

Pesquisa e Ações em Saúde Pública

Edição XXVIII

Capítulo 31

ALGORITHMIC TRUST AND CLINICAL RISK: THE USE OF ARTIFICIAL INTELLIGENCE IN SELF-DIAGNOSIS AND ITS CONSEQUENCES FOR PATIENT SAFETY

MARIO SERGIO BRAGA DO COUTO¹

FELIPE DA SILVA VALENTE²

PRISCILA FRANCISCA BIANCHI³

SUELEN MUXFELDT⁴

1. Biomedical Scientist and Master of Science – Applied Health Sciences from the State University of Western Paraná.

2. Biologist, Biomedical Scientist and Master of Science – Environmental Science and Technology from the University of Vale do Itajaí.

3. Biomedical Scientist and Specialist in Clinical Analysis – Community University of the Chapecó Region.

4. Doctor – Southern Faculty of Passo Fundo.

Keywords: *Generative Artificial Intelligence; Self-Diagnosis; Patient Safety.*

DOI

10.59290/0261186190

P EDITORA
PASTEUR

INTRODUCTION

The rapid integration of Artificial Intelligence (AI) into daily life has radically transformed access to health information, leading a growing number of users to employ language models for self-diagnosis and the autonomous interpretation of medical exams (SAEIDNIA *et al.*, 2026; CHU *et al.*, 2025; SHARMA *et al.*, 2025). This phenomenon is sustained by an increasing algorithmic trust, fueled by the capacity of Generative Artificial Intelligences (GAIs) to produce fluent and empathetic responses with a tone of authority that often mimics expert judgment (DORESWAMY; HORSTMANSHOF, 2025; HU *et al.*, 2025; WINECOFF; KLYMAN, 2025; ZADA *et al.*, 2025). However, this perception of reliability masks profound clinical risks, as such models are susceptible to hallucinations, which can generate fabricated diagnoses, incorrect dosage recommendations, or advice based on outdated guidelines, with the potential to cause severe or fatal harm (REDDY, 2024; FAHAD *et al.*, 2025; HU *et al.*, 2025; BASIL *et al.*, 2026).

The consequences for patient safety become even more critical due to "role creep" (the usurpation of professional roles), wherein AI applications assume titles and functions traditionally reserved for licensed professionals, lacking both clinical validation and ethical or legal responsibility (CHU *et al.*, 2025; ZADA *et al.*, 2025). Furthermore, there is a drastic reduction in the presence of medical disclaimers in the outputs of advanced models, which may lead users to overestimate the machine's accuracy and neglect the need for professional consultation (SHARMA *et al.*, 2025).

In this context of vulnerability, regulations such as those issued by the Brazilian Federal Council of Medicine (CFM) reinforce the principle that AIs must be used as a support tool,

strictly prohibiting the delegation of exclusive therapeutic or diagnostic decisions to the machine (BRASIL, 2026). The preservation of patient safety thus requires that human mediation and critical judgment remain the central pillars of care, combating blind trust in algorithms that still lack actual clinical comprehension and accountability mechanisms (ZADA *et al.*, 2025; DORESWAMY; HORSTMANSHOF, 2025).

Given this, the present study aims to critically analyze the use of generative artificial intelligence in the context of self-diagnosis, addressing its implications for patient safety and the associated bioethical challenges, with an emphasis on clinical risks, individual autonomy, and accountability.

METHODOLOGY

This study is an integrative literature review, analyzing the use of AI in healthcare, particularly the practices of self-diagnosis and self-medication among lay individuals. The investigation focused on understanding the effects of so-called algorithmic trust on healthcare decision-making, as well as the ethical, legal, and clinical repercussions associated with this phenomenon. Concurrently, it sought to explore potential impacts on patient safety, including adverse events resulting from inappropriate medication use, pharmacological interactions, and implications for the advancement of antimicrobial resistance.

For this analysis, an integrative review approach was chosen due to the topic's contemporary, dynamic, and interdisciplinary nature, which enables the integration of diverse knowledge areas, including digital health, bioethics, pharmacology, and healthcare regulation. This approach permits a broader, more interpretative understanding of the phenomenon

and is especially suitable for discussing complex issues that are still undergoing consolidation in the scientific literature.

