient developing sepsis” or which therapy has the greatest chance of preventing sudden death in a specific case¹¹. It can also be used to analyze drug interactions, optimize the fight against hospital infections, triage patients, develop public policies, provide personalized care based on genetic health data, interpret exams, monitor patients, manage health data, among other uses.

An interesting proposal is the use of AI to assist in decision-making about incapacitated patients who cannot express their will. This can be advantageous because advance directives are often inconclusive or non-existent, and the patient's guardians or family members may be influenced by their own emotions when making decisions on behalf of the patient¹². Algorithms can be used to calculate the most likely preferred treatment for incapacitated patients. Annette Rid and David Wendler first proposed this idea in 2010, using patients' sociodemographic data (such as age, sex, marital status and medical history) to predict preferred treatment options. This method became known as “PPP—Patient Preference Predictor”¹³. The development of AI to support medical decision-making, such as in cardiopulmonary resuscitation, has advantages such as reducing stress, time pressure, personal bias, conflicts of interest, and legal concerns that can influence decisions¹⁴. In addition, there are health systems that rely on the concept of “nudging,” such as apps that send notifications to prevent the progression of cognitive impairment in elderly patients. A study conducted in three hospitals in the United States revealed that both ChatGPT-3 and ChatGPT-4 were successful in the admission test for the specialization in neurological surgery, with a 60% correct answer rate on the questions on the national medical qualification exam¹⁵.

¹⁰ Benanti, Paolo. *Oráculos: Entre Ética e Governança dos Algoritmos*. São Leopoldo: Unisinos, 2020.

¹¹ Goodman, Katherine E., et al. "Preparing Physicians for the Clinical Algorithm Era." *The New England Journal of Medicine* 389, no. 6 (August 2023): 483–87.

¹² Ferrario, Andrea, Sophie Gloecker, and Nikola Biller-Andorno. "Ethics of the Algorithmic Prediction of Goal of Care Preferences: From Theory to Practice." *Journal of Medical Ethics* 49 (November 2023): 165–74.

¹³ Ferrario, Andrea, Sophie Gloecker, and Nikola Biller-Andorno. "Ethics of the Algorithmic Prediction of Goal of Care Preferences: From Theory to Practice." *Journal of Medical Ethics* 49 (November 2023): 165–74.

¹⁴ Biller-Andorno, Nikola, et al. "AI Support for Ethical Decision-Making Around Resuscitation: Proceed with Care." *Journal of Medical Ethics* 48 (March 2021): 175–83.

¹⁵ Nogaroli, Rafaella. "Implicações da IA na Medicina: ChatGPT Já Faz Diagnósticos e É Aprovado para Residência Médica." *Gazeta do Povo*, April 13, 2023.

Thus, AI seems to be aligned with a new era of medicine, focused on prevention and personalized care. However, these technological advances have also triggered a range of new risks that were previously unthinkable and unknown. Therefore, new technologies should not be valued only for their benefits, but also for the harm they can cause. In the words of Klaus Schwab, “the changes are so profound that, from the perspective of human history, there has never been a moment as potentially promising or dangerous”¹⁶. It can be observed that changes and technological advances are occurring in society and new technologies are being applied very quickly and that the State, in its normative role, does not keep up with the speed of social facts.

II. MAIN RISKS IDENTIFIED THROUGH A SYSTEMATIC LITERATURE REVIEW

The identification of the main risks of using AI in medical activity was carried out through a systematic literature review. Reviewing the literature means covering published studies that provide an assessment of the literature related to specific subjects¹⁷ (Galvão; Ricarte, 2019). To this end, the MethodiOrdinatio was used, which has nine steps (P1 to P9), namely: (P1) Establishing the research intention; (P2) Preliminary exploratory research with keywords in the databases; (P3) Defining and combining keywords and databases; (P4) Searching the databases; (P5) Filtering procedures; (P6) Identification of the Impact Factor, year and number of citations of each article; (P7) Ordering the articles through InOrdinatio; (P8) Locating the articles in full format; (P9) Reading and systematic analysis of the articles.

The MethodiOrdinatio formula is presented here:

$$\text{InOrdinatio} = (F_i / 1000) + \alpha * [10 - (A_t - A_r)] + (\sum C_i)(1)$$

In the formula:

- F_i is the impact factor;
- α is equal to 10;
- A_t is the current year of the systematic review;
- A_r is the year of publication of the article;
- C_i is the number of citations of the article.

