Towards HCXAI: Explainability Preferences of Healthcare Professionals

Mishell Cadena-Yanez, Angela Bernardini, and Marisol Gómez

NAIR Center-Universidad Pública de Navarra

2026/05/05

1. Introduction
2. Methodology
3. Results
4. Limitations

XAI Cultures

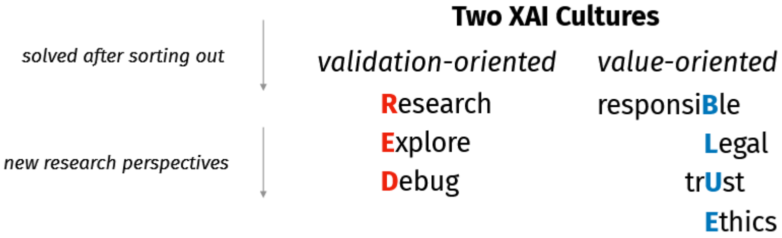


Figure 1: Blue Red XAI Biecek and Samek

- (RQ1) How do HCP's perceptions of current AI limitations influence their differentiated trust and what role does explainability play in fostering justified trust?
- (RQ2) What type of explanations do HCP require regarding AI system recommendations?

Structure of questionnaire

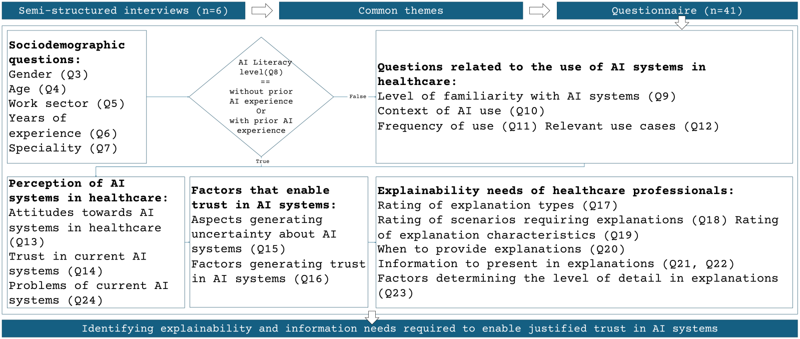
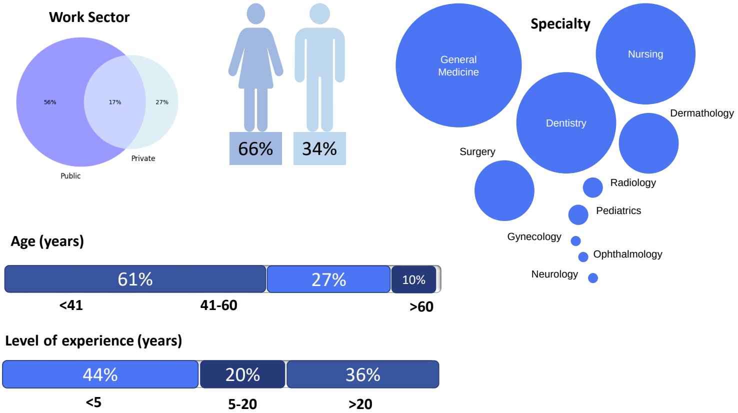


Figure 2: Structure of questionnaire, grouped according to thematic blocks.

Sociodemographic profile

N = 41

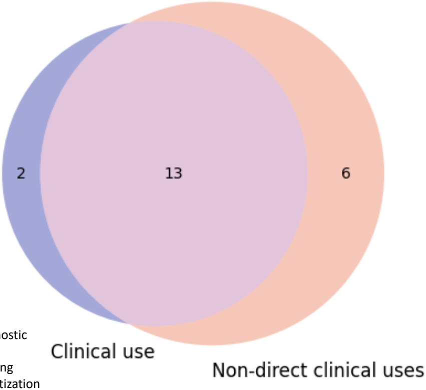


AI Literacy level

With prior experience

n=21

Weekly use (33.3\%)



- Early detection
- Support for diagnostic confirmation
- Treatment planning
- Triage case prioritization
- Second clinical opinion

