CHERIE: User-Centred Development of an XAI System for Chest Radiology through Co-Design

Shereen Fouad
Applied AI and Robotics
Aston University, UK

Lilit Hakobyan
Engineering and Physical Sciences
Aston University, UK

Manolya Kavakli-Thorne
Aston Digital Futures Institute
Aston University, UK

Sarah Atkins
Social Sciences and Humanities
Aston University, UK

Antonio Fratini
Izegbua E. Ihongbe
Engineering and Physical Sciences
Aston University, UK

Bahadar Bhatia
Arvind Rajasekaran
Pankaj Nagori
John Morlese
Sandwell And West Birmingham Hospitals NHS Trust, UK

Modern medical imaging systems increasingly use AI for analysis and diagnosis, yet the "black-box" nature of current deep learning algorithms limits their practical use in radiology. Explainable AI (XAI) aims to address this by making AI decisions more transparent and interpretable. In medical imaging, XAI tools often highlight critical regions in images to explain AI decisions, but their complex visual explanations and poor UI design impede their clinical adoption. This study introduced CHERIE, an XAI prototype designed to enhance transparency in AI-assisted chest radiology. Using our pre-developed XAI diagnostic tool for chest radiology, we adopted a user-centered design (UCD) methodology to develop user interfaces for the AI-enabled diagnostic tool. In particular, we engaged medical practitioners, AI developers, and HCI experts in a multidisciplinary co-design workshop. This collaborative effort was crucial in identifying requirements from the user perspectives, aiming to boost understanding and trust in AI-driven diagnostics. Our findings emphasise the need for UCD for the adoption of XAI systems, proposing user requirements to seamlessly integrate these systems into clinical workflows and effectively address end-user needs.

User-centred design, Co-design, Explainable AI, Medical imaging

1. INTRODUCTION

Modern medical imaging systems increasingly depend on AI-based computer vision for analysis, diagnosis, and treatment (Fouad et al., 2021), (Fouad et al., 2017b). However, their complexity and black box nature have hindered adoption in healthcare. XAI is an emerging concept that focuses on developing methods to improve the transparency, interpretability, and trustworthiness of complex AI models (Arrieta et al., 2020), (Tjoa and Guan. 2020). In the context of medical imaging, current XAI tools predominantly offer post-hoc explainability, which visually interprets AI results by identifying and highlighting the critical regions in the image that influence the AI's decisions (de Vries et al. 2023). While visual explainability in AI-based tools for medical imaging shows significant promise, several usability issues may impede their effective adoption in clinical practice. As noted by (Schoonderwoerd et al. 2021), visual XAI tools frequently suffer from suboptimal UI design, which can adversely affect their usability, making interaction and interpretation of the provided information difficult.

Most current XAI researchers focus on the computational aspects of generating explanations, with insufficient emphasis on the user-centred design (UCD) of the user interface. These limitations risk rendering current XAI tools for medical imaging incomprehensible to medical professionals, thereby reducing their adoption, clinical relevance and trustworthiness (Schoonderwoerd et al. 2021). This paper investigates the application of a co-design methodology for

Proceedings of . . .

designing the UI of visual XAI systems in radiology imaging, denoted here as CHERIE (CHEst RadIology Explainable). We detail the co-design process, which involved medical practitioners, AI developers and HCI experts to integrate domain-specific knowledge and usability considerations into the UI design, proposing a UI for CHERIE that has the potential to enhance comprehension and foster trust in AI-assisted diagnostic processes.

2. BACKGROUND

Within the UCD framework, co-design, a subset of participatory design (PD), focuses on actively engaging and integrating diverse stakeholders into the design process, which is essential for AI development (Neuhauser et al. 2013). The primary objective of co-design is to leverage the collective knowledge and insights of all stakeholders, particularly end-users, to effectively innovate and resolve problems. In this paper, the term co-design will be used to refer to this inclusive design practice. The global market for AI-based healthcare solutions is projected to reach $208 billion by 2030, expanding at a compound annual growth rate of 38.5 per cent from 2022 to 2030 (Zhukovska et al. 2023). The integration of AI technologies in healthcare systems holds the potential to enhance efficiency and improve care quality (Raparthi, M. 2020). This advancement, however, raises concerns about fairness, bias, and inclusivity, as the lack of equity, diversity, and inclusion considerations can exacerbate existing inequalities and perpetuate systemic biases (Nyariro et al. 2023).

