Human-Centered Design for Accessible and Sustainable XAI in Healthcare

Giovanni Arras[1] Tommaso Turchi[1] Giuseppe Prencipe[1] and Giuseppina Sgandurra[2,3]

[1] *Department of Computer Science University of Pisa, Italy*
[2] *Department of Clinical and Experimental Medicine, University of Pisa, Italy*
[3] *Department of Developmental Neuroscience, IRCCS Fondazione Stella Maris, Italy*

Abstract

Artificial Intelligence (AI) is reshaping healthcare by offering innovative tools for clinical decision-making, personalized treatments, precision diagnostics, and efficient resource management. However, the adoption of AI poses challenges related to transparency, informed consent, and socio-economic and ethical sustainability. To address these issues, we present ongoing work on a general framework driven by Human-Centered Design (HCD) and integrating Accessible Explainable AI and sustainability. This framework aims to foster trust, acceptance, and equity by providing inclusive explanations tailored to diverse stakeholders (clinicians, patients, caregivers, and ethics committees), actively involving stakeholders through HCD methodologies, and designing accessible, sustainable AI solutions that address ethical considerations and promote equity.

Keywords

Artificial Intelligence, Accessible XAI, Human-Centered Design, Sustainability, Pediatric Rehabilitation, Co-Design, Equity, Digital Health

1. Introduction

Artificial Intelligence (AI) is rapidly transforming healthcare, offering advanced tools for clinical decision support, personalized treatments, and efficient resource management. AI-based algorithms are already demonstrating remarkable capabilities in various areas, such as analyzing medical images for faster and more accurate diagnoses, optimizing treatment plans in oncology, and predicting patient outcomes based on real-time data [4, 7]. These advances hold immense potential to revolutionize clinical workflows and significantly improve patient care. The urgency of addressing AI challenges in healthcare becomes particularly evident when we consider the critical need for transparency and trust. While AI-based algorithms demonstrate remarkable capabilities in areas like medical imaging and treatment optimization, their potential remains constrained by fundamental concerns about comprehensibility and user acceptance. The opacity of these systems poses a significant barrier, potentially undermining the very technological advances that promise to revolutionize clinical workflows and patient care. [5].
However, the integration of AI in healthcare also presents significant challenges, particularly regarding transparency, informed consent, and both ethical and economic sustainability. Human-Centered Design (HCD) offers a powerful approach to address these issues. By actively involving stakeholders such as clinicians, patients, ethics committees, and technical experts in the design of AI systems, HCD ensures that these technologies are tailored to end-user needs and promote transparency, usability, and trust [10, 11]. The lack of transparency in AI algorithms poses substantial problems for both patients and clinicians. Patients may hesitate to accept treatment

[1]Joint Proceedings of the ACM IUI Workshops 2025, March 24-27, 2025, Cagliari, Italy
giovanni.arras@phd.unipi.it (G. Arras); tommaso.turchi@unipi.it (T. Turchi); giuseppe.prencipe@unipi.it (G. Prencipe); giuseppina.sgandurra@unipi.it (G. Sgandurra)

recommendations when the reasoning behind them remains unclear, while clinicians may be reluctant to rely on AI-driven insights if they cannot understand the underlying logic. This opacity risks hindering the widespread adoption of potentially beneficial AI solutions [3, 4].

Accessible Explainable AI emerges as a key element to address these challenges [1, 2, 3]. By ensuring that algorithmic decisions are understandable to diverse user groups, Accessible XAI fosters trust and facilitates informed decision-making. Techniques such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) provide valuable methods for elucidating model mechanisms [8, 9]. However, to be truly effective in clinical practice, these methods often require customization to meet the specific needs and literacy levels of different users, including clinicians, patients, and families. Tailored explanations not only build trust but also contribute to sustainability by ensuring that diverse user groups, including those with limited health literacy, can access and benefit from AI technologies.

Sustainability complements these efforts by addressing economic, social, and environmental dimensions. For instance, equitable access to AI technologies requires mitigating healthcare disparities [14, 15], implementing multilingual interfaces [16, 17], and adopting culturally sensitive design practices [18, 19]. Promoting the long-term cost-effectiveness of AI solutions necessitates minimizing the environmental impact of computational processes, optimizing resource use, and establishing partnerships that support ongoing updates and maintenance [19].

To achieve this level of customization and inclusivity, a Human-Centered approach is crucial. Human-Centred Design (HCD), through methodologies like co-design and meta-design, emphasizes collaborative and participatory approaches, which are especially critical in healthcare contexts like pediatric telerehabilitation [12, 13].