- Searching for medical information
- Report writing
- Personal consultations
- Learn

Trust and Perception of HCP towards AI systems

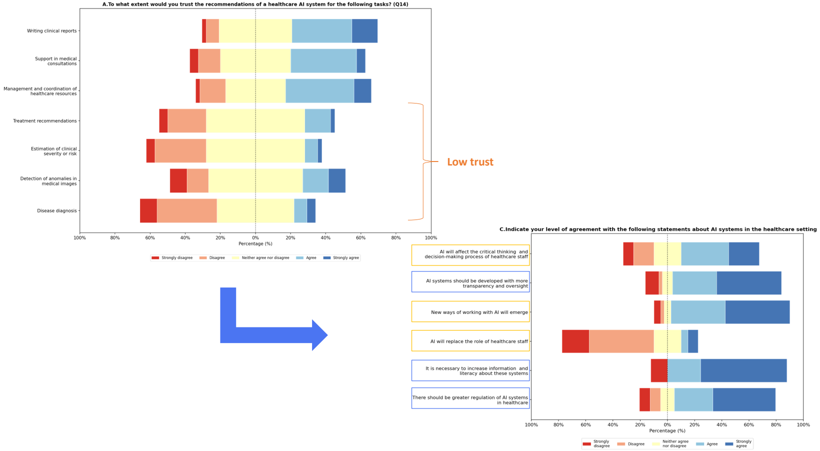
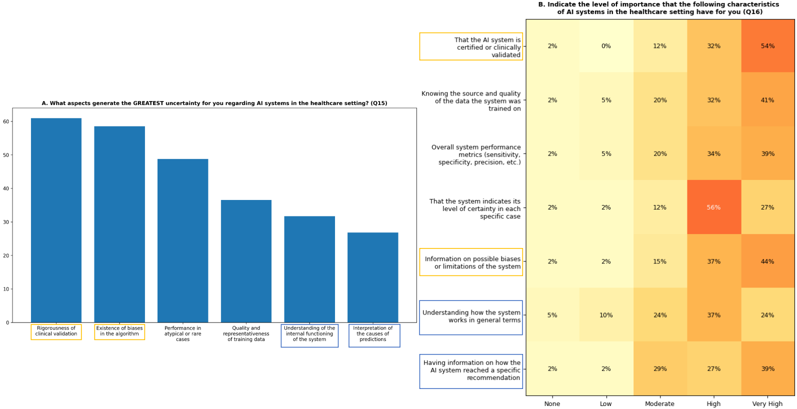


Figure 5: Level of trust in AI systems (Q14), Attitudes towards AI healthcare systems(Q13)

Factors generating trust



Generic information and lack of clinical context push HCP to demand clinical validation and bias rigour above all. Explainability is complementary. Explainability supports (but does not replace) that foundation, and only when it is contextualized and clinically relevant

Figure 6: Aspects generating uncertainty about AI systems (Q15). Factors generating trust in AI systems (Q16)

Perception of HCP towards AI systems information

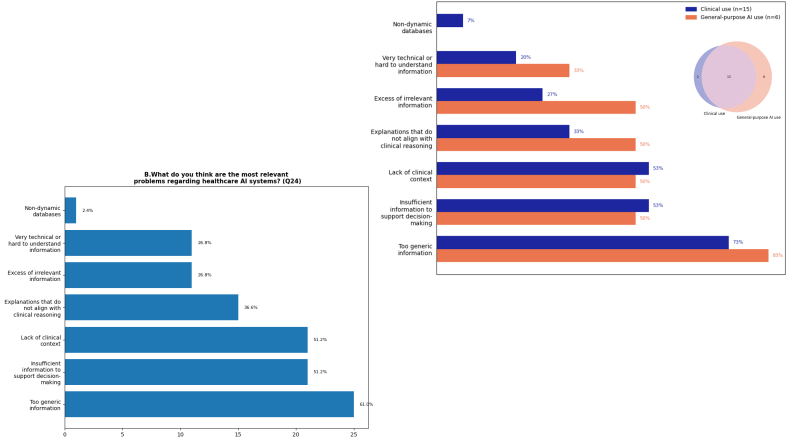


Figure 7: Most relevant problems of current AI systems (Q24)

