

HEALTH POLICY TRANSPARENCY ON ARTIFICIAL INTELLIGENCE

DOINA-MONICA AGHEORGHIESEI

Alexandru Ioan Cuza University of Iași

Iași, Romania

agheorghiesei.doina@feaa.uaic.ro

Abstract

In the context of profound digital transformations that mark healthcare systems globally, artificial intelligence (AI) has become an essential tool in supporting clinical research, therapeutic decisions and the optimization of medical processes. Although the technological promise of AI is substantial, its implementation generates a series of essential challenges, among which the lack of transparency, the difficulties of explainability of algorithmic decisions and the risk of bias stand out. Transparency is thus emerging not only as a technical requirement, but as an ethical and normative principle, essential for maintaining public trust, protecting patients' rights and ensuring responsible technology governance.

According to the European Commission (2021), the transparency of an AI system implies its ability to provide clear and intelligible explanations of the logic of automated decisions, along with the traceability of decision-making processes and the data used. In this sense, transparency implies not only technological explainability, but also auditability, clearly defined responsibility and the active participation of all actors involved in the life cycle of AI systems: developers, medical professionals and patients.

This study brings together the findings of several recent articles to explore the transparency of artificial intelligence (AI) in public health policies. The analysis highlights both convergences and differences in approach, offering an integrated perspective on how transparency is understood and applied in this field.

Key themes such as accountability, equity and trust in AI systems are outlined, alongside contrasts between technical and holistic views of implementation.

The study proposes a structured synthesis of the current context and suggests future directions for research and policy development in the field of ethical use of artificial intelligence in health.

Keywords: *artificial intelligence; transparency; public health policy; ethics.*

JEL Classification: I18; O33; K32; D83.

1. INTRODUCTION

Artificial intelligence (AI) is playing an increasingly important role in healthcare, helping to improve clinical research, quality of care and patient

outcomes. However, challenges related to bias, lack of transparency and difficulty in interpreting AI models requires rigorous regulation.

This study analyses peer-reviewed articles to highlight common themes and distinct perspectives on transparency in the use of AI, with the aim of informing future strategies. Integrating AI into public health policies brings both benefits and risks, and transparency is becoming a key element for maintaining public trust and respecting ethical principles. They offer an overview of the complexity of transparency, addressing technical, ethical and normative dimensions.

Transparency in the use of artificial intelligence (AI) in public health policy is essential to ensure trust, ethics and efficiency in decision-making. Here are some relevant definitions taken from specialized sources:

1. The European Commission defines transparency in the context of AI as the ability to guarantee that AI systems used in the European Union are "secure, transparent, ethical, impartial and under human control". This involves classifying AI systems according to the level of risk and applying specific rules for each category, thus ensuring the protection of citizens and respect for fundamental rights (Regulation on Artificial Intelligence, (EU) 2024/1689).

2. Samuela Rostas, in the article "The impact of artificial intelligence in decision-making processes in public administration" (2024), emphasizes that transparency in the use of AI in public administration implies clarity and openness in the way these technologies are integrated into decision-making processes, allowing citizens to understand and evaluate decisions made with the help of AI.

3. Navid Toosi Saidy, in the article "Artificial Intelligence (AI) in health: advantages & risks" (n.d.), published on Citizen 47.biz, highlights that transparency in the use of AI in health involves providing clear and accessible explanations about the decision-making processes of algorithms, essential for the acceptance and effective use of these systems in public health.

2. MATERIALS AND METHODS

The present study adopts a theoretical-comparative methodology, aiming at the analysis of a corpus of up-to-date scientific literature, with direct relevance to the issue of AI transparency in the field of public health.

Five academic papers that offer complementary perspectives on the topic are analyzed in detail: the article by He *et al.* (2023), which proposes an ethical and normative vision of explainable AI; the systematic study by Payrovnaziri *et al.* (2020), focused on explainable models applied to real data from electronic medical records; the contribution by Kiseleva *et al.* (2022), who propose a layered framework of accountability and transparency; the article by Hernandez-Boussard *et al.* (2020), which introduces the MINIMAR reporting standards; and the research by Rajpurkar *et al.* (2022), which reviews the applicability of AI in medicine.

These papers are examined to identify common themes, methodological and conceptual differences, and how health policy influences standards of transparency in AI.

By maximizing key ideas and optimizing conclusions, this research synthesizes a comprehensive view that supports ongoing research and policy formulation, based on a larger number of documents, with the most relevant cited.

