



Artificial intelligence models in primary care: performance, transparency, and safety in patient triage .

Lucas Pedroza Daniel - Physician, graduated in Medicine from the University of Medical Sciences of Cuba and postgraduate in Family and Community Medicine (UFCSPA) - lucasomni@gmail.com

Summary

The incorporation of artificial intelligence (AI) into healthcare systems has advanced significantly in the last decade, reaching non-hospital settings such as primary care and emergency services. This article critically analyzes international experiences (United States, Canada, United Kingdom, and Brazil) involving AI applications in automated triage, risk stratification, and clinical decision support, focusing on low- and medium-complexity contexts. The main risks associated with the use of these technologies are discussed – algorithmic bias, opacity, interoperability failures, weaknesses in data governance, and privacy issues – in light of international regulatory and ethical frameworks proposed by organizations such as the World Health Organization (WHO), the Food and Drug Administration (FDA), and the National Institute for Health and Care Excellence (NICE). Based on this analysis, a set of minimum criteria for the safe and ethical adoption of AI in primary care and emergency services is proposed, including local clinical validation, transparency, bias control, data governance, system integration, team training, and post-implementation monitoring. In conclusion, AI can strengthen primary care, improve patient flow management, and support complex clinical decisions, provided it is implemented under robust clinical governance, with continuous professional supervision and respect for patients' rights.

Keywords: artificial intelligence in healthcare; primary care; clinical governance; medical ethics; automated triage; emergency services.

Abstract

The incorporation of artificial intelligence (AI) into healthcare systems has significantly progressed in recent years, expanding into non-hospital settings such as primary care and emergency services. This article critically analyzes international experiences (United States, Canada, United Kingdom, and Brazil) involving AI applications for automated triage, risk stratification, and clinical decision support, with a focus on low- and medium-complexity healthcare settings. It outlines key risks associated with these technologies—algorithmic bias, opacity, interoperability failures, data governance weaknesses, and privacy issues—in light of international regulatory and ethical frameworks proposed by institutions such as the World Health Organization (WHO), the Food and Drug Administration (FDA), and the National Institute for Health and Care Excellence (NICE). Based on this analysis, the article proposes a set of minimum criteria for the safe and ethical implementation of AI in primary care and emergency contexts, including local clinical validation, transparency, bias control, data governance, systemic integration, staff training, and post-deployment monitoring. It concludes that AI can strengthen primary care, improve patient flow management, and support complex clinical decisions, provided it is implemented under robust clinical governance, with continuous professional oversight and full respect for patients' rights.

Keywords: artificial intelligence in healthcare; primary care; clinical governance; medical ethics; automated triage; emergency care.

1. Introduction

Artificial intelligence (AI) has been identified as one of the main forces transformative in contemporary healthcare, with the potential to increase the efficiency of services, To enhance diagnostic reasoning and strengthen the ability to manage complex systems. Although



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

although early experiences may have focused predominantly on high-risk hospitals.

Given the complexity, there is a growing trend of these technologies migrating towards attention.

Primary healthcare and emergency services, which handle initial contact, handle the largest volume of cases. of care and some of the main organizational bottlenecks in health systems.

In this context, AI applications focused on automated screening and risk stratification. population-based healthcare and clinical decision support are frequently presented as solutions for Organize queues, prioritize serious cases, anticipate decompensations, and guide the work of teams. multidisciplinary. At the same time, these tools introduce new challenges: risk of bias. algorithmic, model opacity, mismatches between the local context and the training data, difficulties in interoperability between information systems and vulnerabilities in privacy and Security of sensitive data.

In light of these risks, international organizations have been consolidating ethical principles and frameworks. Regulatory guidelines to guide the responsible adoption of AI in healthcare. The World Health Organization (WHO) published specific guidelines on ethics and governance in AI applied to health, emphasizing themes such as autonomy, transparency, equity, and accountability. In parallel, agencies regulators such as the Food and Drug Administration (FDA) and the National Institute for Health and Care Excellence (NICE) develops its own frameworks for evaluating software as devices. Medical devices (SaMDs) and other digital health technologies, with an emphasis on clinical evidence, continuous monitoring, and human supervision. It is in this scenario of opportunities and risks that... This raises the discussion about the use of AI in primary care and in low- and medium-level emergency services. complexity.