With the aforementioned data tabulated in an electronic spreadsheet, the Index Ordinatio (InOrdinatio) is calculated, which in turn makes it possible

¹⁶ Schwab, Klaus. *A Quarta Revolução Industrial*. Translated by Daniel Moreira Miranda. São Paulo: Edipro, 2016.

¹⁷ Galvão, Maria Cristiane Barbosa, and Ivan Luiz Marques Ricarte. "Revisão Sistemática da Literatura: Conceituação, Produção e Publicação." *Logeion: Filosofia da Informação* 6, no. 1 (September 2019): 57–73.

to order the articles according to their relevance.

A total of 20 steps (E1 to E20) were also performed, as follows: (E1) Defining the research question: "How are ethical issues of AI in healthcare being addressed in the Web of Science from 2020 to 2023?"; (E2) Preliminary exploratory search with keywords in the database; (E3) Definition of keywords/terms: "Ethical use of artificial intelligence in healthcare"; (E4) Definition that the search would be by article title; (E5) Conducting the first search: 291 articles found; (E6) Definition of the time cut: 2023, 2022, 2021 and 2020; (E7) Definition of the language searched: English and Portuguese; (E8) Definition that only complete articles available in the database would be considered; (E9) Filter by area of knowledge (Ethics + Medical Ethics + Computer Science + Artificial Intelligence + Robotics), partial result: 19 articles; (E10) Generation of .xls file with all available data; (E11) Generation of .Ris and .txt files; (E12) Generation of graphs by authors in VOSviewer using the Ris file; (E13) Generation of graphs by institution and country in VOSviewer using the txt file; (E14) Analysis of the adherence of article titles to the research question: partial result 16 articles; (E15) Analysis of the adherence of article abstracts to the research question: partial result 12 articles; (E16) With the 12 selected articles available, data were sought for the MethodiOrdinatio: impact factor, year and number of citations of each article; (E17) Ordering of articles using InOrdinatio; (E18) Decision to select articles with InOrdinatio higher than 90 points, thus classifying 10 articles; (E19) Critical analysis of the 10 selected articles, using VOSviewer and Nvivo software for network and content analysis respectively; (E20) Writing of the section presented below where the articles are presented in the sequence indicated by the InOrdinatio ranking. Through the application of the method, we arrived at the ten best classified works on the topic: 1) "Ethics of the algorithmic prediction of goal of care preferences: from theory to practice"¹⁸; 2) "A systematic review of artificial intelligence impact assessments"¹⁹; 3) "A smarter perspective: Learning with and from AI-cases"²⁰; 4) "Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives"²¹; 5) "AI support for ethical decision-

¹⁸ Ferrario, Andrea, Sophie Gloecker, and Nikola Biller-Andorno. "Ethics of the Algorithmic Prediction of Goal of Care Preferences: From Theory to Practice." *Journal of Medical Ethics* 49 (November 2023): 165–74.

¹⁹ Stahl, Bernd Carsten, et al. "A Systematic Review of Artificial Intelligence Impact Assessments." *Artificial Intelligence Review* 24 (2023): 1–33.

²⁰ Ossa, Laura Arbalaez, et al. "A Smarter Perspective: Learning with and from AI-Cases." *Artificial Intelligence in Medicine* 135 (January 2023).

²¹ Aquino, Yves Saint James, et al. "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics*, February 2023.

making around resuscitation: proceed with care”²²; 6) “Responsible nudging for social good: new healthcare skills for AI-driven digital personal assistants”²³; 7) “Multi Scale Ethics—Why We Need to Consider the Ethics of AI in Healthcare at Different Scales”²⁴; 8) “Evaluation of artificial intelligence clinical applications: Detailed case analyses show value of healthcare ethics approach in identifying patient care issues”²⁵; 9) “Limiting medical certainties? Funding challenges for German and comparable public healthcare systems due to AI prediction and how to address them”²⁶; 10) ““Just” accuracy? Procedural fairness demands explainability in AI based medical resource allocations”²⁷.

