# Human in the loop requirement and AI healthcare applications in low-resource settings: A narrative review

**F Kabata,** LLD,[1]; **D Thaldar,** PhD,[1,2]

[1] *School of Law, University of KwaZulu-Natal, Durban, South Africa*
[2] *Petrie-Flom Center for Health Law Policy, and Bioethics, Harvard Law School, Harvard University, Cambridge, USA*

***Corresponding author:*** *F Kabata (faithkabata@gmail.com)*

**Background.** Artificial intelligence (AI) applications in healthcare provision have the potential to universalise access to the right to health, particularly in low-resource settings such as rural and remote regions in which AI is deployed to fill in medical expertise gaps. However, a dominant theme in evolving regulatory approaches is the human in the loop (HITL) requirement in AI healthcare applications to ensure the safety and protection of human rights.
**Objective**. To review HITL requirements in AI healthcare applications and inform how best to regulate AI applications in low-resource settings.
**Method.** We conducted a narrative review on HITL requirements in AI healthcare applications to assess its practicality in low-resource settings.
**Results**. HITL requirements in low-resource settings are impractical as AI applications are deployed to fill in gaps of insufficient medical experts.
**Conclusion.** There is a need for a shift in regulatory approaches from primarily risk-based to an approach that supports the accessibility of AI healthcare applications in low-resource settings. An approach anchored on the human right to science ensures both the safety requirements and access to the benefits of AI systems in healthcare provision.
**Key words.** Human in the loop, low-resource settings, AI healthcare, human right to science, safety, accessibility to the right to health.

*S Afr J Bioethics Law* 2024;17(2):e1975. https://doi.org/10.7196/SAJBL.2024.v17i2.1975

The application of artificial intelligence (AI) in healthcare has shown potential to universalise the right to the highest attainable standard of health (right to health) by improving access, particularly in low-resource settings such as rural and isolated areas where specialised expertise is unavailable.[1] Illustratively, in Africa, the AI application *ubenwa* has been used in rural and remote areas of Nigeria to detect birth asphyxia in the absence of medical experts.[2] In India, malaria detection algorithms have been deployed for clinical diagnosis and treatment in remote and hard-to-reach regions with high malaria incidences and a shortage of specialised microscopists and pathologists.[3] However, these benefits of universal access, efficiency and cost savings in healthcare provision are largely clouded by the regulatory frameworks requiring human oversight.

Emerging international regulations for AI, such as the 2019 European Commission Ethics Guidelines for Trustworthy AI (EC Ethics Guidelines),[4] the 2021 UNESCO Recommendation on the Ethics of AI (UNESCO Recommendation)[1] and the 2023 EU Artificial Intelligence Act (EU AI Act),[6] emphasise human oversight as a key requirement in AI development and deployment. In healthcare, the EU AI Act explicitly requires human oversight by natural persons for all high-risk AI systems including medical applications.[6] The UNESCO Recommendation endorses human oversight and determination in healthcare applications, stating that 'life and death decisions should not be ceded to AI'[5] and recommending that final decisions on diagnosis and treatment should be made by human persons.[5]

The main research objective is to review human in the loop (HITL) requirements in AI healthcare applications to determine the best regulatory approach for AI health applications in low-resource settings. To address the research objective, we answered three questions: Why HITL? Who should be the human agent in HITL? What factors support or limit HITL? The study uses the term low-resource settings narrowly, referring specifically to the inadequacy of medical experts for individual and public health intervention, as identified by Zyl *et al.*[7] in their review of low-resource settings in healthcare-related inequities. This narrow application is justified as the study focuses on AI healthcare applications that fill the gaps of unavailable medical experts.

The key contribution of this article is a review of HITL requirements in AI healthcare applications and a discussion on the practicality of HITL requirements in low-resource settings, proposing an alternative regulatory framework.

## Models of human oversight in AI applications

The EC Ethics Guidelines describe the three models of human oversight in AI systems: human in the loop, human on the loop and human in command.[4] In the HITL model, human persons interact in all the steps, hence assuming control of every decision throughout the AI system's life cycle.[4] In essence, human persons can intervene in the decision cycle where necessary.[8] The human on the loop (HOTL) model envisages lesser human control, and limits intervention by the human person to the design and monitoring cycles of the AI system.[4] Human in command (HIC) envisages the ability of the human to control the overall activity of the system, including deciding when and how to deploy it.[4] A fourth model, the human out of the loop, closes out human intervention owing to lack of expertise, skills or the inability to effectively respond to time-critical operations.[4]

## Methods

We applied narrative review methodology to conduct a review of the current status of HITL requirements and inform how best to regulate AI healthcare applications in low-resource settings. A narrative review is a form of literature review aimed at conducting a subjective examination and critique of existing literature.[9] Narrative reviews describe what is known about a topic and provide new insights into the current state of knowledge when viewed from a different perspective.[9] Narrative reviews are thus useful in providing new insights into existing literature.[9] The narrative review methodology is appropriate for this article as it describes the current state of knowledge on HITL requirements in AI healthcare applications and explores its practicality from the perspective of low-resource settings, thus generating new insights.