2. AI applications in risk screening and stratification

2.1 Automated triage in primary and emergency care

AI tools designed for automated sorting have the primary goal of classifying quickly assess the severity of cases and guide referrals, supporting decisions that Traditionally, these models rely solely on the initial clinical assessment. In emergency services, however, models... Machine learning-based tools combine vital signs, reported symptoms, and data. Demographic data can be used to suggest service priorities, with the potential to reduce waiting times. first medical contact and full-time stay in the service.

Studies in American emergency rooms indicate that triage algorithms can to identify, with satisfactory performance, patients at risk of early clinical deterioration, assisting in the decision regarding allocation to observation beds or hospitalization. In some centers, Predictive models are being integrated into already established screening protocols, such as... The Emergency Severity Index (ESI) functions as a kind of ongoing "second opinion" for



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

The nursing team, without replacing their decision-making responsibility.

In other contexts, such as in Canada, there have been experiences with demand forecasting in emergency situations.

Monitoring of hospital patient flows shows that AI can anticipate peak arrival times.

Patients and congestion in critical sectors, allowing adjustments to staff schedules and

Temporary opening of beds. Hospitals that have adopted digital command centers use dashboards.

powered by predictive algorithms to track, near real-time, indicators of

Occupancy rates, waiting times, and risks of decompensation, articulating clinical and logistical decisions.

In the United Kingdom, digital platforms that integrate triage chatbots with the 111 service.

They aim to guide users on the most appropriate entry point — self-care at home,

Consultation in primary care, care in an emergency service, or ambulance dispatch.

Although the results are still heterogeneous and depend on regulation and flow design, there are

evidence suggests that such solutions can reduce unnecessary consultations and reorganize demand when embedded in well-structured systems.

In Brazil, more recent experiences indicate the use of AI in clinical screening.

qualified in the active identification of cases in primary care. The TAMIS project, by

For example, it uses algorithms to analyze electronic medical records from primary care units and flag them.

patients with a higher probability of underdiagnosed rare or chronic diseases, referring them-

for detailed medical evaluation. In this configuration, AI functions as a search engine.

active, expanding the team's ability to identify risks that might otherwise go unnoticed in

Overburdened care routines.

In summary, automated screening has the potential to accelerate the recognition of

severe cases, reduce waiting times in emergency services, and support the early detection of clusters.

risk assessment in primary care. However, its effectiveness depends on the quality and completeness of the data.

that feed the models, from calibration in relation to the local epidemiological profile and capacity

integration with workflows and the infrastructure available in each service.

2.2 Population risk stratification

AI-powered risk stratification utilizes large clinical and administrative databases.

to classify individuals according to the likelihood of adverse outcomes, such as hospitalization,

readmission or mortality. In chronic condition monitoring programs — such as

Diabetes, heart failure, and chronic obstructive pulmonary disease — this approach aims to...

directing more intensive resources (more frequent consultations, home visits,

telemonitoring, pharmaceutical support) for higher-risk users, with a view to reducing

Avoidable complications and associated costs.

Literature reviews and cohort studies indicate that such models may present good results.



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

Accuracy in controlled environments, demonstrating the ability to identify at-risk subgroups.

a substantially larger number of events. In some scenarios, interventions guided by stratification of Risk factors were associated with a reduction in hospital admissions and, in certain groups, with a decrease in hospital admissions. mortality, especially when combined with structured management programs illnesses.

However, the transposition of these models to different contexts or their widespread adoption... Large-scale testing, without careful validation, often results in modest or no effects. In some cases... In some cases, an increase in resource use due to overtriage was observed—that is, when an excessive number of people are triaged. The number of patients is classified as high-risk — either due to difficulties integrating with the workflow. caregiving, which prevents the provision of interventions proportionate to the assigned classification. Stratification by AI, therefore, is not sufficient on its own: it requires that teams have the necessary conditions. concrete ways to offer differentiated responses to prioritized groups, otherwise we risk simply redistributing [resources]. bureaucratic processes, without relevant clinical impact.