An analysis of the above works, in addition to consulting other bibliographic sources, highlighted the following risks, classified below for purely didactic reasons, since the same circumstance can fit into more than one category: a) biases, b) discrimination, c) social consequences (in addition to individual ones), d) denial of the existence of biases, e) black box, f) reinforcement of prejudices, g) explainability; h) transparency; i) intelligibility; j) privacy.

a. Biases

In the context of AI in medical practice, biases represent a significant concern, as they can negatively influence treatment results and the quality of patient care. In other words, if the data used to train an algorithm is biased towards certain demographic groups, such as gender or ethnicity, the system may generate inaccurate diagnoses or unequally applicable treatment recommendations. Therefore, it is essential to identify, mitigate and correct biases in AI algorithms, adopting transparent and inclusive approaches in the selection and interpretation of data, taking into account the context in which the system will be used. This includes conducting regular audits to detect and

²² Biller-Andorno, Nikola, et al. "AI Support for Ethical Decision-Making Around Resuscitation: Proceed with Care." *Journal of Medical Ethics* 48 (March 2021): 175–83.

²³ Capasso, Marianna, and Steven Umbrello. "Responsible Nudging for Social Good: New Healthcare Skills for AI-Driven Digital Personal Assistants." *Medicine, Health Care and Philosophy* 25 (2021): 11–22.

²⁴ Smallman, Melanie. "Multi-Scale Ethics – Why We Need to Consider the Ethics of AI in Healthcare at Different Scales." *Science and Engineering Ethics* 28, no. 63 (2022).

²⁵ Rogers, Wendy, Heather Draper, and Stacy Carter. "Evaluation of Artificial Intelligence Clinical Applications: Detailed Case Analyses Show Value of Healthcare Ethics Approach in Identifying Patient Care Issues." *Bioethics* 35, no. 7 (2021): 623–33.

²⁶ Ulmenstein, Ulrich Von, et al. "Limiting Medical Certainties? Funding Challenges for German and Comparable Public Healthcare Systems Due to AI Prediction and How to Address Them." *Frontiers in Artificial Intelligence* 5 (2022).

²⁷ Rueda, Jon, et al. "Just Accuracy? Procedural Fairness Demands Explainability in AI-Based Medical Resource Allocations." *Open Forum*, December 21, 2022.

correct existing biases. In addition, it is important to promote diversity and representation in the team responsible for developing and implementing AI, ensuring a broad perspective that is sensitive to the cultural and social nuances of the patients served²⁸.

b. Discrimination

In addition to the risk categorized in the previous item, the discrimination factor found in the systematic literature review means that algorithms may present calculation “biases” that, when applied on a large scale, may result in significant injustices. There is growing evidence that benefits are not equitably distributed due to AI replicating or amplifying existing biases in society²⁹, which must be combated through measures already analyzed, such as data quality, equity, audits and impact reports.

c. Social consequences

Existing ethical guidelines on the use of AI in health and medicine currently focus on the impact of technology on the individual, in an ethical approach based on rights, without taking into account the power that technology exerts over social structures themselves. There is a neglect of the power of AI to truly shape social arrangements. AI acts as a major driver of structural changes in society and cannot be considered a simple tool for use in health.

In an analogy with automobiles, these can be seen as mere means of transportation—in the same way that AI can be thought of as a mere tool for use in different areas. However, “we only need to look outside our windows to realize that cars have shaped every decision in our lives”: where we live, who we spend our time with, where we work, where we eat. All of these decisions are shaped by whether or not we own a car³⁰.

“Advanced technologies such as AI and robotics present powerful forces for much broader change as well.” For example, studies of robotics have found that “the sheer cost of technologies means that health care needs to become more centralized, often at the expense of more local” and traditional care, resulting in more difficult access to health care for low-income families

²⁸ Ferrario, Andrea, Sophie Gloecker, and Nikola Biller-Andorno. "Ethics of the Algorithmic Prediction of Goal of Care Preferences: From Theory to Practice." *Journal of Medical Ethics* 49 (November 2023): 165–74.

²⁹ Aquino, Yves Saint James, et al. "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics*, February 2023.

³⁰ Smallman, Melanie. "Multi-Scale Ethics – Why We Need to Consider the Ethics of AI in Healthcare at Different Scales." *Science and Engineering Ethics* 28, no. 63 (2022).

who tend to have less access to transportation, potentially further deepening existing health inequalities and generating different treatments for different groups, exacerbating existing inequalities³¹.

d. Denial of the existence of bias

The research demonstrated that attention needs to be paid to the idea that denies the existence of possible biases in AI. The literature review showed that there is divergence on the following points: 1) whether biases actually exist in AI in healthcare. The majority agrees that they do, a minority denies the existence of biases, and a third group understands that the benefits of technology in healthcare outweigh any harm caused by biases; 2) what are the best strategies to combat these biases; and 3) whether or not sociocultural data, such as race and gender, should be excluded in the development of AI in an attempt to mitigate biases³².