This paper discusses ongoing work to define a comprehensive framework that integrates Accessible XAI, HCD, and sustainability to holistically address these challenges in AI-based healthcare. Although applicable to various healthcare settings, the framework is particularly relevant for pediatric rehabilitation, where co-design with families, children, and clinicians is instrumental in developing transparent, user-friendly, and sustainable AI solutions. The next sections detail the ethical, practical, and socio-economic barriers to AI adoption and illustrate how the proposed framework can help overcome them. by emphasizing the importance of designing AI solutions that are both user-friendly and ethically grounded. Building on recent insights [20], the framework seeks to extend co-design principles with a specific focus on pediatric telerehabilitation, while also incorporating sustainability considerations (economic, social, environmental) into the AI development lifecycle.

2. Context and Challenges

The integration of AI in healthcare presents a range of challenges that span ethical, practical, and socio-economic domains. These challenges must be carefully addressed to ensure that AI systems are not only effective but also equitable and widely adopted. This section explores these critical barriers and contextualizes their implications for the successful implementation of AI technologies in healthcare. Addressing these challenges proactively through a Human-Centered Design approach can contribute to the development of more ethical, equitable, and sustainable AI systems.

2.1. Ethical Considerations

The integration of AI in healthcare raises several ethical concerns, particularly the potential for bias and discrimination. AI algorithms, trained on large datasets, may inadvertently reflect and perpetuate societal biases present in the data, leading to unequal access to care, misdiagnoses, or inappropriate treatment recommendations [15, 16]. In pediatric rehabilitation, ensuring fairness and avoiding bias is essential, given the wide variability in developmental needs. Another ethical challenge is safeguarding patient autonomy and ensuring informed consent. In pediatric settings, children may have limited understanding of AI systems, so parents or guardians must be adequately informed about AI-driven interventions. Clear and accessible communication is vital to enable guardians to make decisions prioritizing the child's well-being [8, 11]. Co-design practices can facilitate the development of mechanisms for ensuring informed consent and addressing ethical concerns in collaboration with ethicists, clinicians, and families.

2.2. Practical Challenges

Integrating AI into clinical workflows introduces practical barriers, such as clinician training and system interoperability. Healthcare professionals need support to effectively use AI tools, interpret outputs, and integrate them into decision-making [7, 10]. Ensuring interoperability with existing electronic health records and healthcare infrastructures adds complexity.
Continuous monitoring and evaluation of AI systems are also crucial. As models evolve and encounter new data, their performance must be regularly assessed for biases or drops in accuracy. This is particularly important in pediatric telerehabilitation, where long-term impacts of AI tools are not yet fully understood [8].

2.3. Socio-Economic Barriers

The development and deployment of AI technologies can be costly, potentially limiting access and exacerbating health disparities. In pediatric rehabilitation, where families already face financial burdens, affordability is crucial [12, 13]. Collaboration among policymakers, healthcare providers, and tech developers is needed to plan equitable distribution of AI benefits, avoiding new forms of inequality. Public-private partnerships and subsidized AI solutions can help ensure broader accessibility [10, 11]. By involving diverse stakeholders, including underserved communities, in the design process, participatory methods can help identify and address socio-economic barriers to access from the outset.

3. Our Proposed Framework

The proposed operational framework establishes Human-Centered Design (HCD) as its foundational element: it informs the development of Accessible XAI Guidelines, which in turn contribute to achieving Sustainable AI in Healthcare. This framework provides a systematic approach to designing inclusive, effective, and socially responsible AI solutions.

3.1. Main Objectives

The framework pursues three interconnected objectives:

1. Driving trust and transparency through HCD enhances AI system functionality by ensuring explanations are both comprehensible and actionable. This transparent communication enables clinicians, patients, and families to make informed decisions [6, 8].
2. Fostering equity and accessibility through participatory co-design helps identify and address barriers to adoption, such as health literacy or cultural and linguistic differences. The framework develops tailored solutions to these challenges, ensuring equitable distribution of AI benefits [10, 11].
3. Promoting sustainability by involving diverse stakeholders ensures that AI solutions address disparities in access to technology. The framework generates multilingual interfaces, affordable deployment strategies, and culturally sensitive adaptations through the co-design process [14, 15].

3.2. Replicability and Ecological Validity

To enhance the replicability of AI development processes, the framework incorporates ecological validity assessments. These provide structured guidelines ensuring AI technologies remain contextually relevant and adaptable across diverse healthcare settings. By grounding AI solutions in specific cultural, social, and clinical contexts, ecological validity ensures practical applicability and widespread acceptance. Key metrics include user satisfaction surveys, adoption rates, and the absence of biases in AI outputs [18, 19].

In pediatric telerehabilitation contexts, ecological validity involves evaluating AI system performance in real-world clinical settings with diverse patient populations, considering factors like cultural background and varying levels of health literacy.