Therefore, the use of AI in population risk stratification only translates into gains. Measurable when linked to care models that combine: local validation of algorithms, Clear definition of care pathways for patients identified as high-risk, availability of compatible healthcare resources and continuous monitoring of outcomes achieved.

3. AI in supporting clinical decision-making in primary and emergency care.

Clinical decision support systems (CDSS)

They constitute one of the most established areas for the application of artificial intelligence in medical practice. In general terms, these systems offer security alerts, diagnostic suggestions, reminders of Screening and prevention, as well as therapeutic recommendations based on protocols and guidelines. updated.

In emergency settings, predictive models have been used to signal risk of... sepsis, hemodynamic decompensation, or the need for admission to intensive care, starting from Dynamic analysis of vital signs and laboratory tests in the first hours of care. When Properly calibrated and integrated into the care flow, these systems can anticipate activation of therapeutic bundles or mobilizing rapid response teams, reducing delays in critical interventions.

In primary care, applications based on natural language processing are emerging. being tested for automated analysis of clinical notes. Among the most frequent objectives These include identifying early signs of depression or suicide risk in written records, Suggestion of differential diagnoses for nonspecific complaints and verification of interactions.



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

Complex medication regimens in patients taking multiple medications. Thus, AI tends to assume the role of a systematic "second reading" of data already present in the medical record, expanding the capacity to risk and inconsistency detection.

Despite this potential, there is a consensus that AI should act as a complementary tool and not as a substitute for clinical judgment. The ultimate responsibility for the decision remains with the doctor, doctors and other health professionals, who need to integrate algorithmic recommendations into key elements of clinical practice: physical examination, detailed history, social context, preferences of the patient and their actual ability to adhere to the treatment plan.

Qualitative studies with healthcare professionals reveal a combination of enthusiasm and caution. On one hand, the usefulness of AI in reducing repetitive tasks is recognized, maintaining updated guidelines support diagnostic reasoning in complex cases. On the other hand, they are frequent concerns about the reliability of algorithms, the risk of "alert fatigue" (saturation due to excess of warnings) and the possibility of erosion of professional autonomy if the tools are implemented in a prescriptive and opaque manner. From this perspective, the interface design – frequency and relevance of alerts, clarity of justifications, and possibility of critical review of recommendations – become a central component of the clinical governance of these systems.

4. Patient flow management and systems organization

Beyond the strictly clinical dimension, AI is being incorporated into workflow management and resources in emergency services and healthcare networks. In large hospitals, predictive algorithms are used to estimate the number of admissions in different windows of time, supporting decisions regarding shift schedules, allocation of observation beds, and need of backup beds in other units.

In high-volume emergency care units, forecasting models for "outbreaks of demand" was associated with the reorganization of physical flows and the reallocation of teams between sectors and adjustments to the provision of quick-run tests. In certain experiences, particularly in Canada, initiatives of this type have been linked to a significant reduction in average travel time. Expected, with no documented increase in adverse events.

In primary care, AI-based solutions can support absenteeism prediction in consultations, adjusting schedules according to risk profiles, identifying territories with higher risk levels, probability of seasonal outbreaks and monitoring of continuity of care indicators, such as loss to follow-up in chronic conditions. In resource-limited settings, this type of tool can contribute to using healthcare capacity in a more strategic way.

However, the effective use of AI in organizing workflows depends on certain conditions, minimum structural requirements. Among them, the following stand out: integrated electronic medical records, connectivity



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

adequate data storage and processing capacity and technology teams

information capable of maintaining, updating, and monitoring systems. In middle-income countries, such as

Brazil, the heterogeneity of platforms, the fragmentation of information systems, and the coexistence

Legacy systems represent significant obstacles to the full adoption of these technologies.

especially in large, decentralized public networks.