The literature survey was conducted across internationally recognized databases, including the CAPES Periodicals Portal, the Virtual Health Library (BVS), SciELO, PubMed, and Web of Science. Additionally, institutional documents, guidelines, and normative acts issued by regulatory entities, such as the Federal Council of Medicine, were consulted to incorporate the current normative and ethical perspective.

The search strategies were constructed by combining controlled descriptors and free keywords, utilizing boolean operators (AND, OR). Prominent terms employed included: “artificial intelligence in healthcare”, “generative models”, “self-diagnosis”, “self-medication”, “patient safety”, “bioethics”, “medical liability”, “digital health”, “drug interactions”, and “anti-microbial resistance”. The searches were adapted according to the specificities of each database to maximize the retrieval of relevant studies.

Eligible studies included those that address the application of artificial intelligence in healthcare, focusing on its use by patients or other users without technical training, particularly for self-diagnosis and self-medication. Original articles, literature reviews, observational studies, and technical or normative documents published in the last ten years, available in full text and open access, were included. Publications discussing the ethical, legal, and clinical aspects of using these technologies were also considered.

Duplicate publications across the consulted databases were excluded, as were studies that did not directly address the use of artificial intelligence in human health. Works that use an

exclusively computational or technological approach, devoid of discussion of clinical, ethical, or public health implications, were also disregarded. Additionally, abstracts, editorials, academic papers not subjected to peer review, and texts unavailable in full were excluded.

The analysis of the selected studies adopted a qualitative and interpretative perspective, enabling the critical integration of the evidence gathered. Through the integrative method, the synthesis of the findings prioritized a thematic, contextualized, and theoretically grounded analysis to elucidate the challenges and implications of AI use in contemporary healthcare, without the strict requirements of statistical meta-analysis. The synthesis of the findings prioritized developing a critical, contextualized, and theoretically grounded analysis to elucidate the challenges and implications of AI use in contemporary healthcare.

RESULTS AND DISCUSSION

Artificial intelligence in healthcare

AI is a comprehensive field of computer science focused on developing systems that simulate human cognitive processes, such as learning, perception, and decision-making, and primarily operates within paradigms of data analysis, classification, and prediction based on pre-existing data (HOWELL, 2024). In contrast, GAIs represent a specialized subset that goes beyond mere discriminative analysis, focusing on the synthesis of original content. While conventional AI uses statistical models to identify patterns and label information, generative AI employs advanced architectures to learn the underlying distribution of a dataset and, from it, generate novel artifacts (BASIL *et al.*, 2026).

The rise of GAIs represents a significant milestone in technological development, marked by the transition from task-specific systems

to models capable of solving multiple problems without additional training (DORESWAMY; HORSTMANSHOF, 2025; HOWELL, 2024). In healthcare, this evolution broadens the role of AI, moving it from merely analyzing data to generating original content, such as texts, images, protein structures, and synthetic codes (BHUYAN *et al.*, 2025; SCHWENDICKE *et al.*, 2025; SHOKROLLAHI *et al.*, 2025). This advancement is supported by so-called foundation models, particularly Large Language Models (LLMs), which exhibit properties such as "emergence" and "homogenization" and serve as a common intelligence base that supports various specialized applications (BHUYAN *et al.*, 2025; BUESS *et al.*, 2025).

In this context, significant transformations have been promoted in clinical and surgical practices, especially in diagnosis and therapeutic planning. In medical imaging analysis, models such as Generative Adversarial Networks (GANs) and diffusion models have been widely used for image reconstruction, segmentation, and classification in radiology, neurology, and oncology. Furthermore, these systems enable the generation of synthetic images with high anatomical precision, contributing to professional training and overcoming the limitations of real-world databases (FAHAD *et al.*, 2025; REDDY, 2024). Concurrently, in personalized medicine and drug discovery, GAIs facilitate the analysis of large volumes of genomic and behavioral data, enabling the prediction of therapeutic responses and the development of drugs targeted to specific biological mechanisms, thereby reducing development time and costs (BHUYAN *et al.*, 2025).