The analysis reveals a common focus on the need for transparency to ensure fairness, accountability and trust in AI systems, while differing in approaches, focusing on technical versus holistic implementation. This article formulates a structured understanding of the current landscape, offering insights for future directions in the ethical implementation of AI in health.

2.1. What are the main challenges in integrating AI technologies into public health policies?

Integrating artificial intelligence (AI) technologies into public health policy brings with it a number of significant technical and systemic challenges. One of the most pressing difficulties is related to the quality and availability of medical data, which are often fragmented, incomplete or demographically uneven. The lack of interoperability between health systems exacerbates this problem, making it difficult to effectively train and implement AI models (Jiang *et al.*, 2017). Algorithmic models can also reflect or even amplify existing biases in the data, leading to inequitable results for certain population groups (Obermeyer *et al.*, 2019). In these conditions, AI integration cannot be considered purely technological, but requires a systemic approach, adapted to the complexity of the medical field.

In addition to technical challenges, ethical and social aspects play a crucial role in the responsible adoption of AI in public health. Algorithmic transparency, explainability of automated decisions, and privacy protection are essential for building trust among health professionals and the general public (Floridi *et al.*, 2018). At the same time, AI technologies need to be developed in a way that promotes equity and equal access to healthcare, especially for vulnerable or marginalized groups. Without such measures, there is a risk that AI will contribute to reinforcing existing disparities in healthcare systems (Benjamin, 2019). Thus, social and cultural aspects cannot be ignored in the definition and application of AI integration policies. The regulation of AI technologies in healthcare is a field that is constantly evolving, but faces multiple uncertainties. Many of the current legislative frameworks, such as those in the European Union or the United States, do not keep up with the accelerated pace of technological innovation, which creates gray areas in terms of accountability, certification and safety (Morley *et al.*, 2020). Clear global standards that are adaptable to local circumstances are needed to facilitate the effective implementation of AI. At the same time, the lack of investment in digital infrastructure and the training of healthcare professionals

are major obstacles, especially in low- and middle-income countries (Whitelaw *et al.*, 2020). Finally, public acceptance and social legitimacy of AI are essential for the success of technology-based public health policies. Concerns about privacy, job losses, and patient alienation can reduce public trust in automated decisions. Therefore, transparent communication, community engagement, and public education are key elements to foster a culture of trust in AI (Meskó *et al.*, 2017). At the same time, international cooperation is crucial to reduce global disparities in access to technologies and to ensure that AI solutions are culturally and linguistically inclusive. Addressing these challenges in a proactive way will enable the responsible and sustainable integration of AI into public health policies, with real benefits for the quality and equity of health services on a global scale.

2.2. How public health policies influence the development of AI transparency standards

Public health policies play a critical role in shaping artificial intelligence (AI) transparency standards, influencing both the design of algorithmic systems and their regulation. The relationship between public policy and AI is bidirectional: policies establish requirements for explainability and accountability, while technological progress prompts the revision of normative frameworks. For example, Article 22 of the General Data Protection Regulation (GDPR) of the European Union requires “significant information” about the logic of automated decisions, requiring the use of methods such as SHAP or LIME to increase the interpretability of algorithms (Voigt and Von dem Bussche, 2017). In parallel, ethical initiatives, such as the World Health Organization (WHO) guidelines, emphasize the link between transparency and equity, calling for audits to identify bias in datasets (World Health Organization, 2021).

Differences in policy approaches across countries are leading to divergent standards of AI transparency, with significant implications for developers and healthcare institutions. The European Union promotes strict, risk-based regulation that requires detailed technical documentation for high-risk systems. In contrast, agencies such as the FDA in the United States are adopting voluntary guidelines that allow for iterative post-launch surveillance (Floridi *et al.*, 2018). Canada, through its algorithmic impact assessments, emphasizes fair transparency, requiring demographic reporting and bias audits, even at the risk of compromising model interpretability. These strategies, summarized in Table 1, reflect a structural tension between the desire for rigorous regulation and the need to stimulate innovation.