5. Risks, biases and technical challenges

Algorithmic bias and health inequalities

One of the most debated risks in the literature on AI in healthcare is algorithmic bias. When

The data used to train the models do not adequately represent the diversity of the population.

– for example, through the underrepresentation of racial groups, low-income regions, or high-income contexts.

Social vulnerability – performance tends to be lower precisely for these groups. In health,

This can translate into underdiagnosis, delayed referrals, and overestimation of stability.

clinical or false guarantees of safety.

A classic study by Obermeyer et al. (2019) demonstrated how a widely used algorithm

used in population management overestimated the health of black patients compared to

white people, when using past healthcare costs as a proxy variable for care needs. In

In contexts where vulnerable populations have historically had less access to services, this choice

The methodology reproduces and deepens structural inequalities.

Recognizing this problem, the FDA recommends that developers and healthcare services

incorporate, at all stages of the model lifecycle – from the design of training databases

post-implementation monitoring involves the systematic analysis of biases and impacts on groups.

vulnerable. The WHO, in turn, has emphasized that AI technologies in healthcare should be

designed to promote equity, preventing historically marginalized groups from being

further harmed by the automation of clinical and administrative decisions.

Opacity and explainability

Another key challenge concerns the opacity of many AI models, especially

Those based on deep neural networks and architectures that are difficult to interpret. A

the impossibility, in practice, of understanding how the model arrived at a particular conclusion.

The recommendation limits auditability, hinders the identification of systematic errors, and reduces

The trust of professionals and patients.

In response, there is growing interest in explainable AI (XAI) approaches, which

They seek to offer visual or textual representations of the most relevant factors for each prediction.

such as rankings of variable importance, heat maps, or simplified textual justifications.



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

From an ethical standpoint, recent WHO documents argue that AI technologies in healthcare be accompanied by clear documentation about their general operating logic, evidence of Performance, recommended use cases, and known limitations. This transparency is a condition. so that professionals can judge when to follow, when to relativize, and when to challenge a algorithmic recommendation.

Interoperability, privacy and information security

The quality and reliability of AI models depend directly on the quality of... Input data. Interoperability barriers between electronic health record platforms and systems. Laboratory systems, monitoring devices, and administrative bases can create gaps and inconsistencies compromise the accuracy of forecasts. Standards such as FHIR seek to mitigate this. Some of these obstacles exist, but its adoption is uneven, especially in large-scale public networks.

Furthermore, the intensive use of sensitive data raises significant challenges in terms of Privacy, personal data protection, and cybersecurity. In countries like Brazil, the use of AI in healthcare must adhere to regulatory frameworks such as the General Data Protection Law (LGPD), which It establishes principles of data minimization, legitimate purpose, necessity, and transparency, in addition to... to require clear legal bases for the handling of health information. The specialized literature recommends that AI solutions in healthcare incorporate audit trails and access controls in multiple levels, structured incident response plans, and explicit governance policies. data.

Finally, AI systems are subject to technical flaws and cyberattacks that can directly affecting critical care flows. In this sense, resilience and continuity requirements are essential. Care needs to be considered from the design phase to daily operation, including system redundancy, contingency plans, and mechanisms for safe shutdown of models in case of malfunction.

Clinical governance and international regulatory frameworks

Given the identified risks, the understanding has solidified that AI in healthcare should be... Implemented under strong clinical governance. In 2021, the WHO proposed six general principles to guide it. The development and ethical use of AI in healthcare: protecting human autonomy; promoting well-being. being and security; ensuring transparency and intelligibility; fostering accountability and possibility of appeal; ensuring inclusion and equity; and promoting responsive and sustainable AI. These principles serve as a reference for health systems that wish to incorporate them. AI technologies without relinquishing fundamental rights and minimum criteria for distributive justice.



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

On the regulatory front, the FDA has made progress in building a specific framework for AI and machine learning-based software classified as medical devices (SaMD), including the proposal of a “total product cycle”, with emphasis on post-monitoring, marketing and mechanisms for continuous updating of models. NICE, in turn, developed the Evidence Standards Framework for digital health technologies, classifying them by risk level and establishing gradual requirements for clinical and economic evidence. proportionally to the expected impact on care.