Co-design allows navigating complex ethical and societal considerations and seeks to promote transparency, accountability, and trustworthiness in AI systems by incorporating diverse viewpoints in the co-design process. This inclusive approach helps address ethical dilemmas such as data privacy and bias, ensuring that the AI systems are developed in a manner that aligns with societal values and clinical standards.

(Zicari et al. 2021) present an exemplary model of co-design in their development of an AI system component aimed at explaining decisions made by deep learning networks analysing images of skin lesions. Their approach, known as Z-inspection®, involved a multidisciplinary team working collaboratively to create AI systems. This inclusive approach facilitated the alignment of technical development with clinical needs and workflows, ensuring that the AI tools are not only advanced but also practically applicable in medical settings.

(Panigutti et al. 2023) explored the co-design approach for developing user-centred XAI interfaces, focusing on effectively presenting explanations to users. Traditional XAI methodologies often introduce user involvement late in the development process, but recent initiatives emphasise engaging domain experts early, particularly in medical fields, to evaluate the impact of explanations on clinical decision-making. Their study involved an online experiment with healthcare providers, who performed tasks and provided feedback via questionnaires. Findings indicated that user feedback is crucial in designing explanation interfaces, significantly influencing trust and decision performance. Participants highlighted a need for explanations that address uncertainties in AI-generated advice.

Another study employed a design thinking approach augmented with theoretical frameworks to develop a prognostic tool predicting the likelihood of admission to an intensive care ward based on disease severity in chest x-ray images, particularly in the context of the COVID-19 pandemic (Shulha et al. 2024). Stakeholder involvement proved essential across several research phases (including ideating and testing). The researchers highlighted the importance of continuous clinician engagement and a robust design thinking framework in developing transparent and explainable AI decision support systems.

In a related study, (Gerlings, Jensen, and Shollo, 2022) examined XAI's role in AI-driven solutions for respiratory diseases, particularly for COVID-19 diagnosis using lung X-rays. They investigated how the need for XAI arises throughout an AI application's development lifecycle through a case study of an AI startup. Methods included semi-structured interviews, online workshops, and document analysis, focusing on the development team and subject matter experts.

Distinctive XAI requirements for various stakeholder groups were identified as follows: development teams require explanations for understanding datasets and model functions; subject matter experts stress domain knowledge and validation; decision-makers need clear output interpretations for trust and accountability; and audience concerns revolve around patient communication and decision impact management.

These studies underscore the multidisciplinary requirements for addressing XAI requirements, balancing explanation accuracy with comprehensibility tailored to varying technical expertise and involvement levels. Despite this need, comprehensive UI design frameworks for XAI in medical imaging remain underexplored.

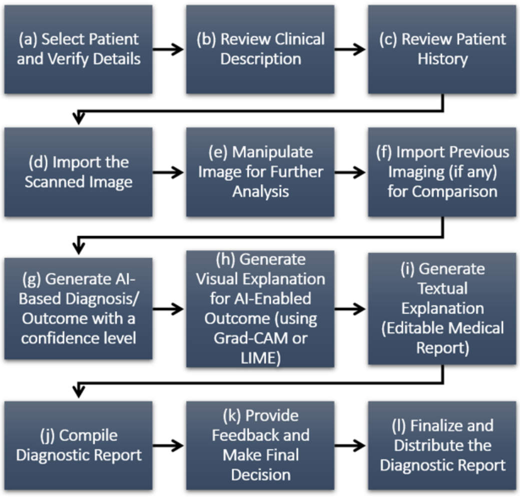


Figure 1: *Journey Map for Diagnosing Patients Using Explainable AI diagnostic tool for Chest Radiology Imaging*