Table 1. Divergences in policy approaches and technical implications

Policy approach	Standard AI outcome	Compensations
Risk-based regulation (EU)	Detailed documentation for high-risk systems	High compliance costs, potential obstacle to innovation
Voluntary Guidelines (US FDA)	Post-market surveillance, iterative updates	More flexibility, but uneven application
Equity focused mandates (Canada)	Demographic and bias audits	Promotes fairness, but complicates interpretability

Source: own computation

However, implementation of transparency policies is often limited by technical constraints, time lags between regulation and technological development, as well as tensions between global standards and local priorities. For example, legal "right to explanation" requirements (such as in the GDPR) favour models that are easy for users to understand, but they can sacrifice algorithmic precision, especially in the case of neural network-based systems. Furthermore, many health policies predate the advance of AI, which creates legal gaps – an example being Ghana's legislation, which does not include mandates on AI transparency in public health. Concomitantly, attempts at global standardization, promoted by the WHO, sometimes conflict with local requirements, such as those in India, where algorithmic explainability must be linguistically and culturally adapted (World Health Organization, 2021). In response to these challenges, emerging transparency standards have emerged, developed in parallel with public policies. Documentation frameworks such as MINIMAR require details of data sources, validation protocols, and failure modes, providing a technical basis for assessing AI responsibility. Tools such as model cards (Mitchell *et al.*, 2019) facilitate reporting of AI performance by subpopulation. At the same time, laws such as California's Automated Decision Systems Accountability Act promote algorithmic impact assessment, and the European IVDR Regulation requires continuous monitoring of AI performance in medical diagnosis. However, the implementation of these initiatives is affected by the ambiguity of policy terms, infrastructure disparities between developed and developing countries, and misalignment between the interests of practitioners, developers and policy makers (Leslie, 2019).

3. RESULTS

The comparative analysis of the revealed literature shows that transparency is universally recognized as an essential condition for the acceptance and responsible use of AI systems in health. The studies of He *et al.* (2023) and Kiseleva *et al.* (2022) converge in arguing that transparency is fundamental for building and maintaining trust between healthcare providers and patients, having

a direct impact on the quality and ethics of the medical act. Also, explainable models (XAI), such as LIME and SHAP, are highlighted by Payrovnaziri *et al.* (2020) as indispensable tools in ensuring the validity and interpretability of AI results. The World Health Organization (2021) similarly advocates the need to introduce AI auditing mechanisms to detect systematic biases and ensure fairness in the delivery of health services.

At the same time, compliance with European regulations, in particular the General Data Protection Regulation (GDPR) and the proposed Artificial Intelligence Act (AI Act), is recognized as essential. Article 22 of the GDPR, which stipulates the right to a meaningful explanation of automated decisions, becomes a fundamental pillar in shaping transparency policies. He *et al.* (2023) and Kiseleva *et al.* (2022) make explicit reference to these legal instruments, underlining the importance of clear and enforceable regulation.

Differences in approach between the reviewed studies are evident in terms of methodology and thematic focus. Thus, He *et al.* (2023) offer a holistic approach, emphasizing the ethical and social dimensions of AI transparency. In contrast, Payrovnaziri *et al.* (2020) adopt a technical and systematic perspective, focusing on the classification and evaluation of XAI methods used in the analysis of data from electronic health records. Kiseleva *et al.* (2022) propose a layered model of transparency, differentiating between external (for patients), internal (for professionals) and own (for developers) transparency, thus providing a comprehensive architecture of responsibilities in AI.

Public policies are proving to play a decisive role in shaping and enforcing transparency standards. The AI Act and GDPR impose clear requirements on the explainability and traceability of AI systems, while WHO guidelines emphasize the correlation of transparency with equity and patient safety. Concrete examples of best practices include South Korea's initiative to publish the performance of AI systems used in pandemic triage, California's legislation on algorithmic auditing, and the provisions of the European IVDR Regulation that require continuous evaluations of the performance of AI in diagnosis.

Despite these advances, significant obstacles persist. Terms such as “meaningful transparency” remain ambiguous, leading to uneven and sometimes inconsistent application of the regulations. Also, the lack of a standardized technical framework for assessing the explainability of AI models makes their comparability across institutions or jurisdictions difficult. Technological infrastructure disparities between developed and developing countries magnify inequalities in the ability to implement these standards.

4. DISCUSSIONS, RECOMMENDATIONS AND PROPOSALS

To ensure an effective and equitable implementation of AI in healthcare, a contextualized approach to transparency is needed. Thus, the level of information and explainability must be adapted according to the category of users: patients

need clear and accessible explanations, medical professionals need relevant technical details, and developers need access to training data and algorithms. Kiseleva's proposed stratified model (Kiseleva *et al.*, 2022) provides a useful framework for this differentiation.