The existing divergences demonstrate the barriers in combating biases. In any case, even denying the existence of biases, the parties are responsible for addressing them in algorithmic systems, and empirical studies are needed to understand algorithmic biases and strategies for the development of AI with participatory and diverse involvement in research³³.

e. Black box

Black box refers to the ability of AI programs to generate skills or provide responses in unexpected ways. Considering that AI algorithms often use many variables to arrive at a specific result, the complex mathematical representation is usually incomprehensible to humans. When developers create a program in a “traditional” way, the lines of code inserted are clearly reflected in the result that the software obtains. However, in AI development, engineers work to arrive at a system that imitates the “neural networks” of human intelligence. This involves a large number of interconnected processors that can handle large amounts of data, detect patterns among millions of variables using machine learning and, most importantly, adapt in response to what they are doing. The complex form of mathematical representation is, for the most part, unintelligible to humans—which is why

³¹ Smallman, Melanie. "Multi-Scale Ethics – Why We Need to Consider the Ethics of AI in Healthcare at Different Scales." *Science and Engineering Ethics* 28, no. 63 (2022).

³² Aquino, Yves Saint James, et al. "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics*, February 2023.

³³ Aquino, Yves Saint James, et al. "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics*, February 2023.

algorithms are commonly referred to as “black box” systems³⁴.

In medical practice, where AI algorithms are fed by an ever-increasing amount of both patient health data—due to the increase in data production in digitized records and wearables—and due to the massive intellectual production in the medical field—currently, medical knowledge doubles every 73 days, so that the doctor would need to dedicate 29 hours a day to absorb all the new information³⁵—the lack of explainability and the presence of black boxes presents a risk to rights such as autonomy, consent, among others.

f. Reinforcement of bias

In the context of AI in healthcare, the risk of reinforcing bias represents a significant challenge that can negatively impact the quality of healthcare. AI algorithms can inadvertently perpetuate existing biases present in the data used to train them, resulting in misdiagnoses, inappropriate treatments, or discrimination against certain groups of patients. For example, if historical data reflects inequalities in access to healthcare services or is based on social stereotypes, algorithms can reproduce and amplify these discrepancies. To mitigate the risk of reinforcing bias, it is essential to adopt measures that promote fairness and impartiality in AI systems, which include the use of representative and diverse data and the implementation of bias detection and correction techniques. In addition, it is essential to empower healthcare professionals to recognize and address biases in the use of AI, promoting fair and inclusive clinical practice.

g. Explainability

The risk identified in the doctrine refers to the lack of explainability, as this factor is closely tied to the reliability of the systems and is a prerequisite for discussions on transparency. An explainable artificial intelligence is one that “produces details that make its operation clear or easy to understand”³⁶, and the right to explanation refers to the guarantee that all individuals have the right to understand how AI-based decisions impact their lives and how they are made. Explainability stands out as an important element in enabling

³⁴ Selbst, Andrew, and Julia Powles. "Meaningful Information and the Right to Explanation." *International Data Privacy Law* 7, no. 2 (November 2017): 233–42.

³⁵ Paranjape, Ketan, et al. "Short Keynote Paper: Mainstreaming Personalized Healthcare—Transforming Healthcare Through New Era of Artificial Intelligence." *IEEE Journal of Biomedical and Health Informatics* 24, no. 7 (July 2020).

³⁶ Arrieta, Alejandro Barredo, et al. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI." Cornell University, 2019.

the existence of other basic principles of AI in health, such as the fundamental principles outlined in the Code of Medical Ethics (CFM Res. 2.217/2018).

The structural elements of the "principle of explainability" are: the effectiveness of AI systems is limited by the machine's inability to explain its thoughts and actions to human users. Therefore, explainable AI (XAI) refers to methods and techniques that generate high-quality, interpretable, intuitive, and understandable explanations of AI decisions. This component is essential for different stakeholders, including regulators, data scientists, business sponsors, and end consumers, to trust and effectively manage local government AI systems. To instill confidence in AI systems, people must be able to analyze the underlying models, explore the data used to train them, expose the reasoning behind each decision, and promptly provide coherent explanations to all stakeholders; ensuring individuals' right to know and providing users with sufficient information about the purpose, function, limitations, and impact of the AI system³⁷.