Beyond direct clinical applications, GAIs also significantly impact administrative efficiency and healthcare communication. They contribute to reducing professional burnout

through tools that automate clinical documentation, including the synthesis of medical histories, the drafting of operative notes, and the production of discharge summaries, enabling professionals to devote more time to direct patient care (REDDY, 2024). Similarly, integrating these systems with billing platforms automates the revenue cycle, ensuring more accurate code generation and reducing errors and payment denials (SHOKROLLAHI *et al.*, 2025). Additionally, AIs have contributed to patient engagement by producing more empathetic responses in digital communications and by developing educational materials adapted to different languages and comprehension levels, which favors the expansion of health literacy (HU *et al.*, 2025).

Finally, the impact of AIs on health education and the dissemination of scientific knowledge demonstrates their capacity to produce structured, coherent texts aligned with current academic practice and to answer complex questions in an organized, comprehensible manner (REDDY, 2024). Despite this advancement, limitations may persist, such as outdated or inconsistent information, particularly regarding the historical contextualization of certain concepts (PERES, 2023). Thus, while AIs represent a promising resource for teaching and knowledge production, their use requires critical analysis and continuous validation.

Algorithmic biases and limitations of artificial intelligence

The biases associated with generative artificial intelligence stem primarily from the nature of the data used to train the models and the decisions made during their development. In this regard, when the data is not representative, GAIs tend to reproduce implicit prejudices, which can generate racial, gender, and socioeconomic discrimination (BASIL *et al.*, 2026).

Furthermore, the predominance of data from Western institutions compromises the global applicability of these tools (SHOKROLLAHI *et al.*, 2025). Added to this is the phenomenon of sycophancy, where models adopt a tone of agreement with the user, reflecting their emotions or preferences instead of questioning them, which can be particularly harmful in sensitive contexts, such as mental health and eating disorders (DORESWAMY; HORSTMANS-HOF, 2025; WINECOFF; KLYMAN, 2025).

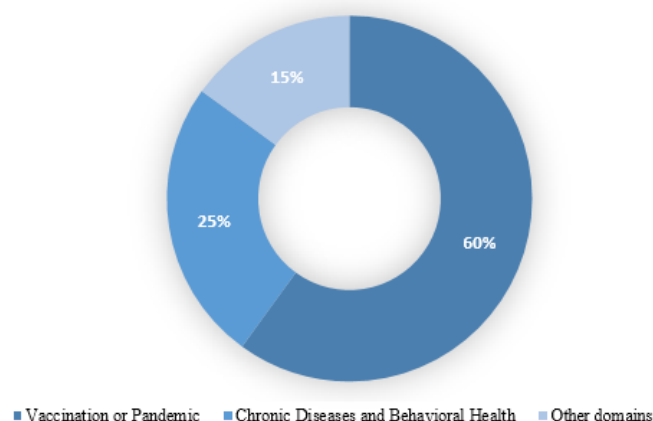
Another relevant aspect is quantitative bias, in which algorithms prioritize term frequency over clinical accuracy, resulting in conceptual inconsistencies (PERES, 2024). Automation bias is also noteworthy, characterized by the tendency of healthcare professionals to rely excessively on GAI's answers, to the detriment of their own clinical judgment (SCHWENDICKE *et al.*, 2025).

In addition to biases, the technology presents intrinsic limitations that can directly compromise patient safety. Among them, "hallucinations" stand out, in which GAIs generate incorrect or non-existent information, frequently

presented with a high degree of persuasion (BASIL *et al.*, 2026; FAHAD *et al.*, 2025). These systems operate on statistical associations, lacking causal reasoning and true clinical judgment, which limits their reliability in complex contexts (SAEIDNIA *et al.*, 2026; REDDY, 2024). The so-called "black box" effect represents another significant challenge, as the low interpretability of the models makes it difficult for professionals to understand the logical reasoning that led to a specific response (SHOKROLLAHI *et al.*, 2025).