3.3. Research Roadmap: Establishing Accessibility Guidelines

The framework guides the development of standardized accessibility guidelines for AI systems. Inspired by accessibility standards in web and application design, the roadmap aims to:

- Define accessibility benchmarks for explainability, usability, and inclusivity tailored to healthcare AI applications [20].
- Develop open resources, creating publicly available toolkits for developers, clinicians, and patients to co-design AI solutions [12, 13].
- Implement pilot programs to test these guidelines in diverse clinical environments, such as pediatric rehabilitation [14, 15].
- Iterate and scale using feedback from pilot programs to refine the guidelines.

These guidelines will ultimately support the development of sustainable AI healthcare solutions by ensuring they are accessible, understandable, and beneficial to all stakeholders.

4. Evaluating the Framework

The effectiveness of the proposed HCD-driven framework hinges on its ability to address practical challenges while fostering trust, equity, and sustainability in healthcare AI. This section outlines the methods and metrics for evaluating the framework's impact in real-world clinical settings.

4.1. Key Evaluation Metrics

The evaluation framework encompasses four key dimensions:

1. Engagement measures the diversity and active participation of stakeholders in co-design and validation sessions [11, 12], including tracking the number of iterative feedback cycles to assess sustained collaboration [14, 15].
2. Acceptance is evaluated through structured surveys and qualitative feedback focusing on usability and relevance [13, 16], alongside monitoring adoption rates of AI tools in clinical workflows [17, 18].
3. Clinical effectiveness analyzes therapeutic outcomes, such as improved diagnosis accuracy and enhanced treatment efficiency [19], while assessing how AI systems influence clinicians' decision-making confidence and patient trust.
4. Equity is measured through regular audits to identify and address biases in AI recommendations, evaluating accessibility improvements such as multilingual interfaces and equitable distribution of AI benefits across socio-demographic groups.

4.2. Validation through Pilot Studies

The framework will undergo rigorous validation in real-world clinical settings through two main approaches: (i) Pediatric Rehabilitation pilots will test the framework's applicability in co-designing AI tools for pediatric care, emphasizing tailored explanations and sustainable practices while engaging families, clinicians, and policymakers to ensure solutions align with cultural and clinical contexts. (ii) Broader Healthcare Applications will expand testing to other domains, such as oncology and chronic disease management, to evaluate the framework's scalability [18, 20].

4.3. Iterative Refinement and Feedback

An iterative feedback loop will guide continuous improvement through:

- Collection of qualitative and quantitative data through surveys, focus groups, and usability tests [15, 20].
- Refinement of AI solutions based on feedback, ensuring alignment with user needs and ecological validity [17, 19].
- Documentation of lessons learned to update guidelines and inform future implementations [19, 20].

4.4. Expected Outcomes and Impact

Through the application of HCD principles, this framework is expected to achieve several key outcomes:

- Enhanced trust through transparent and tailored explanations will foster greater confidence in AI systems [14, 16], strengthening informed consent and allowing better patient engagement in care decisions [6, 7].
- Improved clinical effectiveness will result from integrating Accessible XAI and HCD, supporting more informed therapeutic choices and improved care pathways [8, 14].
- Greater equity will emerge through inclusive design, ensuring AI benefits are accessible to underserved populations [16, 19], addressing health literacy, cultural differences, and socioeconomic barriers [10, 11].
- Looking ahead, the framework aims to develop comprehensive guidelines based on HCD principles, expand interdisciplinary collaborations, and scale applications to address global health disparities. These objectives will strengthen the practical impact of AI while promoting equitable, transparent, and sustainable integration into healthcare ecosystems.

5. Discussion and Conclusions

This paper has presented a holistic framework for designing and implementing AI systems in healthcare, rooted in Accessible XAI, Human-Centered Design, and Sustainability. Unlike existing models focusing on isolated aspects, this integrated approach offers a comprehensive solution reflecting the complexity of clinical environments [1, 2, 3, 4].

A central contribution is the framework's capacity to enhance trust and equity through Accessible XAI. Traditional XAI models often provide static explanations, whereas this framework emphasizes tailoring explanations to users' profiles [6, 7, 8]. The participatory design approach embedded in HCD ensures that AI solutions are developed collaboratively with stakeholders, addressing real-world needs [12].

On the sustainability front, this framework emphasizes operational longevity alongside social and economic dimensions. By integrating equity-focused design principles with green technologies and continuous updates, the framework aligns innovation with environmental and economic

goals. While it builds on established concepts, the integrated structure and strong focus on inclusivity mark a significant step forward.
Future research will focus on empirical validation in real-world clinical settings, particularly in pediatric rehabilitation, evaluating trust, transparency, and clinical effectiveness. Building on insights from Turchi et al. [20], interdisciplinary collaboration will be prioritized to refine the model's compliance with regulations and alignment with societal expectations. This validation process will expand across multiple healthcare settings to deepen clinical impact and ensure long-term sustainability and inclusivity.