In general terms, the debate on explainability is currently divided into two interpretations. On one hand, there are those who argue for the viability and scope of the right to explanation only concerning the general functionality of the system, rather than specific decisions and individual circumstances³⁸. On the other hand, there is the understanding that the explanation should also include specific decisions, with transparency limited only by the inherently black-box nature of the algorithms³⁹. In all cases, however, there is consensus that the lack of explainability can challenge the pillars of evidence-based medicine, being a requirement for exercising autonomy and for combating the application of biased algorithms in decision-making that could promote unjustified discrimination. It is expected that the systems be understandable and explainable not only to developers but also to healthcare professionals, patients, users, and regulators, taking into account each group or individual's capacity to understand.

³⁷ Cf. Gunning, David, and David W. Aha. "DARPA's Explainable Artificial Intelligence Program." *AI Magazine*, Association for the Advancement of Artificial Intelligence, 2019, 44–59; Yigitcanlar, Tan, Juan M. Corchado, and Rashid Mehmood, et al. "Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda." *Journal of Open Innovation: Technology, Market, and Complexity* 7, no. 71 (2021); Corrêa, Nicholas Kluge, Camila Galvão, and James William Santos, et al. "Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance." *Patterns* 4 (October 13, 2023): 100857.

³⁸ Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7, no. 2 (2017).

³⁹ Selbst, Andrew, and Julia Powles. "Meaningful Information and the Right to Explanation." *International Data Privacy Law* 7, no. 2 (November 2017): 233–42.

h. Transparency

The risk identified by the doctrine is the lack of transparency. Transparency is the ethical principle most frequently found in general guidelines for the use of AI, and it is also a key principle for AI in healthcare. Implementing the transparency of algorithms is necessary for other key principles of the use of AI in healthcare to be effective, such as human autonomy, so that people remain in control of their medical decisions, and equity, in the sense of ensuring social inclusion and so that algorithms do not reproduce any type of prejudice and discrimination. Thus, it can be said that the expression of other principles presupposes transparency of AI systems⁴⁰.

Currently, the main mechanism for expressing algorithmic transparency has been precisely the right to explanation regarding automated decisions, considered a fundamental element in the regulation of algorithms. In addition to receiving an intelligible explanation, the right to be heard, to question and request review of the automated decision is created—which has been called “algorithmic due process.”

i. Intelligibility

As for the intelligibility factor, in the context of the use of AI in medical practice, it refers to the ability to understand and explain how AI systems reach their conclusions. It is crucial that algorithms are transparent and interpretable by health professionals, so that they can trust the recommendations and make informed decisions. Understanding the internal workings of algorithms allows doctors to assess the reliability and accuracy of the information provided by AI, in addition to identifying possible biases or flaws in the data. Intelligibility is also important to maintain responsibility and ethics in the use of AI, ensuring that patients understand how their information is used and have confidence in the security and privacy of their data. Systems are expected to be intelligible and explainable to developers, health professionals, patients, users and regulators⁴¹.

j. Privacy

The privacy factor means that the right to privacy of personal data must

⁴⁰ Dourado, Daniel de Araujo, and Fernando Mussa Abujamra Aith. "A Regulação da Inteligência Artificial na Saúde no Brasil Começa com a Lei Geral de Proteção de Dados Pessoais." *Revista Saúde Pública* 56, no. 80 (2022).

⁴¹ Dourado, Daniel de Araujo, and Fernando Mussa Abujamra Aith. "A Regulação da Inteligência Artificial na Saúde no Brasil Começa com a Lei Geral de Proteção de Dados Pessoais." *Revista Saúde Pública* 56, no. 80 (2022).

be guaranteed throughout the entire life cycle of the AI system, especially because sensitive data is used in medical activity. That is, the predictive analysis that AI performs in systems for medical activity may include sensitive data, which is why privacy must be a concern, in accordance with what is also determined by the LGPD, when it establishes that personal data must be treated in a specific way, with appropriate technical and administrative measures to guarantee its security and confidentiality, avoiding leaks or undue access.

Especially in the health area, the development of AI has created systems that are much more autonomous and complex than one might imagine and, as a consequence, challenges are presented that are profoundly more sensitive than in other areas because they involve the health of human beings. This circumstance highlights the need for monitoring and regulation by the responsible public entities.

Although the dilemma of regulating new technologies demonstrates the need for caution in decisions about when, how and why to regulate, the risks of using AI in medical practice are already widely known, so that the informational challenge regarding the need for regulation or not has been overcome. It is important to remember that, when we talk about medicine, we are dealing with fundamental rights such as life, health and human dignity.