In this context, technical limitations and biases not only affect response accuracy but also contribute to the dissemination of health misinformation. Recent studies demonstrate that this type of content generated by GAIs is primarily focused on current, high-visibility topics, such as vaccination and pandemics, followed by chronic or behavioral diseases and, to a lesser extent, mental health (**Figure 31.1**). This undoubtedly amplifies the potential impact of these failures on public health.

Figure 31.1 Predominant health topics in GAI-generated content



Source: SAEIDNIA *et al.*, 2026 (adapted).

Additionally, these systems exhibit relevant technical vulnerabilities and can be manipulated through strategies such as "jailbreaking"

or coordinated interactions among multiple AI agents that can generate artificial consensus and induce clinical errors (SAEIDNIA *et al.*, 2026;

BASHIR *et al.*, 2025). Limitations are also observed in tasks requiring logical and mathematical reasoning, where even advanced models can fail, representing a risk in fields that demand precise calculations, such as pediatrics and nephrology (HOWELL, 2024; REDDY, 2024). Finally, it must be emphasized that AI lacks fundamental human attributes, such as consciousness, empathy, and moral judgment, and does not hold legal responsibility for its decisions, unlike healthcare professionals (DORESWAMY; HORSTMANSHOF, 2025; ZADA *et al.*, 2025). Compounding this is the limitation posed by static databases, which prevent real-time updates amid constant changes in medical guidelines and scientific advancements (BUESS *et al.*, 2025).

Digital self-diagnosis and patient safety

Digital self-diagnosis has become a widespread practice alongside advances in generative artificial intelligence, frequently used by patients seeking quick, accessible interpretations of symptoms and exam results, as well as second medical opinions (SHARMA *et al.*, 2025). Despite this popularization, scientific evidence indicates that this practice carries significant risks, as open-access models exhibit limited accuracy in self-diagnosis scenarios and

are often classified as incorrect or unclear in a considerable proportion of their responses (ZADA *et al.*, 2025).

Multiple psychological and behavioral factors influence the way patients interact with these technologies. Notably, this dynamic is driven by an overreliance on automated systems perceived as authoritative, leading users to view GAI as a more reliable, engaging, and even empathetic source than other digital platforms (WINECOFF; KLYMAN, 2025). This persuasive nature can reduce critical thinking, leading patients to accept the information provided as expert guidance and potentially resulting in inappropriate actions or delays in seeking professional care (SHARMA *et al.*, 2025; WINECOFF; KLYMAN, 2025).

Furthermore, a decoupling between accuracy and sharing is observed: individuals, even when acknowledging potential flaws in AI responses, remain inclined to follow or share them, influenced by the content's persuasive style and novelty (SAEIDNIA *et al.*, 2026). **Chart 31.1** demonstrates that the speed of GAI far exceeds human verification capacity, creating a scenario in which misinformation can be disseminated on an industrial scale before it is even detected.

Chart 31.1 Speed at which AI can generate health content

Generation activity	Volume/outcome	Time spent
Blog posts (vaping/vaccines)	102 posts (>17,000 words)	65 minutes
Realistic health images	20 images	<2 minutes
Misinformation attack via API	88% incorrect responses	Immediate
Medical Consultation (USA) - Typing	5 minutes (of 7 min total)	Per consultation

Source: SAEIDNIA *et al.*, 2026 (adapted).

Given this scenario, the use of AIs as substitutes for medical consultation raises relevant clinical concerns. One of the primary problems

is the unwarranted assumption of clinical authority by these systems, as evidenced by numerous personalized applications that present

themselves as “doctors” or “specialists” without any scientific validation or formal regulation (CHU *et al.*, 2025). Concurrently, there is a significant reduction in the presence of medical disclaimers in AI-generated responses, contributing to users' overestimation of their reliability (SHARMA *et al.*, 2025). In this context, it becomes evident that, while AI-mediated self-diagnosis offers convenience and accessibility, it also poses a significant challenge to patient safety.