### Data sources and search strategy

The study retrieved articles from two databases, Google Scholar and Medline via PubMed, as well as through internet browsing using the Google search engine. The keywords and search strings used were: ('human in the loop' AND 'AI healthcare' OR 'AI medical care' OR 'AI clinical care'), 'human in the loop'. The survey included recent articles published between 2019 and 2024 that were in English and open access.

## Results

The article addresses three questions that provide an analysis of HILT requirements in AI healthcare applications: (*i*) Why HITL? (*ii*) Who should be the human agent in HITL? (*iii*) What factors support or limit HILT? By answering these questions, the article provides knowledge on HITL against which its practicality in low-resource settings is assessed.

### Why HILT?

We analysed the literature on AI applications in healthcare and synthesised the purposes of HITL. Most of the literature identified safety and accuracy and upholding human values as the central purposes of HILT requirements in healthcare applications.[10-13] From considerations of safety, accuracy and upholding human values, we refined four sub-purposes of HILT: correcting errors and biases, ethical considerations, interpretation and explainability and accountability. An essential element of the right to health is quality, which refers to scientifically and medically safe health services and products.[14] In AI healthcare, issues of quality arise from data and algorithmic aspects that affect the accuracy and correctness of AI decisions.[15] Data aspects relate to the features of the data, such as incompleteness, unrepresentativeness, and errors in measurements or labelling of data.[16] Algorithmic aspects refer to human or data biases embedded in the design of AI applications.[15,16] Human biases, whether intentional or unintentional, reflect the moral and value biases of the AI designer. Data biases stem from data deficiencies such as the unrepresentativeness of the training dataset.[15] Gerybaite *et al.*[16] argued that while regulatory standards such as the EU AI Act prescribe technical standards to ensure data quality, HITL is the 'last resort' to ensure the safety and accuracy of AI health applications. In this context, the purpose of HILT is to correct errors and biases. Similarly, Bodén *et al.*[11] demonstrated in their study on digital pathology that HILT addressed errors that resulted from the use of poor-quality slide images.

On ethical considerations, the principles of autonomy, nonmaleficence, beneficence and justice are accepted values of medical ethics. These principles mirror the WHO ethical principles for the development of AI in healthcare, which include protecting human autonomy, human well-being and safety, responsibility and inclusiveness.[17] Autonomy refers to an individual patient's right to self-determination, allowing them to make free choices regarding their treatment options. Savulescu *et al.*[18] described AI health applications that prioritised treatment options based on the value system of the designer and argued that such inbuilt prioritisation offended patient autonomy. While noting that AI applications can be designed to incorporate patient values, they acknowledged that HILT ensures respect for autonomy by translating AI treatment recommendations into natural language and letting the patient participate in and consent to the treatment options.[18] Beneficence requires that actions and decisions should safeguard the best interest of the patient, hence ensuring their well-being and safety. Lederman *et al.*[19] pointed out that natural language processing models, when deployed in real-world clinical settings, tend to simplify medical issues to binary questions and outputs. This approach often fails to link these outputs with other medical conditions. This can occasionally cause harm to the patient. According to Maadi *et al.*,[8] such scenarios represent novel situations not captured by AI training datasets. HITL addresses this gap by providing context and knowledge for scenarios not captured by datasets.