XAI in clinical scenarios

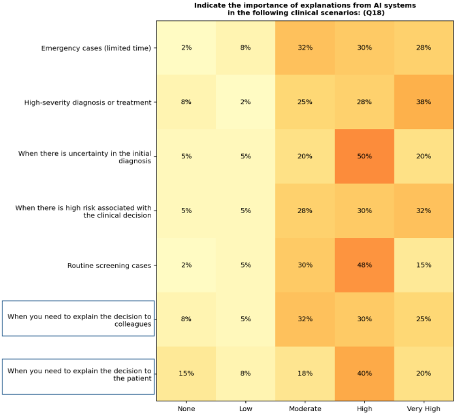


Figure 8: Importance of XAI in different clinical scenarios (Q18)

Level of detail in explanations

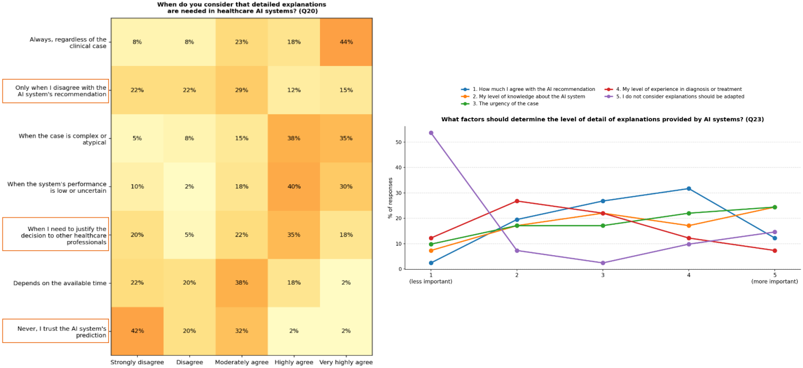


Figure 9: Level of details(Q20) and characteristics of details in different scenarios(Q23)

Types of explanations

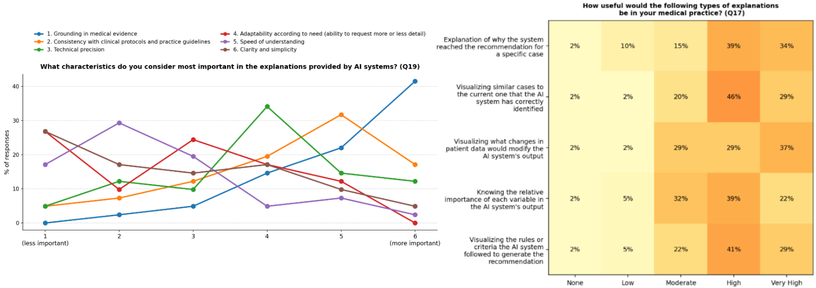


Figure 10: Types (Q17) and characteristics of explanations(Q19)

Types and characteristics of explanations in agree (Q22) or disagree scenarios(Q21)

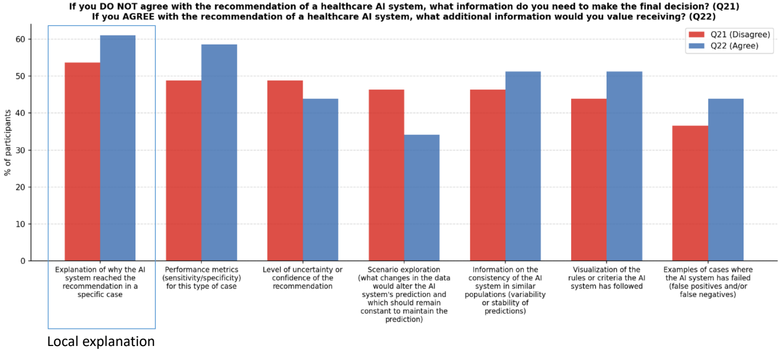


Figure 11: Types of explanations in case of agree or disagree output AI

- E-mail recruitment may have introduced self-selection bias, potentially overrep- resenting professionals with a prior interest in AI.
- Sample is geographically concentrated in Navarra, Spain

Bibliografía I

Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, and Vince I. Madai. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. 20(1):310. ISSN 1472-6947. doi: 10.1186/s12911-020-01332-6. URL https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6.