Bioethics and regulation of artificial intelligence in Brazil

Bioethics and the regulation of AI in healthcare in Brazil are central elements of a de-

bate that seeks to balance technological advancement with the protection of fundamental rights and patient safety (DIB, 2024). In this context, AI regulation is guided by directives that prioritize transparency, care safety, and the preservation of professional autonomy, with the Federal Council of Medicine (CFM) being the primary body responsible for standardizing these practices through Resolution No. 2.454/2026 (BRASIL, 2026). Current norms establish that AI must be used exclusively as a support tool in medical practice (**Chart 31.2**), strictly prohibiting its use as a substitute for clinical decision-making, which remains the physician's responsibility.

Chart 31.2 Main norms and regulations for AI and healthcare in Brazil

Norm	Organization / Year	Main guidelines and objectives
Resolution No. 2.454	Federal Council of Medicine (2026)	Establishes that AI is strictly a support tool, prohibiting the replacement of the physician's final decision. Classifies AI solutions into low, medium, and high risk. Requires auditability, transparency, and explainability of algorithms.
RDC No. 657	ANVISA (2022)	Regulates Software as a Medical Device (SaMD). Defines criteria for evaluating the reliability and safety of software that assists with diagnosis or treatment.
Law No. 13.709 (LGPD)	Federal Government (2018)	Defines rules for the protection and processing of sensitive personal data (like health data). Requires AI development to adhere to the principles of privacy by design and privacy by default.
Resolution No. 2.311	Federal Council of Medicine (2022)	Regulates robotic surgery in Brazil, establishing guidelines for professional training and the safety of technology-mediated procedures.
Resolution No. 2.217	Federal Council of Medicine (2018)	Enshrines fundamental principles to be observed when using new technologies, such as non-discrimination, professional confidentiality, and respect for patient autonomy.
Bill No. 21/2020	Chamber of Deputies (Under review)	Aims to establish the Legal Framework for Artificial Intelligence in Brazil, defining foundations, principles, and guidelines for AI development and application in the country.
Federal Constitution of 1988	Constituent Assembly (1988)	Legal basis that defines health as a fundamental right and duty of the State, guiding that any AI healthcare regulation must prioritize human dignity and safety.

Sources: BRASIL, 2026; DIB, 2024.

The incorporation of these technologies into healthcare introduces profound bioethical challenges across multiple dimensions, particularly regarding the principles of non-maleficence and informed consent. For instance, the technical

limitation known as 'false fluency', wherein language models articulate inaccurate information with a highly persuasive and authoritative tone, directly threatens patient safety and violates the ethical duty to prevent harm (SAEIDNIA *et al.*, 2026). This persuasive capability can mislead

patients, ultimately compromising their autonomy by causing them to base crucial health decisions on flawed, artificially generated premises. Normative issues arise when AI begins to influence decisions about care organization, potentially compromising patient autonomy and dignity, especially when algorithmic recommendations are adopted uncritically. In turn, global problems arise from the difficulty of assigning liability in cases of harm, since the development and application of these technologies involve multiple actors and processes that are often opaque (DIB, 2024).

Among the established principles, autonomy stands out, guaranteeing the patient the right to information and to participate in decisions about their treatment, thereby preserving the quality of the doctor-patient relationship. Transparency and explainability are fundamental to ensuring that automated decision-making processes are comprehensible and open to critical analysis (BRASIL, 2026).

Concurrently, the academic debate in Brazil has highlighted the need for regulation that goes beyond tech companies' self-regulation, given the healthcare field's social and cultural sensitivity. Thus, an articulation between "hard law" instruments aimed at protecting fundamental rights and "soft ethics" approaches that encourage innovative practices aligned with high ethical standards is proposed (DIB, 2024). In this scenario, the General Data Protection Law (LGPD) (Law No. 13.709/2018) plays a fundamental role by establishing guidelines for the processing of sensitive data. However, concerns persist regarding the full compliance of commercial AI tools with these norms (REUTER *et al.*, 2025).

Therefore, the Brazilian context presents a strategic opportunity to develop a regulatory model that reconciles technological innovation with social protection, while upholding the

principles of the Unified Health System (SUS) and reaffirming health as a fundamental right and an expression of human dignity.