In the European Union, the proposed Regulation of Artificial Intelligence introduces a an explicitly risk-based approach, imposing more stringent requirements for systems. considered "high risk," which includes various medical applications. The European debate highlights the need for extensive testing, robust security documentation, and mechanisms for Human oversight is required for technologies that may affect fundamental rights, including access to... health and non-discrimination.

In Brazil, in addition to the sanitary regulation of medical devices by Anvisa, there are initiatives This is being discussed in bodies such as the National Supplementary Health Agency and in groups. Interministerial committees dedicated to AI regulation are still in the process of consolidation. Projects such as... TAMIS-IA signals possible paths for the gradual and responsible incorporation of technologies. AI in primary care within the Brazilian Unified Health System (SUS) is acceptable, provided it is accompanied by clear governance protocols. Clinical training, team training, and systematic evaluation of clinical and organizational impact.

Technical and ethical criteria for adoption in low and medium complexity contexts.

Based on a review of international experiences and governance guidelines, it is It is possible to outline a set of minimum criteria to guide the adoption of AI in services. primary care and low- and medium-complexity emergency services.

Local clinical validation

AI models should be validated on databases that represent the context in which they are used. will be used. This implies testing performance in primary care units, UPAs or emergency services with patient profiles, epidemiological patterns, and resources distinct from those that originated the algorithm. Metrics such as sensitivity, specificity, predictive values, area under The ROC curve and calibration should be documented and compared with existing alternatives. including usual clinical judgment.

**Transparency and explainability**

Whenever possible, it is recommended that AI systems offer mechanisms for Explainability, such as indicating the most relevant factors for a given forecast or graphs. Importance of variables. Essential information about how the model works and its limitations. The training context and information should be made available to professionals in accessible language. to patients and managers, in order to allow a critical evaluation of the recommendations produced.

Equity and bias control

The adoption of AI should include routines for monitoring bias, evaluating the performance of model in different subgroups (by sex, race/color, age, socioeconomic status, territory, among others) others). When relevant differences are identified, strategies should be implemented to corrections, such as data reweighting, inclusion of underrepresented samples, recalibration segmented or adjustments to decision thresholds, aiming to reduce the reproduction of pre-existing inequalities. existing.

Data governance and privacy protection

Services that adopt AI need to have explicit data governance policies in place. compatible with legal frameworks such as the LGPD and the GDPR, defining legal bases for processing. data, anonymization or pseudonymization procedures, access criteria and Responsibilities in case of incidents. Audit trails should record accesses and modifications. based on the models used, ensuring traceability, accountability, and the ability to... Fault investigation.

Integration into workflows and interoperability

AI systems are only effective when they are integrated into the actual workflow of teams. This means that the technology must connect to the electronic health record and other systems. information, avoiding duplicate records, unnecessary interruptions in service or Overload of administrative tasks. The adoption of interoperability standards, such as FHIR, It contributes to reducing data fragmentation and increasing the reliability of forecasts.

Training, organizational culture, and governance committees.

The implementation of AI must be accompanied by ongoing training programs for healthcare professionals, managers, and IT teams, addressing both technical and ethical aspects and

Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

Organizational. The creation of internal AI governance committees is recommended, composed of Clinical, legal, technology, and user representatives responsible for evaluating projects, Monitor risks, deliberate on incidents, and guide adjustments in implementation.

Post-implementation monitoring

Even after validation and deployment, AI models are subject to what is called *model drift*. That is, the loss of performance over time due to changes in the population served, in epidemiology or in clinical practice. Therefore, it is necessary to continuously monitor indicators of accuracy, record adverse events, periodically review performance, and decide when relevant, due to the updating, reconfiguration, or discontinuation of the models.

Consent, autonomy, and the doctor-patient relationship.