Interdisciplinary collaboration becomes indispensable in the development and application of transparency policies. Informatics, health professionals, ethicists, and policy makers must work together to define realistic and enforceable standards. Initiatives such as IBM AI Fairness 360 (IBM, 2021) demonstrate the usefulness of private-public sector collaborations in developing fair and transparent models.

Another necessary step is the global standardization of transparency assessment tools. Proposals such as MINIMAR and Model Cards, promoted by researchers and technology companies such as Google, can serve as starting points for the development of international guidelines. At the same time, continuous monitoring of AI systems and their post-implementation audit, as required by IVDR, must become common practice.

5. CONCLUSIONS

AI transparency is an essential element for the responsible operation of artificial intelligence systems in public health. Its integration requires a complex approach, which combines technical, ethical and legal dimensions. Public policies play a decisive role, not only by imposing normative requirements, but also by facilitating an organizational culture oriented towards responsibility and equity. For the future, harmonized policies, clear evaluation mechanisms, the participation of all stakeholders and investment in infrastructure and education are needed. Only through such an integrated approach will AI be able to sustainably contribute to the improvement of public health.

References

- 1) Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Cambridge: Polity Press.
- 2) European Commission (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. [online] Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206> [Accessed 01.04.2025].
- 3) Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V. and Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), pp. 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- 4) He, Z., Zhang, R., Diallo, G., Huang, Z. and Glicksberg, B.S. (2023) Editorial: Explainable artificial intelligence for critical healthcare applications. *Frontiers in Artificial Intelligence*, 6, Article 1282800. <https://doi.org/10.3389/frai.2023.1282800>.

- 5) Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J.P.A. and Shah, N.H. (2020). MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *Journal of the American Medical Informatics Association*, 27(12), pp. 2011–2015. <https://doi.org/10.1093/jamia/ocaa088>.
- 6) IBM (2021). *AI Fairness 360 toolkit*. [online] Available at: <https://aif360.mybluemix.net> [Accessed 01.04.2025].
- 7) Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S. and Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. <https://doi.org/10.1136/svn-2017-000101>.
- 8) Kiseleva, A., Kotzinos, D. and De Hert, P. (2022). Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations. *Frontiers in Artificial Intelligence*, 5, Article 879603. <https://doi.org/10.3389/frai.2022.879603>
- 9) Leslie, D. (2019). *Understanding artificial intelligence ethics and safety*. The Alan Turing Institute. [online] Available at: <https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety> [Accessed 10.04.2025].
- 10) Meskó, B., Drobni, Z., Bényei, É., Gergely, B. and Györffy, Z. (2017). *Digital health is a cultural transformation of traditional healthcare*. *mHealth*, 3, Article 38. <https://doi.org/10.21037/mhealth.2017.08.07>
- 11) Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19)*, pp. 220–229. <https://doi.org/10.1145/3287560.3287596>.
- 12) Morley, J., Machado, C.C., Burr, C., Cows, J., Taddeo, M. and Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, 260, article 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>.
- 13) Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), pp. 447–453. <https://doi.org/10.1126/science.aax2342>.
- 14) Payrovnaziri, S.N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J.H., Liu, X. and He, Z. (2020) Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *Journal of the American Medical Informatics Association*, 27(7), pp. 1173-1185. <https://doi.org/10.1093/jamia/ocaa053>.
- 15) Rajpurkar, P., Chen, E., Banerjee, O. and Topol, E.J. (2022). AI in health and medicine. *Nature Medicine*, 28(1), pp. 31–38. <https://doi.org/10.1038/s41591-021-01614-0>.
- 16) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828 (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689 (12 July 2024).

- 17) Rostas, S. (2024). The impact of artificial intelligence in decision-making processes in public administration. *Revista română pentru protecția și securitatea datelor cu caracter personal*, 5(1-2), pp. 24–37.
- 18) Saidy, N.T. (n.d.). Artificial Intelligence (AI) in health: advantages & risks. *Citizen47.biz*. [online] Available at: <https://citizen47.biz/inteligenta-artificiala-ai-sanatate-avantaje-riscuri/> [Accessed 23.04.2025].
- 19) Voigt, P. and Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-57959-7>
- 20) Whitelaw, S., Mamas, A.M., Topol, E. and Van Spall, H.G.C. (2020). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health*, 2(8), e435–e440. [https://doi.org/10.1016/S2589-7500\(20\)30142-4](https://doi.org/10.1016/S2589-7500(20)30142-4).
- 21) World Health Organization (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization. [online] Available at: <https://www.who.int/publications/i/item/9789240029200> [Accessed 27.04.2025].