In cases where the regulator intends to influence the technological arrangement in order to minimize harmful consequences—such as risks to the health of patients or discrimination against excluded groups, it is necessary for intervention to occur early, at an early stage. Given that these risks are already known, non-intervention may be seen as state inertia in its power and duty to ensure health and equity.

CONCLUSION

This study investigated the main risks associated with the application of AI in medical practice. Through a systematic literature review and consultation of other bibliographic and legislative sources, ten key risks were identified and categorized for didactic purposes, as the same circumstance may fit into more than one category: a) biases, b) discrimination, c) social consequences, d) denial of bias, e) black-box, f) reinforcement of prejudice, g) explainability, h) transparency, i) intelligibility, j) privacy.

The existence of these risks, already identified by specialized doctrine, highlights the need for monitoring and regulation of the topic by the responsible public entities. It is emphasized that the oversight of medical practice and the punishment of misconduct, due to its punitive nature, requires prior norms that clearly establish the rules of conduct that professionals must follow.

It is imperative to foster a future scenario in which AI predominantly serves a supportive role rather than replacing physicians. There must always be space for specialists to validate the accuracy of algorithmic recommendations and assess their effectiveness in real-world contexts. The final clinical assessment and professional decision-making cannot be automated, as there will be situations where, based on solid and scientific grounds, the physician should not follow the algorithm's suggestion⁴².

It is crucial to emphasize the preservation of human autonomy as one of the predominant ethical principles in the application of AI in medicine. The World Medical Association adopts the term "Augmented Intelligence" to replace "Artificial Intelligence," thus emphasizing the assistive role of these technologies. In other words, such technologies enhance the intellectual capacity of healthcare professionals rather than seeking to supplant their role⁴³. Given the volatile nature of technology, the creation of an Interdisciplinary Committee is suggested to monitor and regulate this area.

REFERENCES

- Aquino, Yves Saint James, et al. "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics*, February 2023. Accessed June 2, 2024. <https://pubmed.ncbi.nlm.nih.gov/36823101/>.
- Arrieta, Alejandro Barredo, et al. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI." Cornell University, 2019. Accessed March 3, 2024. <https://arxiv.org/pdf/1910.10045>.
- Benanti, Paolo. *Oráculos: Entre Ética e Governança dos Algoritmos*. São Leopoldo: Unisinos, 2020.
- Biller-Andorno, Nikola, et al. "AI Support for Ethical Decision-Making Around Resuscitation: Proceed with Care." *Journal of Medical Ethics* 48 (March 2021): 175–83. Accessed December 12, 2023. <https://www.medrxiv.org/content/10.1101/2020.08.17.20171769v1>.
- Capasso, Marianna, and Steven Umbrello. "Responsible Nudging for Social Good: New Healthcare Skills for AI-Driven Digital Personal Assistants." *Medicine, Health Care and Philosophy* 25 (2021): 11–22. Accessed April 3, 2024. <https://link.springer.com/article/10.1007/s11019-021-10062-z>.

⁴² Nogaroli, Rafaella. "Implicações da IA na Medicina: ChatGPT Já Faz Diagnósticos e É Aprovado para Residência Médica." *Gazeta do Povo*, April 13, 2023.

⁴³ Nogaroli, Rafaella. "Implicações da IA na Medicina: ChatGPT Já Faz Diagnósticos e É Aprovado para Residência Médica." *Gazeta do Povo*, April 13, 2023.