In many situations, especially in public health and low-risk screening, the use of AI It can be based on general legal grounds. However, it is still recommended to state this explicitly. informing patients about the automated technologies being used in their care, explaining Expected benefits, limitations, and privacy protection measures. In high-severity decisions. or involving particularly sensitive data, specific, informed, and renewable consent is required. It contributes to strengthening patient autonomy and trust in the relationship with the healthcare team.

Final considerations

Analysis of the available evidence indicates that artificial intelligence has potential. significant for improving primary care and emergency services, especially in what concerns This refers to triage, risk stratification, and clinical decision support. In high-demand settings. and with limited resources, these technologies can help in the early identification of severe cases, in organizing patient flows, reducing waiting times, and supporting decisions that, from Alternatively, they would depend exclusively on the individual memory and experience of the professionals.

At the same time, the mapped risks – algorithmic bias, opacity, gaps in Interoperability, privacy vulnerabilities, and potential impacts on autonomy. Professionals – demonstrate that AI cannot be treated as a neutral or merely technical solution. Safe and ethical adoption requires robust clinical governance, careful local validation, and vigilance. Continuous performance and active participation of professionals and patients in defining priorities. and limits.

For countries with universal systems like the SUS (Brazilian Unified Health System), AI represents an opportunity to... Strengthen primary care, support emergency networks, and improve the management of scarce resources. This



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

This promise, however, will only be fulfilled if the technology is incorporated with a critical spirit, aligned with principles of equity and social justice, accompanied by consistent investments in Infrastructure, team building, and clear regulatory frameworks.

Far from replacing clinical judgment, AI should be understood as a tool for Expanding care capacity. Used with transparency, responsibility, and a focus on the process. The person being cared for can contribute to more responsive health systems, capable of combining Technological innovation with ethical and clinical commitment in the daily routine of primary care and urgency.

References

WORLD HEALTH ORGANIZATION. *Ethics and governance of artificial intelligence for health*. Geneva: World Health Organization, 2021.

FOOD AND DRUG ADMINISTRATION. *Proposed regulatory framework for modifications to AI/ML-based software as a medical device (SaMD): discussion paper and request for feedback*. Silver Spring: FDA, 2019.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE. *Evidence standards framework for digital health technologies*. London: NICE, 2021.

BRAZIL. Ministry of Health. *TAMIS-IA: National initiative for artificial intelligence in health*. Brasilia: DATASUS, 2023.

PEREIRA, JP; DINIZ, MAA; LIMA, JG. *Artificial intelligence in health: ethical risks and perspectives for the SUS (Brazilian Public Health System)*. *Bioethics Journal*, v. 30, n. 2, p. 264-273, 2022.

SHORTLIFFE, EH; SEPÚLVEDA, MJ *Clinical decision support in the era of artificial intelligence*. *JAMA*, vol. 320, no. 21, p. 2199-2200, 2018.

OBERMEYER, Z. et al. *Dissecting racial bias in an algorithm used to manage the health of populations*. *Science*, vol. 366, n. 6464, p. 447-453, 2019.

TOPOL, E. *Deep medicine: how artificial intelligence can make healthcare human again*. New York: Basic Books, 2019.

RAJKOMAR, A.; DEAN, J.; KOHANE, I. *Machine learning in medicine*. *New England Journal of Medicine*, vol. 380, p. 1347-1358, 2019.

SILVA, A.; CARVALHO, DB; FARIA, LF *Interoperability in health: challenges and perspectives for the adoption of AI in the SUS*. *Cadernos de Saúde Pública*, v. 37, n. 4, 2021.

IBM WATSON HEALTH. *Transparency and trust in AI for health*. Whitepaper. IBM, 2021.



Year III, v.2 2023 | Submission: October 12, 2023 | Accepted: October 14, 2023 | Publication: October 16, 2023

Federal District. Health Secretariat. *TAMIS: Advanced Medical Triage with AI in the SUS (Brazilian Public Health System)*.
Brasilia, 2023.

WORLD ECONOMIC FORUM. *AI governance in healthcare: ethical and legal challenges*.
Geneva: WEF, 2020.

EUROPEAN COMMISSION. *Proposal for a regulation on artificial intelligence*. Brussels: European Union, 2021.