Acknowledgements

This work has been partially funded by the AInCP Project, supported by the European Commission under Horizon Europe Research and Innovation Action (GA n. 101057309). Additional funding was provided by the AInCP Consortium to integrate and validate AI-driven solutions in pediatric rehabilitation across European centers.
This work was also produced with the co-funding of the European Union – Next Generation EU, in the context of The National Recovery and Resilience Plan, Investment 1.5 Ecosystems of Innovation, Project Tuscany Health Ecosystem (THE), ECS00000017. Spoke 3.

References

[1] C. H. Nwokoye, M. J. P. Peixoto, A. Pandey, L. Pardy, M. Sukhai, and P. R. Lewis, "A Survey of Accessible Explainable Artificial Intelligence Research," arXiv:2407.17484, 2024.
[2] M. L. Tielman, M. C. Suárez-Figueroa, A. Jönsson, M. A. Neerincx, and L. C. Siebert, "Explainable AI for All: A Roadmap for Inclusive XAI for People with Cognitive Disabilities," Technology in Society, vol. 79, 2024.
[3] C. T. Wolf and K. E. Ringland, "Designing Accessible, Explainable AI (XAI) Experiences," SIGACCESS Access. Comput., no. 125, 2020.
[4] R. Ahmad, R. Ahmad, N. Ahmad, and R. Aggarwal, "Artificial intelligence (AI) in medicine, current applications and future role," Journal of Laboratory Physicians, vol. 7, no. 1, pp. 1-8, 2021.
[5] AInCP Proposal. (2022). Horizon Europe call HORIZON-WIDERA-2022-ACCESS-07, Proposal ID 101119320.
[6] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," BMC Medical Informatics and Decision Making, vol. 20, no. 1, 2020.
[7] M. Anderson, S. Albala, B. Casadei, et al., "Health information technology and digital innovation for national learning health and care systems," The Lancet Digital Health, vol. 3, no. 6, pp. e383-e396, 2021.
[8] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems, pp. 4765-4774, 2017.
[9] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144, 2016.
[10] .N. Farmer, F. Osei Baah, F. Williams, E. Ortiz-Chapparo, V. M. Mitchell, L. Jackson, and A. Johnson, "Use of a community advisory board to build equitable algorithms for participation in clinical trials: A protocol paper for HoPeNET," BMJ Health & Care Informatics, vol. 29, no. 1, e100453, 2022.
[11] R. Fehr, A. Renz, B. Prainsack, and A. Tsamados, "A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare," The Lancet Digital Health, vol. 6, no. 1, pp. e42-e50, 2024.

[12] S. Filogna, A. Malizia, D. Mazzei, G. Prencipe, G. Sgandurra, and T. Turchi, "Telemedicine and AI: From Co-Design to Explainability," in 2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI), pp. 363-368, 2024.
[13] J. C. Bjerring and J. Busch, "Artificial intelligence and patient-centered decision-making," Journal of Medical Ethics, vol. 47, no. 1, pp. 58-64, 2021.
[14] C. P. Helenason, C. Attwood, J. Allen, et al., "Exploring the feasibility of an artificial intelligence-based clinical decision support system for cutaneous melanoma detection," The British Journal of General Practice, vol. 74, no. 739, pp. e899-e907, 2024.
[15] A. Sauerbrei, A. Kerasidou, F. Lucivero, and N. Hallowell, "The impact of artificial intelligence on the person-centered, doctor-patient relationship: some problems and solutions," BMC Medical Informatics and Decision Making, vol. 23, no. 1, 73, 2023.
[16] S. L. Rose and D. Shapiro, "An ethically supported framework for determining patient notification and informed consent practices when using artificial intelligence in health care," Chest, vol. 166, no. 3, pp. 572-578, 2024.
[17] A. Harper, N. Mustafee, and S. Burns, "Participatory design research for the development of real-time simulation models in healthcare," Simulation in Healthcare, vol. 12, no. 4, pp. 375-386, 2023.
[18] D. Morana, T. Balduzzi, and F. Morganti, "La salute 'intelligente': eHealth, consenso informato e principio di non-discriminazione," Federalismi.it, vol. 34, 2022.
[19] M. Romanello, C. Di Napoli, P. Drummond, et al., "The 2023 report of the Lancet Countdown on health and climate change," The Lancet, vol. 402, no. 10419, pp. 2346-2394, 2023.
[20] T. Turchi, G. Prencipe, A. Malizia, S. Filogna, F. Latrofa, and G. Sgandurra, "Pathways to democratized healthcare: Envisioning human-centered AI-as-a-service for customized diagnosis and rehabilitation," Artificial Intelligence in Medicine, vol. 151, no. 102850, 2024.