- Corrêa, Nicholas Kluge, Camila Galvão, and James William Santos, et al. "Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance." *Patterns* 4 (October 13, 2023): 100857. <https://doi.org/10.1016/j.patter.2023.100857>.
- Dourado, Daniel de Araujo, and Fernando Mussa Abujamra Aith. "A Regulação da Inteligência Artificial na Saúde no Brasil Começa com a Lei Geral de Proteção de Dados Pessoais." *Revista Saúde Pública* 56, no. 80 (2022). Accessed June 8, 2024. <https://www.scielo.org/pdf/rsp/2022.v56/80/pt>.
- Facchini Neto, Eugenio, and Roberta Scalzilli. "Pode a Ética Controlar o Desenvolvimento Tecnológico? O Caso da Inteligência Artificial, à Luz do Direito Comparado." In *Tutela Jurídica do Corpo Eletrônico*, edited by Cristiano Colombo, Wilson Engelmann, and José Luiz de Moura Faleiros Junior. Indaiatuba: Foco, 2022.
- Ferrario, Andrea, Sophie Gloecker, and Nikola Biller-Andorno. "Ethics of the Algorithmic Prediction of Goal of Care Preferences: From Theory to Practice." *Journal of Medical Ethics* 49 (November 2023): 165–74. Accessed May 15, 2024. <https://jme.bmj.com/content/medethics/49/3/165.full.pdf>.
- Galvão, Maria Cristiane Barbosa, and Ivan Luiz Marques Ricarte. "Revisão Sistemática da Literatura: Conceituação, Produção e Publicação." *Logeion: Filosofia da Informação* 6, no. 1 (September 2019): 57–73. Accessed December 21, 2023. <https://sites.usp.br/dms/wp-content/uploads/sites/575/2019/12/Revis%C3%A3o-Sistem%C3%A1tica-de-Literatura.pdf>.
- Goodman, Katherine E., et al. "Preparing Physicians for the Clinical Algorithm Era." *The New England Journal of Medicine* 389, no. 6 (August 2023): 483–87. Accessed May 10, 2024. <https://pubmed.ncbi.nlm.nih.gov/37548320/>. <https://doi.org/10.1056/NEJMp2304839>.
- Gunning, David, and David W. Aha. "DARPA's Explainable Artificial Intelligence Program." *AI Magazine*, Association for the Advancement of Artificial Intelligence, 2019, 44–59.
- Markus, Aniek F., Jan A. Kors, and Peter R. Riknbeek. "The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies." *Journal of Biomedical Informatics*, 2020. Accessed May 15, 2024. <https://pubmed.ncbi.nlm.nih.gov/33309898/>.
- Moreno-Sanchez, Pedro A. "An Automated Feature Selection and Classification Pipeline to Improve Explainability of Clinical Prediction Models." In *IEEE 9th International Conference on Healthcare Informatics (ICHI)*, Finland, 2021.

- Nogaroli, Rafaella. "Implicações da IA na Medicina: ChatGPT Já Faz Diagnósticos e É Aprovado para Residência Médica." *Gazeta do Povo*, April 13, 2023. Accessed May 15, 2024. https://www.gazetadopovo.com.br/opinioao/artigos/implicacoes-da-ia-na-medicina-chatgpt-ja-faz-diagnosticos-e-e-aprovado-para-residencia-medica/?fbclid=PAAab1Qw_lqKO0fGnZ1OYMioIOmw-ZC_m2lGpWJyyi5kYZra8IYpCIpgKnPCM.
- Ossa, Laura Arbalaez, et al. "A Smarter Perspective: Learning with and from AI-Cases." *Artificial Intelligence in Medicine* 135 (January 2023). Accessed May 10, 2024. <https://www.sciencedirect.com/science/article/pii/S0933336572200210X>.
- Paranjape, Ketan, et al. "Short Keynote Paper: Mainstreaming Personalized Healthcare—Transforming Healthcare Through New Era of Artificial Intelligence." *IEEE Journal of Biomedical and Health Informatics* 24, no. 7 (July 2020). Accessed April 10, 2024. <https://ieeexplore.ieee.org/document/8988253>.
- Rajkumar, Radhika. "AI Risks Are Everywhere – and Now MIT Is Adding Them All to One Database." *ZDNet*, August 14, 2024. Accessed August 15, 2024. <https://www.zdnet.com/article/ai-risks-are-everywhere-and-now-mit-is-adding-them-all-to-one-database/>.
- Rogers, Wendy, Heather Draper, and Stacy Carter. "Evaluation of Artificial Intelligence Clinical Applications: Detailed Case Analyses Show Value of Healthcare Ethics Approach in Identifying Patient Care Issues." *Bioethics* 35, no. 7 (2021): 623–33. Accessed April 10, 2024. <https://onlinelibrary.wiley.com/doi/10.1111/bioe.12885>.
- Rueda, Jon, et al. "Just Accuracy? Procedural Fairness Demands Explainability in AI-Based Medical Resource Allocations." *Open Forum*, December 21, 2022. Accessed April 10, 2024. <https://link.springer.com/article/10.1007/s00146-022-01614-9>.
- Sánchez-Caro, Javier, and Fernando Abellán. *Telemedicina y Protección de Datos Sanitarios*. Granada: Comares, 2002. Accessed May 15, 2024. <https://protecciondata.es/wp-content/uploads/2021/10/Sanchez-Caro-Javier-y-Abellan-Fernando-Telemedicina-y-Proteccion-de-Datos-Sanitarios-Edit.-El-Partal-Granada-2002.pdf>.
- Schwab, Klaus. *A Quarta Revolução Industrial*. Translated by Daniel Moreira Miranda. São Paulo: Edipro, 2016.
- Selbst, Andrew, and Julia Powles. "Meaningful Information and the Right to Explanation." *International Data Privacy Law* 7, no. 2 (November 2017): 233–42. Accessed May 15, 2024. <https://academic.oup.com/idpl/article/7/4/233/4762325>.
- Skinovsky, James, Maurício Chibata, and Daniel Emílio Dalledone Siqueira. "Realidade Virtual e Robótica em Cirurgia – Aonde Chegamos e para