3. METHOD

Grad-CAM (Gradient-weighted Class Activation Mapping) (Selvaraju et al. 2017) and LIME (Local Interpretable Model-Agnostic Explanations) (Ribeiro et al. 2016) are widely used XAI algorithms in medical imaging that adapt to various deep learning models, providing localised explanations by highlighting influential image regions (e.g. Barzas et al. (2023)). This capability aids radiologists in validating AI decisions and enhances radiology report quality, while also supporting clinicians in patient consultations by involving them in the decision-making process. Despite their utility, these algorithms have not been extensively validated in real-life clinical scenarios, relying instead on automated evaluations which may not fully meet clinical explainability requirements. We developed an XAI system CHERIE (CHEst RadIology Explainable) using a range of widely available deep learning algorithms.

In designing the UI for CHERIE, we adopted a UCD approach by initially conducting a comprehensive requirement analysis (Ihongbe et al. 2024) through a literature review and focus group meetings with healthcare professionals, AI experts, and human-computer scientists. This analysis aimed to understand the interaction needs and design constraints for AI diagnostic tools, focusing on clinical workflows and current system limitations. Consequently, we identified fourteen initial design requirements (DRs) for our clinical use case, that can be categorised into four categories of audiences, depending on the stakeholders' background. This includes (I) Development Teams (e.g., AI developers), focusing on presentation elements. (II) Subject Matter or Domain Experts (e.g. radiologists), focusing on clinical descriptions and order of workflow elements. (III) Decision-Makers (e.g. clinicians), focusing on AI-enabled outcomes, and (IV) Audience (e.g. nurses, Healthcare providers, and patients). The DRs for each of audiences are listed in Figure 2. As shown in the figure, development teams focus on data understanding and model functions, emphasizing presentation, while domain experts prioritize domain knowledge and validation, stressing clinical descriptions and workflow order. On the other hand, decision-makers need clear output interpretations for trust and accountability, focusing on AI outcomes, and audience (such as nurses and patients) are concerned with communication and decision impacts.

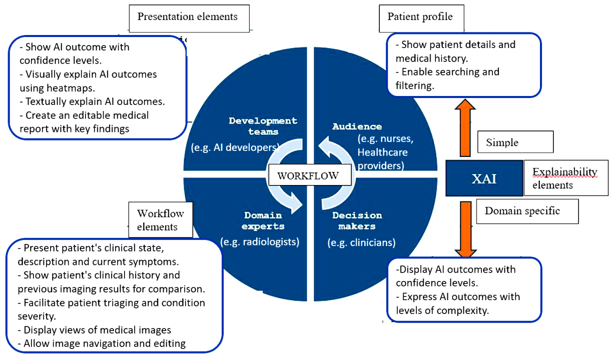


Figure 2: *User Requirement Framework for CHERIE XAI*

Building on this, we developed a journey map (see Figure1) and wireframes to visualise the conceptual UI. The journey map outlines steps from selecting patients and reviewing details to integrating AI insights into a diagnostic report. A detailed discussion of these initial phases is beyond the scope of this paper. Instead, we focus on the final co-design phase of the research, which is discussed below.

3.1. Multidisciplinary co-design workshop

To refine and assess the initial prototype developed in earlier research phases, a one-day multidisciplinary co-design workshop was organised. This event was supported by the Alan Turing Institute (March 2024, AI UK Fringe)[1]. The workshop brought together 30 key stakeholders from various disciplines, including AI scientists, medical image analysis, clinical medicine, radiology, human-computer interaction (HCI) and social sciences. Around one third of the participants had a relevant healthcare profession, while the remaining participants were scientists from other disciplines, including AI and HCI.

Participants were divided into four groups, each composed of an equal mix of expertise (engineering/computer science experts, researchers, and

[1] `https://www.turing.ac.uk`

medical professionals) to ensure diverse discussions. The first part of the workshop included an introduction to the investigated problem as well as a review of the proposed UI prototype (developed in the previous phases of the research) for CHERIE. Participants discussed and validated the UI design by addressing the a number of questions about preferred format for displaying diagnostic predictions, target audience for the AI prediction, required information, visual explanations, imaging workflow and consultation with patients.