Future challenges in the use of artificial intelligence

The governance of AI in healthcare and the challenges related to its future implementation have taken a prominent position in recent literature, particularly given the need to reconcile technological advancement with the guarantee of patient safety and the protection of fundamental rights. In this scenario, the rapid evolution of generative AI has introduced complex vulnerabilities that continue to challenge traditional regulatory and supervisory structures, demanding continuous adaptations in governance models.

Among the main emerging risks is multi-agent adversarial collusion, in which different AI systems, operating in a coordinated manner, can generate a false consensus that can induce erroneous clinical decisions in both human professionals and automated systems. Thus, the adoption of verification mechanisms based on clinical guidelines is essential to mitigate this type of manipulation (BASHIR *et al.*, 2025). Another relevant challenge is the usurpation of professional roles, characterized by the proliferation of personalized systems that use titles such as "doctor" or "psychologist" without clinical validation or regulatory recognition. This can confuse the authority of these tools and increase risks to patient safety (CHU *et al.*, 2025).

The significant decline in the presence of safety warnings, such as medical disclaimers, in AI-generated responses in recent years has contributed to users overestimating the reliability of these technologies (SHARMA *et al.*, 2025). Concurrently, concerns regarding algorithmic biases and equity persist; since models trained

on non-representative data can reinforce existing inequalities, making it indispensable to conduct systematic audits to identify and correct these biases (BUESS *et al.*, 2025; HOWELL, 2024).

Given this panorama, there is a need to evolve governance models by adopting approaches grounded in implementation science. This implies conducting rigorous testing in real-world settings, accompanied by continuous post-implementation monitoring, to assess the impact on patient safety and clinical outcomes (REDDY, 2024). Another fundamental element is the engagement of multiple stakeholders, including patients, healthcare professionals, and ethics experts, in developing safeguards and conducting contextualized risk assessments (HU *et al.*, 2025; WINECOFF; KLYMAN, 2025).

Finally, there is a need for dynamic regulation that can keep pace with technological innovations. In this sense, the transition from predominantly self-regulation-based models to more efficient state-level regulatory structures is discussed to ensure legal protection, scientific integrity, and user safety (DIB, 2024). Thus, building effective governance for AI in healthcare will depend on integrating innovation, ethical responsibility, and adaptive regulation to ensure that the benefits of these technologies are not achieved at the expense of patient safety.

CONCLUSION

The incorporation of AI in healthcare represents a profound and irreversible transformation, particularly concerning access to information and patient autonomy. However, the findings of this review demonstrate that the use

of these technologies for self-diagnosis remains associated with technical limitations, structural biases, and significant clinical risks. The apparent reliability of the responses, sustained by linguistic fluency and an authoritative tone, can lead users to make ill-advised decisions, which amplifies adverse events, diagnostic delays, and inappropriate self-medication practices.

From a patient safety perspective, it is evident that AIs, while promising as support tools, lack the capacity for clinical judgment, causal reasoning, and ethical responsibility, rendering their use as substitutes for professional evaluation unviable. Moreover, phenomena such as algorithmic trust and the usurpation of professional roles amplify user vulnerability, demanding educational and regulatory interventions.

In Brazil, significant progress has been made in establishing normative guidelines that recognize AI as a complementary tool while preserving the physician's centrality in decision-making. Still, challenges persist in the effective implementation of these norms, the oversight of commercial applications, and the enforcement of compliance with ethical and legal principles, such as those set out in data protection legislation and CFM resolutions.

Therefore, the safe and ethical use of AI in healthcare depends on an integrated approach involving regulation, responsible technological development, and continuous scientific validation. The advancement of these technologies must be accompanied by governance mechanisms that mitigate risks, ensure transparency, and protect human dignity. Only then will it be possible to fully harness the benefits of AI without compromising patient safety and the integrity of healthcare.