In relation to interpretation and explanation, interpretability in AI systems is defined as the understanding of the internal operations of the AI algorithm, while explainability involves reconstructing and explaining why the algorithm arrived at a given decision in human-understandable terms.[20,21] The inherent tension between predictive accuracy and explainability implies that AI systems with higher predictive accuracy are less interpretable.[8,21] This poses a dilemma regarding whether AI algorithms should be constrained as a trade-off for human interpretability. In the context of healthcare, it raises additional concerns about the quality of medical care, if less accurate AI systems were to be used to ease understanding.[18,22] Tied to this is the right to explanation, which encompasses the right to receive an explanation of the algorithm's output decisions, particularly when the decisions significantly impact the individual. HITL provides the necessary human interaction to fulfil this right, ensuring that patients can make informed decisions based on sufficient information.[8,21]

Accountability refers to the attribution of responsibility and liability for AI decisions, raising the question of who should assume wrongdoing for a wrong diagnosis or treatment recommendation. If AI healthcare applications are similar to other medical devices or drugs, should the medical expert not be solely responsible for evaluating the algorithm's performance, explaining its benefits to the patient and conveying the confidence level of the algorithm?[18] Alternatively, given the involvement of different actors, such as doctors, designers of the AI, medical institutions where the AI is deployed and the AI algorithm itself, should liability extend beyond the doctor to include these other stakeholders.[15] Jarrahi *et al.*[12] argued that expert in the loop systems promote accountability by enabling the medical expert to assume overall responsibility. This approach allows the medical expert to consider all relevant parameters and contextual medical information to verify, accept or reject the AI application's decision.

### Who should be the human agent in HITL?

The issue concerns the intervention capabilities that the human agent in HITL should possess. The EU AI Act in article 14 specifically mandates human oversight by natural persons when medical applications are in use.[6] HITL models enable researchers, domain experts, data scientists and ordinary people to be the human agent. In addition, article 14 requires that the human agent should be trained, competent and qualified to be in the loop.[6] Further, it stipulates that the human agent should be able to interpret, accept, reject, disregard, override or reverse the AI system output decisions.[6] Diyasena *et al*.[13] identified doctor in the loop HITL models, which incorporate a domain expert as the human agent who holds the main authority over the AI system. The authors pointed out that the doctor-in-the-loop model resulted in improved healthcare outcomes by addressing novel situations not captured in the training datasets, increasing social acceptance and patient approval.[13] Notwithstanding, this model raised concerns about data security as domain experts rather than designers have overall control of the AI system.[13] Maadi *et al*.[8] reported that the appropriate human agent in HITL should depend on the nature and complexity of the task. For professional and complicated tasks, the human agent should have a higher level of domain expertise, while for less complicated and non-professional tasks such as identifying objects, basic human abilities suffice. Accordingly, for medical applications, the human agent should be medical experts, to ensure improved quality, safety and accuracy.[9]

### What factors support or limit HITL?

The question concerns whether human agents can effectively exercise overall control over AI systems designed to work autonomously and what is needed for this control to be effective. Hille *et al*.[23] pointed out that HITL does not inherently guarantee effective human control, as the presence of a human agent does not necessarily indicate the extent of control exercised or whether the human agent is enabled to exert control. In addition, human control may be impeded by the human agent's inability to understand and respond to the AI system.[24] Haselager *et al*.[24] also argued that HITL does not guarantee effective oversight owing to human reasons, such as concentration deficits and routine boredom, which may result in automation bias. They proposed a reflection machine concept, which avails to the human agent data that supports an alternative output thus requiring the human agent to reflect on the AI decision.[24] Bashkirova and Krpan[25] highlighted confirmation bias by demonstrating that mental health practitioners with higher levels of expertise were more inclined to accept AI recommendations that aligned with their own beliefs and knowledge.

The factors supporting or limiting HITL are; the capacity of the human agent, information and options available to the human agent, time to exercise control and automation and confirmation biases. On the capacity of the human agent, exercising effective human control entails the ability to understand the AI system and respond, including the capacity to override its output decision. The human agent must have the requisite knowledge, qualifications and skills. Hille *et al*.[23] noted that studies on meaningful human control in health systems designated medical experts as the controllers of control, confirming the need for the human agent to have the requisite capacity to exercise effective human control. According to Haselager *et al*.,[24] in the reflection machine concept, the human agent is confronted with data supporting alternative decisions to those recommended by the AI system, thus the human agent must have the necessary competence to consider and deliberate on the alternatives. Besides human factors, the design of the AI system also supports or limits HITL. The information and options available to the human agent are determined by the design of the AI system and the human control measures it supports. Hille *et al*.[23] observed that designers of AI systems in healthcare are enablers of meaningful human control and argued that the incorporation of certification and validation mechanisms creates avenues for the exercise of human control. Similarly, on time to exercise control, AI systems should be designed to allow adequate time for the exercise of human control, including mechanisms such as slow AI systems.[23]

## Discussion

This review revealed that the main purpose of HITL in AI healthcare systems is to ensure safety and accuracy and to protect human rights and values. In addition, the human agent in HITL should be a medical expert with the requisite competence, knowledge and capacity for effective human oversight. These findings have implications for HITL regulatory requirements, particularly in low-resource settings where AI fills gaps rather than complements medical experts. In such scenarios, it may be impractical to fully implement HITL as a regulatory requirement owing to the lack of medical experts. The right to health encompasses both elements of quality and accessibility. Quality requires that healthcare products and services are scientifically approved and medically safe, whereas accessibility requires that healthcare services, goods and technology must be availed to all segments of the population, including marginalised groups and rural populations.[15] The right to health does not therefore envisage a binary choice between the safety and accuracy of healthcare products and services and access to them, including technology. Rather, both elements must be present.