In the second part of the workshop, the groups engaged in a co-design activity, collaboratively proposing interfaces that addressed the shortcomings of the current UI design (see example screens in Figure3 for the proposed UI for the XAI clinical use case). Participants were provided with common office supplies such as post-it notes, papers, and pens to engage in paper prototyping.

3.2. Data Collection and Analysis

The data collected from the workshop included paper sketches and notes for each group, containing handwritten responses to the series of questions (outlined above) (see example Figure 4), as well as feedback on the proposed UI for the XAI clinical use case (Figure 3). The written notes were analysed as a method of exploring key concerns, and insights of stakeholder participants. They were scrutinised for key concepts related to the requirements and feedback/responses to the co-design questions (listed above).

4. RESULTS

The workshop session provided several insights and recommendations regarding the optimal presentation formats for computer-aided diagnostic predictions. A consensus emerged that combining visual XAI output with accompanying textual explanations provides the most effective format to explain the AI decision in a human understandable format. Participants also highlighted the utility of zoomable images, which enable detailed examination of specific areas, crucial for accurate diagnosis.

There was a strong preference for presenting the original radiology image (e.g., X-ray) before the automatic explainable AI diagnosis to avoid potential bias. Moreover, side-by-side presentations of the original X-ray images and the generated XAI image, as well as the option to overlay the visual XAI output (e.g, heatmap) on the original X-ray image with adjustable opacity, were recommended for facilitating quick and effective comparisons. Additional features like greyscale filters with RGB and dark mode enabled viewing were suggested to further improve visual clarity and reduce eye strain, especially in dark environments.

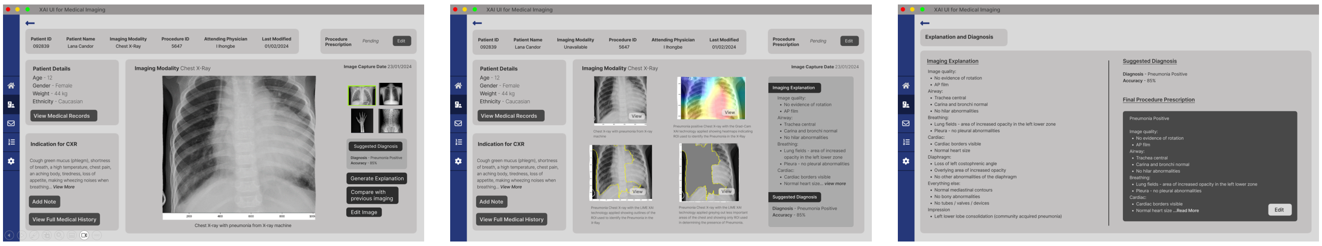
Regarding the target audience for AI explanations, radiologists were identified as the primary recipients. Given their working conditions, the explanations need to be optimised for darkrooms and large screens. Clinicians were also a key end-users, requiring concise and clear information suitable for typical clinical settings. Simplified explanations were deemed necessary for patients to ensure accessibility and comprehension, while researchers would benefit from detailed and comprehensive information.

The essential information to be included in AI explanations was also discussed. Unique identifiers were deemed crucial for radiologists to correlate patient data with the images on their screens. The explanations should be tailored in complexity based on the recipient, whether they are radiologists, clinicians, patients, or researchers. Clear and detailed reasoning behind predictions, integrated patient information, and device-agnostic presentation were emphasised to ensure that the information is accessible and useful across various devices and user needs.

In terms of the preferred modality or format of explanation, visual representation was preferred for its clarity and immediate impact. Textual explanations were also favoured for providing detailed context and reasoning. Audio explanations, while considered inclusive, faced some opposition due to potential sensory overload in clinical settings.

To enhance visual explanations, several improvements were suggested. Incorporating 3D visualisations was recommended to add depth perception to the images, thereby providing a more comprehensive view. The use of greyscale filters with RGB and dark mode enabled viewing were reiterated as beneficial for improving image analysis and reducing eye strain. An improved GRAD-CAM color scheme, specifically blue-toned, was suggested to align with radiologists' familiarity and preferences. Moreover, ensuring that LIME explanations do not obscure images was highlighted as essential for maintaining image clarity.