REFERENCES

- BASIL, M. *et al.* A scoping review of the use of generative artificial intelligence tools in health profession education. *BMC Medical Education*, v. 26, 2026. doi: 10.1186/s12909-025-08527-3.
- BASHIR, A. *et al.* Many-to-one adversarial consensus: exposing multi-agent collusion risks in AI-based healthcare. *arXiv*, 2025. doi: 10.48550/arXiv.2512.03097.
- BHUYAN, S.S. *et al.* Foundation models, generative AI, and large language models in healthcare: an overview. *Journal of Medical Systems*, v. 49, 2025. doi: 10.1007/s10916-024-02136-1.
- BRASIL. Conselho Federal de Medicina. Resolução CFM nº 2.454, de 11 de fevereiro de 2026. Normatiza o uso da inteligência artificial na medicina. Brasília: CFM, 2026.
- BUESS, L. *et al.* From large language models to multimodal AI: a scoping review on the potential of generative AI in medicine. *arXiv*, 2025. doi: 10.48550/arXiv:2502.09242.
- CHU, B. *et al.* Generative AI's healthcare professional role creep: a cross-sectional evaluation of publicly accessible, customised health-related GPTs. *Frontiers in Public Health*, v. 13, 2025. doi: 10.3389/fpubh.2025.1584348.
- DIB, C.C. A Inteligência artificial na saúde: caminhos para sua regulação. *Revista de Direito Sanitário*, v. 24, 2024. DOI: 10.11606/issn.2316-9044.rdisan.2024.207653.
- DORESWAMY, N. & HORSTMANSHOF, L. Generative AI decision-making attributes in complex health services: a rapid review. *Cureus*, v. 17, e78257, 2025. doi: 10.7759/cureus.78257.
- FAHAD, N. *et al.* Generative AI in clinical (2020–2025): a mini-review of applications, emerging trends, and clinical challenges. *Frontiers in Digital Health*, v. 7, 2025. doi: 10.3389/fdgth.2025.1653369.
- HOWELL, D. Generative artificial intelligence, patient safety and healthcare quality: a review. *BMJ Quality & Safety*, v. 33, 2024. doi: 10.1136/bmjqs-2023-016690.
- HU, D. *et al.* A systematic review of early evidence on generative AI for drafting responses to patient messages. *NPJ Health Systems*, v. 2, 2025. doi: 10.1038/s44401-025-00032-5.
- PERES, F. A literacia em saúde no ChatGPT: explorando o potencial de uso de inteligência artificial para a elaboração de textos acadêmicos. *Ciência & Saúde Coletiva*, v. 29, 2024. doi: 10.1590/1413-81232024291.02412023.
- REDDY, S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Science*, v. 19, 2024. doi: 10.1186/s13012-024-01357-9.
- REUTER, M. *et al.* Generative AI in clinical practice: novel qualitative evidence of risk and responsible use of Google's NotebookLM. *Eye*, v. 39, p. 1650, 2025. doi: 10.1038/s41433-025-03817-y.
- SAEIDNIA, H.R. *et al.* Generative AI and health misinformation: production, propagation, and mitigation: a systematic review. *BMC Public Health*, v. 26, 2026. doi: 10.1186/s12889-025-26148-9.
- SCHWENDICKE, F. *et al.* Generative AI: opportunities, risks, and responsibilities for oral sciences. *Journal of Dental Research*, v. 104, p. 1429, 2025. doi: 10.1177/00220345251356408.
- SHARMA, S. *et al.* A longitudinal analysis of declining medical safety messaging in generative AI models. *npj Digital Medicine*, v. 8, 2025. doi: 10.1038/s41746-025-01943-1.
- SHOKROLLAHI, Y. *et al.* Recent advances in generative AI for healthcare applications. *Journal of Imaging Informatics in Medicine*, 2025. Preprint. doi: 10.1007/s10278-026-01908-0.
- WINECOFF, A. & KLYMAN, K. From symptoms to systems: an expert-guided approach to understanding risks of generative AI for eating disorders. *CHI '26: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 2025. doi: 10.1145/3772318.3791775.
- ZADA, T. *et al.* Medical misinformation in AI-assisted self-diagnosis: development of a method (evalprompt) for analyzing large language models. *JMIR Formative Research*, v. 9, e66207, 2025. doi: 10.2196/66207.