Amie J. Barda, Christopher M. Horvat, and Harry Hochheiser. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. 20(1):257. ISSN 1472-6947. doi: 10.1186/s12911-020-01276-x. URL https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01276-x.

Bibliografía II

Przemyslaw Biecek and Wojciech Samek. Position: Explain to question not to justify. URL http://arxiv.org/abs/2402.13914.

Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L. Beam. The false hope of current approaches to explainable artificial intelligence in health care. 3(11):e745–e750. ISSN 2589-7500. doi: 10.1016/S2589-7500(21)00208-9. URL https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00208-9/fulltext. Publisher: Elsevier.

Bibliografía III

Imanol Pinto, Álvaro Olazarán, David Jurío, Borja De la Osa, Miguel Sainz, Aritz Oscoz, Jerónimo Ballaz, Javier Gorricho, Mikel Galar, and José Andonegui. Improving diabetic retinopathy screening using artificial intelligence: design, evaluation and before-and-after study of a custom development. 7. ISSN 2673-253X. doi: 10.3389/fdgth.2025.1547045. URL https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2025.1547045/full. Publisher: Frontiers.

Bibliografía IV

Massimo Salvi, Silvia Seoni, Andrea Campagner, Arkadiusz Gertych, U. Rajendra Acharya, Filippo Molinari, and Federico Cabitza. Explainability and uncertainty: Two sides of the same coin for enhancing the interpretability of deep learning models in healthcare. 197:105846. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2025.105846. URL https://www.sciencedirect.com/science/article/pii/S1386505625000632.

Mark Sendak, Madeleine Clare Elish, Michael Gao, Joseph Futoma, William Ratliff, Marshall Nichols, Armando Bedoya, Suresh Balu, and Cara O'Brien. "the human body is a black box": supporting clinical decision-making with deep learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pages 99–109. Association for

Bibliografía V

Computing Machinery. ISBN 978-1-4503-6936-7. doi: 10.1145/3351095.3372827. URL https://dl.acm.org/doi/10.1145/3351095.3372827.

Hamid Shamszare and Avishek Choudhury. Clinicians' perceptions of artificial intelligence: Focus on workload, risk, trust, clinical decision making, and clinical integration. 11(16):2308. ISSN 2227-9032. doi: 10.3390/healthcare11162308. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10454426/.

Sana Tonekaboni, Shalmali Joshi, Melissa D. McCradden, and Anna Goldenberg. What clinicians want: Contextualizing explainable machine learning for clinical end use. URL http://arxiv.org/abs/1905.05134.

Bibliografía VI

Catharina M. Van Leersum and Clara Maathuis. Human centred explainable AI decision-making in healthcare. 21:100108, 2025. ISSN 26666596. doi: 10.1016/j.jrt.2025.100108. URL https://linkinghub.elsevier.com/retrieve/pii/S2666659625000046.

Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–15. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300831. URL https://dl.acm.org/doi/10.1145/3290605.3300831.

Bibliografía VII

Oskar Wysocki, Jessica Katharine Davies, Markel Vigo, Anne Caroline Armstrong, Dónal Landers, Rebecca Lee, and André Freitas. Assessing the communication gap between AI models and healthcare professionals: explainability, utility and trust in AI-driven clinical decision-making. 316:103839. ISSN 00043702. doi: 10.1016/j.artint.2022.103839. URL http://arxiv.org/abs/2204.05030.

Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang 'Anthony' Chen. CheXplain: Enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376807. URL https://dl.acm.org/doi/10.1145/3313831.3376807.

Bibliografía VIII

Wei Xiong, Hongmiao Fan, Liang Ma, and Chen Wang. Challenges of human—machine collaboration in risky decision-making. 9(1): 89–103, 2022. ISSN 2095-7513, 2096-0255. doi: 10.1007/s42524-021-0182-0. URL https://link.springer.com/10.1007/s42524-021-0182-0.

Thanks!