We propose an alternative AI healthcare regulation framework that shifts from a strictly risk-based model that subordinates AI usage and benefits to potential risks, to a model that favours usage and the benefits of AI while also addressing risk. Roberts *et al*.,[26] while comparing fairness as an ethical value in AI healthcare regulation in China and the European Union (EU), alluded to the differing regulation approaches adopted. The EU approach focuses on controlling the usage of AI in healthcare to prevent and minimise threats to patient safety.[26] On the contrary, China's approach is focused on promoting the usage of AI in healthcare to increase access to healthcare, especially in rural areas where medical expertise is inadequate.[26] Drawing from China's approach, we propose a human rights model anchored on the human right to science.

The human right to science guarantees the right of everyone to enjoy and benefit from the progress in science and its applications and the freedom to scientific research.[27] In relation to AI healthcare applications, states have a core obligation to ensure access to scientific applications, particularly when those applications are key to the enjoyment of economic, social and cultural rights.[14] Accordingly, states should eliminate laws, policies and practices that unjustifiably deny access and also establish a legal and policy framework for legal remedies for any harm occasioned.[14] The obligation to ensure the access also includes an implicit obligation to ensure the safety

and quality of the scientific applications accessible to the public by requiring states to certify and regulate such applications.[14] The human right to science regulatory approach thus addresses the shortcomings of HITL in low-resource settings by ensuring human oversight even when medical experts are absent. This approach shifts the responsibility for human oversight from the end user – medical experts – to the state, which is obligated to certify and regulate the use of AI healthcare applications before their deployment for use. This ensures the safety and accuracy of AI healthcare applications, thereby maintaining the quality of healthcare services and products without impeding accessibility. Shifting oversight from medical experts to public institutions aligns with proposals for broadening the range of actors capable of exercising human oversight.[5,23]

## Conclusion

The study explored HITL as a regulatory requirement for AI healthcare applications. While it demonstrated its value in ensuring the safety, accuracy and protection of human values, it also showed the impracticality of HITL in low-resource settings where AI healthcare applications fill in gaps rather than complement existing expertise. We proposed an alternative regulatory approach based on the human right to science, which balances the safety and accessibility of AI healthcare applications.

**Declaration.** None.
**Acknowledgements.** None.
**Author contributions.** Both authors contributed equally to the article.
**Funding.** Work on this article was supported by the U.S. National Institute of Mental Health and the U.S. National Institutes of Health (award number U01MH127690) under the Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) program. The content of this article is solely our responsibility and does not necessarily represent the official views of the U.S. National Institute of Mental Health or the U.S. National Institutes of Health.
**Data availability statement.** The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.
**Conflicts of interest.** None.

