

XR Buddy: Leveraging Generative AI to Streamline Workflow and Imaging Protocol Access for Radiographers

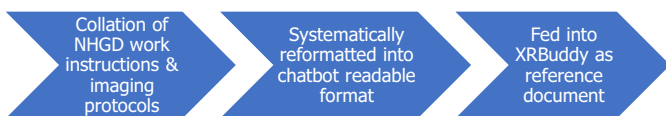
Low Jia Xian Shermaine, Chong Chun Meng, Tan Chia Ling

Problem & Aim

Radiographers often struggle with accessing scattered, site-specific imaging protocols and workflows that are buried deep within the Intranet, potentially leading to inefficiencies and inconsistencies in clinical practice. To address this gap, XR Buddy, a generative AI-powered chatbot built on the Pair platform with a Retrieval-Augmented Generation (RAG) framework was developed to streamline access to NHGD protocols and reference materials. XR Buddy is engineered with strict rules and robust guardrails to ensure responses are strictly drawn from NHGD content. This design mitigates the risk of misinformation and ensures alignment with institutional standards. By delivering accurate, site-specific information in real time, XR Buddy reduces search time, supports new staff and floaters, and promotes standardized imaging practices across NHGD Centres.

Methodologies

1) Content Curation



2) Bot Configuration & Prompts Engineering

Strict guardrails were implemented to control the tone, scope, and safety of responses.

✓ Institution-aligned Responses

- Information drawn from NHGD work instruction & imaging protocol only

🚫 Minimizing Hallucination

- Response word count limit
- No inferencing/ fabricated answers, direct users to colleagues when answer is unavailable

🌐 Cross-Reference Transparency

- Each response anchored with hyperlink directing users to relevant document on Intranet

3) RAG Bot Evaluation

Evaluation 1: Before pilot launch, to establish baseline performance and rating consistency, 30 real user queries from some NHGD junior radiographers were input into XR Buddy. Responses were independently rated by an expert panel (senior to principal radiographers), on a 5-point Likert scale across three key metrics: Faithfulness, Correctness and Relevancy. Mean scores were tabulated.

Evaluation 2: After pilot launch, to further assess real-world usability and ensure governance oversight, some users generated authentic queries and rated the chatbot responses on the same key metrics and rubrics. In parallel, the expert panel reviewed and rated random subset of responses independently.

Pilot sample efficiency test were conducted with some radiographers (n=5) in simulated clinical use to compare protocol retrieval time between manual lookup in Intranet compared to using XR Buddy.

These evaluation approaches were guided by established, validated methods (RAGAS, n.d.; Zhou et al., 2024; Tan et al., 2024; Lee et al., 2024).

Future Works & Impact

XR Buddy exemplifies the potential of generative AI to enhance clinical efficiency and promote standardised practice in radiography. Validated by both junior and expert panel, it streamlines information retrieval, reduce manual searches and support clinical decision-making in NHGD. Ongoing refinement and scalability across modalities, together with expanded testing such as periodic expert audits will further enhance its contextual accuracy and integration as a trusted clinical support tool for radiographers.

Results & Discussion

In evaluation 1 to determine the consistency of the expert panel, inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC), in SPSS software.

Metrics	ICC Scores
Faithfulness	0.804 (Good)
Correctness	0.873 (Excellent)
Relevancy	0.688 (Moderate)

Metrics	Mean Scores (max: 5)
Faithfulness	4.63
Correctness	4.22
Relevancy	4.58

High ICC values for correctness and faithfulness indicated strong agreement among expert raters, demonstrating that XR Buddy consistently produces accurate and source-aligned responses. Moderate ICC for relevancy highlighted variability in contextual interpretation, suggesting an area for future refinement. Additionally, the high mean scores across metrics reinforce XR Buddy's potential to deliver high-quality responses as a reliable clinical support tool.

In evaluation 2, users rated XR Buddy's responses highly, reflecting strong usability and trust. Expert panel oversight confirmed performance with similarly high scores, applying stricter judgement. Together, these findings demonstrate that XR Buddy is both practical and end-users and validated by domain experts, reinforcing governance through dual-layer evaluation.

	Faithfulness	Correctness	Relevancy
User's Mean Scores (max: 5)	4.93	4.70	4.57
Expert Panel Mean Scores (max: 5)	4.80	4.63	4.97

In the pilot sample efficiency test, XR Buddy has shown to reduce the average time to locate protocol and workflow information:

 **5 min → 30 sec**

While preliminary, this suggests a strong potential for improving workflow efficiency and reducing cognitive load during clinical duties, thereby indirectly enhancing patient outcome. Overall, across all evaluations, XR Buddy demonstrated consistently high performance across all three evaluation domains, with particularly strong performance in providing responses that faithfully reflect source material. Using real-world queries and expert panel review strengthened the relevance and validity of the assessment.

References

- RAGAS (n.d.). Available metrics. Retrieved June 24, 2025, from https://docs.ragas.io/en/latest/concepts/metrics/available_metrics/
- Zhou, J., Tan, L. X., & Yeo, W. S. (2024). Evaluating the safety and reliability of large language models in clinical settings: A case study. *Journal of Clinical and Medical Systems*, 48(12), Article 100219. <https://doi.org/10.1016/j.jcms.2024.12.009>
- Tan, H. Y., Lim, A., & Ong, M. J. (2024). Faithfulness and factuality in clinical LLM outputs: A framework for human-aligned evaluation. In *Proceedings of the 2024 ACL Workshop on Dataset and Annotation Standards for Health AI* (pp. 12–21). Association for Computational Linguistics. <https://aclanthology.org/2024.dash-1.2.pdf>
- Lee, C., Ramesh, S., & Wu, T. (2024). SafeBench: A framework for evaluating clinical hallucinations in generative medical AI (arXiv preprint arXiv:2406.14783). arXiv. <https://arxiv.org/pdf/2406.14783>

Acknowledgement

Special thanks to Pair, NHGD T&D team, Expert panel & Junior radiographers for their time and valuable contributions to this project.