- Onde Vamos?" *Revista do Colégio Brasileiro de Cirurgiões* 35, no. 5 (2008): 334–37. Accessed November 15, 2023. <https://www.scielo.br/j/rcbc/a/rFD6rx7BbL5y37gPKJMPDNK/>.
- Smallman, Melanie. "Multi-Scale Ethics – Why We Need to Consider the Ethics of AI in Healthcare at Different Scales." *Science and Engineering Ethics* 28, no. 63 (2022). Accessed November 10, 2023. <https://link.springer.com/article/10.1007/s11948-022-00396-z>.
- Stahl, Bernd Carsten, et al. "A Systematic Review of Artificial Intelligence Impact Assessments." *Artificial Intelligence Review* 24 (2023): 1–33. Accessed February 3, 2024. <https://pubmed.ncbi.nlm.nih.gov/37362899/>.
- Ulmenstein, Ulrich Von, et al. "Limiting Medical Certainties? Funding Challenges for German and Comparable Public Healthcare Systems Due to AI Prediction and How to Address Them." *Frontiers in Artificial Intelligence* 5 (2022). Accessed June 10, 2024. <https://www.semanticscholar.org/paper/Limiting-medical-certainties-Funding-challenges-for-Ulmenstein-Tretter/099516d7d9f7dbe6f3625a03e9b66ee5c05bab46>.
- Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7, no. 2 (2017). Accessed June 12, 2024. <https://academic.oup.com/idpl/article/7/2/76/3860948>.
- Yigitcanlar, Tan, Juan M. Corchado, and Rashid Mehmood, et al. "Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda." *Journal of Open Innovation: Technology, Market, and Complexity* 7, no. 71 (2021). <https://doi.org/10.3390/joitmc7010071>.

* * *

Vanessa Schmidt Bortolini

PhD Candidate and Master in Law from the Pontifical Catholic University of Rio Grande do Sul (PUCRS). Professor of Labor and Social Security Law in the Postgraduate Law Program at the Pontifical Catholic University of Rio Grande do Sul (PUCRS / UOL). Specialist in Labor and Social Security Law from the Verbo Jurídico Educacional Higher School. Professor in the Undergraduate Law Program at the São Judas Tadeu Integrated Colleges. Lawyer.

Email: vsbortolini@gmail.com

ORCID iD: <https://orcid.org/0000-0002-3200-4845>

Wilson Engelmann

PhD and Master's in Public Law (UNISINOS), Brazil; Post-Doctoral Fellowship in Public Law-Human Rights (University of Santiago de Compostela, Spain); Professor and Researcher in the Graduate Law Program - Master's and PhD, and in the Professional Master's in Business and Corporate Law, both at UNISINOS; CNPq Research Productivity Scholar; Leader of the JUSNANO Research Group.

Email: wengelmann@unisininos.br


ORCID iD: <https://orcid.org/0000-0002-0012-3559>

Alexandre de Souza Garcia

PhD in Administration (UNISINOS); Master's in Administration (UNISINOS); Specialist in Business Management (UFRGS); Economist (UFRGS). Researcher in Cooperative Identity, Innovation, Sustainability, ESG, and Innovation in Cooperatives. Guest lecturer without formal ties in Graduate Programs: ESCOOP (RS, BA, CE, and SE), UNILASALLE (RS), ICOOP (MT), FACCAT (RS), UNIAVAN (SC), and URI (RS).

Email: garcia@resultare.com.br

ORCID iD: <https://orcid.org/0000-0002-4177-7612>

 10.59224/bjlti.v2i2.1-18
ISSN: 2965-1549