1. Alami H, Rivard L, Lehoux P, et al. Artificial intelligence in health care: Laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. Glob Health 2020;16(52):2. https://doi.org/10.1186/s12992-020-00584-1
2. Masso A, Chukwu M, Calzati S. (Non) negotiable spaces of algorithmic governance: Perceptions of the Ubenwa health app as a relocated solution. New Media Soc 2022;24 (4):845-865. https://doi.org/10.1177/14614448221079027
3. Nema S, Rahi M, Sharma A, Bharti, PK. Strengthening malaria microscopy using artificial intelligence-based approaches in India. Lancet Reg Health Southeast Asia 2022;5(10054):2. https://doi.org/10.1016/j.lansea.2022.100054
4. European Commission. Ethics guidelines for trustworthy AI: 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 29th January 2024).
5. UNESCO. Recommendation on the ethics of artificial intelligence: UNESCO 2021, para 35 & 36. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (accessed 18 April 2024).
6. European Union. Proposal for AI Act: EU 2023, Art. 14 and Art. 7. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF (accessed 8 February 2024).
7. Van Zyl C, Badenhorst M, Hanekom S, Heine, M. Unravelling 'low-resource settings': A scoping review with qualitative content analysis. BMJ Glob Health 2021;6(e005190):1-14.
8. Maadi M, Akbarzadeh KH, Aickelin U. A review on human-AI interaction in machine learning and insights for medical applications. Int J Environ Res Public Health 2021;18(2121):3. https://doi.org/10.3390/ijerph18042121
9. Sukhera J. Narrative reviews: Flexible, rigorous and practical. J Grad Med Educ 2022;14(4):414. http://dx.doi.org/10.4300/JGME-D-22-00480.1
10. Sezgin E. Artificial intelligence in healthcare: Complementing, not replacing doctors and healthcare providers. Dig Health 2023;9:1-5.
11. Bodén ACS, Molin J, Garvin S, West RA, Lundström C, Treanor D. The human-in-the-loop: An evaluation of pathologists' interaction with artificial intelligence in clinical practice. Histopath 2021;79:210-218.
12. Jarrahi MJ, Davoudi V, Haeri M. The key to an effective AI-powered digital pathology: Establishing a symbiotic workflow between pathologists and machine. J Path Inform 2022;13(100156):1-4. http://dx.doi.org/10.1016/j.jpi.2022.100156
13. Diyasena D, Arambepola N, Munasinghe L. Effectiveness of human-in-the-loop design concept for eHealth systems. PACIS 2022. Proceedings 2022;191:1-9.
14. United Nations. General comment No. 14 (2000) the right to the highest attainable standard of health (article 12 of the International Covenant on Economic, Social and Cultural Rights): Committee on Economic, Social and Cultural Rights 2000, para 12. https://digitallibrary.un.org/record/425041?ln=en&v=pdf (accessed 23 April 2024).
15. Zhang J, Zhang Z. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decision Making 2023;23(7):1-15. https://doi.org/10.1186/s12911-023-02103-9
16. Gerybaite A, Palmieri S, Vign F. Equality in healthcare AI: Did anyone mention data quality? BioLaw J Rivista di BioDiritto 2022;4:385-409.
17. World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models: WHO 2024. https://iris.who.int/bitstream/handle/10665/375579/9789240084759-eng.pdf?sequence=1&isAllowed=y (accessed 18 April 2024).
18. Savalescu J, Giubilini A, Vandersluis R, Mishra A. Ethics of artificial intelligence in medicine. Singapore Med J 2024;65:150-158. https://10.4103/singaporemedj.SMJ-2023-279
19. Lederman A, Lederman R & Verspoor K. Tasks as needs: Reframing the paradigm of clinical natural language processing research for real-world decision support. J Am Med Informatics Assoc 2022;29(10):1810-1817. https://doi.org/10.1093/jamia/ocac121.
20. Rajabi E, Kafaie S. Knowledge graphs and explainable AI in healthcare. Info 2022;13(459):1-10. https://doi.org/10.3390/.
21. Prentzas N, Kakas A, Pattichis CS. Explainable AI applications in the medical domain: A systematic review. 2023. https://doi.org/10.48550/arXiv.2308.05411
22. Gonzàlez-Alday R, Garzía-Cuesta E, Kulikowski CA, Maojo V. A scoping review of the progress, applicability, and future of explainable artificial intelligence in medicine. Appl Sci 2023;13(19):1-23. https://doi.org/10.3390/app131910778.
23. Hille M, Hummel P, Braun M. Meaningful human control over AI for health? A review. J Med Ethics 2023:1-9. https://doi:10.1136/jme-2023-109095
24. Haselager P, Schcraffenberger H, Thill, et al. Reflection machines: Supporting effective human oversight over medical decision support systems. Cambridge Quart Health Ethics 2023:1-10. https://doi.org/10.1017/S0963180122000718
25. Bashkirova A, Krpan, D. Confirmation bias in AI-assisted decision-making: AI triage recommendations congruent with expert judgment increase psychologist trust and recommendation acceptance. Computers in Human Behavior: Artificial Humans 2024;2(100066):1-14. https://doi.org/10.1016/j.chbah.2024.100066
26. Roberts H, Cowls J, Hine E et al. Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. Info Soc 2023;39(2):79-97. https://doi.org/10.1080/01972243.2022.2124565
27. UN Economic and Social Council, Committee on Economic, Social and Cultural Rights, General Comment No. 25 on Science and Economic, Social and Cultural Rights, articles 15 (1) (b), (2), (3) & (4) on the International Covenant on Economic, Social and Cultural Rights. E/C.12/GC/25. 2020. https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2020-article-15-science-and (accessed 13th February 2024).

*Received 23 February 2024. Accepted 2 July 2024.*