The integration of the proposed UI design into the current imaging workflow was another key discussion point. Participants envisioned a seamless connection with existing information sources and necessary systems to enhance workflow efficiency. It was suggested that the new UI should bridge the workflows of radiologists and clinicians into a single cohesive system, thereby improving overall coordination and efficiency.



(a) *Screen for selecting Patients, Verifying Details, Reviewing Clinical Description, and Importing the Scanned Image.*

(b) *Screen for displaying image content alongside AI diagnosis with Visual Explanation using Grad-CAM and LIME*

(c) *Screen for Generating Textual Explanation (Editable Medical Report) with key findings from the image.*

Figure 3: *Example UI prototype design for the Explainable AI diagnostic tool for Chest Radiology Imaging.*

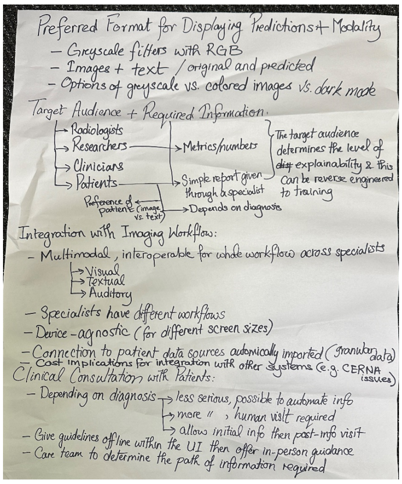


Figure 4: *Example handwritten notes from the co-design workshop*

5. DISCUSSION

The requirements for various stakeholder groups elicited through the co-design workshop are aligned with the XAI requirements suggested by (Gerlings, Jensen, and Shollo, 2022). As demonstrated in Figure 2, XAI Workflow runs from Development Teams to Domain experts, then to Decision Makers and Patients. Development teams require explanations for understanding datasets and model functions; their primary concern is the presentation elements. Domain experts stress domain knowledge and validation; their primary concerns are clinical descriptions and order of workflow elements. Decision-makers need clear output interpretations for trust and accountability; their primary concern is AI-enabled outcomes. Audience concerns revolve around patient communication and decision impact management; their primary concern is patient profile. XAI serves Decision makers and Audience (patients) and the recommendations must be at varying level of complexity depending on the patient profile, ranging from simple to discipline specific interpretations.

In terms of presentation elements, accurate and clear visual representations must be zoomable and complemented by textual information without any occlusion. 3D views, greyscale filters with RGB and dark mode enabled viewing may enhance visual quality. Workflow for these elements should be carefully ordered to prevent bias. Unique identifiers for patients are crucial and detailed reasoning behind predictions should cater for the patient profile.

6. CONCLUSION

In this paper, we developed an XAI prototype (CHERIE) for chest radiology to explore system requirements for making AI decisions more transparent and interpretable. CHERIE highlights critical regions in xRays to explain AI decisions. For large scale clinical adoption, we adopted a UCD approach involving relevant stakeholders such as medical practitioners, AI developers, and HCI experts. This collaborative approach helped identify system requirements for an XAI UI to improve comprehension and trust in AI-assisted diagnostics. Our findings emphasise the need for user-centred design for the adoption of XAI systems to ensure that they fit seamlessly into clinical workflows and meet end-user needs. In future studies, we plan to deploy the prototype for clinical use after incorporating the features identified in the co-design sessions. After deploying the XAI system, it will undergo ongoing evaluation for continuous improvement.

REFERENCES

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R. (2020) *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information fusion, 58, pp.82-115.

Barzas, K., Fouad, S., Jasa, G., and Landini, G., (2023) *An Explainable Deep Learning Framework for Mandibular Canal Segmentation from Cone Beam Computed Tomography Volumes*. The 12th International Conference on Computational Advances in Bio and Medical Sciences. Springer.

Fouad, S., Landini, G., Robinson, M., Song, T.-H., and Mehanna, H. (2021) Human papilloma virus detection in oropharyngeal carcinomas with in situ hybridisation using hand crafted morphological features and deep central attention residual networks. Computerized Medical Imaging and Graphics, 88, 101853.

Fouad, S., Randell, D., Galton, A., Mehanna, H., and Landini, G. (2017) Epithelium and stroma identification in histopathological images using unsupervised and semi-supervised superpixel-based segmentation. Journal of Imaging, 3(4), 61.

Gerlings, J., Jensen, M.S. and Shollo, A. (2022) *Explainable AI, but explainable to whom? An exploratory case study of xAI in healthcare.* Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects, pp.169-198.

Ihongbe, I., Fouad, S., Mahmoud, T., Rajasekaran, A., and Bhatia, B., (2024) *Evaluating Explainable Artificial Intelligence (XAI) Techniques in Chest Radiology Imaging Through a Human-Centered Lens*. Submitted for publication in PLOS ONE.

Lim, C.-P., Vaidya, A., Jain, K., Mahorkar, V. U., and Jain, L. C., (2022) *Handbook of Artificial Intelligence in Healthcare*. Springer.

Neuhauser, L., Kreps, G. L., Morrison, K., Athanasoulis, M., Kirienko, N. and Van Brunt, D. (2013) *Using design science and artificial intelligence to improve health communication: ChronologyMD case example.* Patient education and counseling, 92(2), 211-217.

Nyariro, M., Emami, E., Caidor, P. and Rahimi, S. A. (2023) *Integrating equity, diversity and inclusion throughout the lifecycle of AI within healthcare: a scoping review protocol.* BMJ open, 13(9), e072069.

Panigutti, C., Beretta, A., Fadda, D., Giannotti, F., Pedreschi, D., Perotti, A. and Rinzivillo, S. (2023) *Co-design of human-centered, explainable AI for clinical decision support.* ACM Transactions on Interactive Intelligent Systems, 13(4), pp.1-35.

Raparthi, M. (2020) *AI Integration in Precision Health-Advancements, Challenges, and Future Prospects.* Asian Journal of Multidisciplinary Research and Review, 1(1), 90-96.

Ribeiro, M. T., Singh, S., and Guestrin, C., (2016) *"Why should I trust you?" Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A. and Van Den Bosch, K. (2021) *Human-centered XAI: Developing design patterns for explanations of clinical decision support systems.* International Journal of Human-Computer Studies, 154, 102684.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., (2017) *Grad-cam: Visual explanations from deep networks via gradient-based localization*. Proceedings of the IEEE International Conference on Computer Vision, 618–626.

Shulha, M., Hovdebo, J., D'Souza, V., Thibault, F. and Harmouche, R. (2024) *Integrating Explainable Machine Learning in Clinical Decision Support Systems: Study Involving a Modified Design Thinking Approach.* JMIR Formative Research, 8(1), p.e50475.

Tjoa, E. and Guan, C. (2020) *A survey on explainable artificial intelligence (xai): Toward medical xai.* IEEE transactions on neural networks and learning systems, 32(11), 4793-4813.

de Vries, B. M., Zwezerijnen, G. J. C., Burchell, G. L., van Velden, F. H. P., Menke-van der Houven van Oordt, C. W., and Boellaard, R. (2023) *Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review.* Frontiers in Medicine, 10, 1180773.

Zhukovska, A., Zheliuk, T., Shushpanov, D., Brych, V., Brechko, O., and Kryvokulska, N. (2023) *Management of the Development of Artificial Intelligence in Healthcare.* Proceedings of the 13th International Conference on Advanced Computer Information Technologies (ACIT), 241–247.

Zicari, R.V., Ahmed, S., Amann, J., Braun, S.A., Brodersen, J., Bruneault, F., Brusseau, J., Campano, E., Coffee, M., Dengel, A. and Düdder, B. (2021) *Co-design of a trustworthy AI system in healthcare: deep learning based skin lesion classifier.* Frontiers in Human Dynamics, 3, p.688152.