Designing Safe Hybrid AI–XR Simulations for Healthcare Communication and Interaction Training

Kristina Mikkonen[1,2](✉), Egle Butkeviciute[3], Patrikas Armalis[3], Evelina Aluzaite[1], Rita Baranauskaite[1], Lukas Paulauskas[3], Lina Spirgiene[1], Andrejus Subocius[1], Tomas Blazauskas[3], and Olga Riklikiene[1]

[1] Lithuanian University of Health Sciences, Kaunas, Lithuania
kristina.mikkonen@oulu.fi
[2] University of Oulu, Oulu, Finland
[3] Kaunas University of Technology, Kaunas, Lithuania

Abstract. Effective healthcare education must address not only clinical knowledge but also communication, teamwork, and decision-making under pressure. Extended Reality (XR) combined with artificial intelligence (AI) offers new opportunities for immersive, scalable communication training, yet raises critical concerns related to safety, reliability, and learner trust. This paper presents the design rationale and early insights from VirtualHealEd, a human-centred hybrid AI–XR simulation for healthcare communication and interaction training. The platform integrates conversational AI, speech recognition, and immersive XR while employing a hybrid architecture that combines generative AI with deterministic system components to ensure safety in life-critical scenarios. In parallel, multimodal data, including speech transcripts, performance indicators, and physiological signals—are collected to explore stress and cognitive load during AI-mediated interactions with healthcare students (n = 52). Early observations suggest that AI-driven dialogue increases learners' workload, with substantial inter-individual variability, highlighting the need for adaptive, stress-aware learning designs. The paper contributes design principles for safe hybrid intelligence in healthcare XR education and outlines challenges and future directions for stress-aware adaptive simulations.

Keywords: Extended reality · Hybrid intelligence · Conversational AI · Healthcare education · Stress-aware learning

1 Introduction

Healthcare education has traditionally focused on discipline-specific clinical knowledge and technical skills [1]. However, modern healthcare environments demand strong non-technical competences, including communication, teamwork, situation awareness, and decision-making under stress. Communication failures remain a significant contributor to adverse events, underscoring the need for structured and scalable communication training solutions.

M. Särestöniemi et al. (Eds.): NCDHWS 2026, CCIS 3009, pp. 289–299, 2026.
https://doi.org/10.1007/978-3-032-28812-7_21

Extended Reality (XR) technologies have emerged as promising tools for immersive healthcare education, enabling learners to engage with realistic clinical scenarios in a safe environment [2]. This is consistent with broader evidence from higher education, where immersive virtual reality has been shown to enhance experiential learning and skill acquisition across disciplines [3]. When combined with artificial intelligence (AI), XR simulations can move beyond scripted scenarios toward adaptive, interactive learning experiences. Conversational AI, in particular, enables natural language interaction with virtual patients and team members, increasing realism and learner engagement.

While prior AI-driven XR training systems have demonstrated benefits in simulation realism and learner engagement, many remain limited to pre-scripted dialogue trees or constrained interaction models, reducing their adaptability and ecological validity. In contrast, recent advances in large language models enable more open-ended communication, but introduce new challenges related to reliability, controllability, and pedagogical alignment.

Despite this potential, the use of AI in healthcare education introduces critical challenges [4]. Large language models may generate inconsistent or incorrect information, posing risks when simulations involve medical instructions or treatment decisions. Such limitations are widely discussed in AI safety research, particularly regarding hallucinations, lack of grounding, and risks in high-stakes contexts [5]. Additionally, immersive AI-driven interactions may increase cognitive load and stress, potentially affecting learning outcomes. These risks are well-documented in broader AI safety literature, including concerns related to hallucinations, alignment, and robustness in high-stakes domains. These challenges highlight the importance of hybrid intelligence approaches that balance AI flexibility with human-centred design, transparency, and safety [6]. Hybrid intelligence emphasises the complementary strengths of human and artificial cognition, enabling more reliable and context-aware decision-making in complex environments [7].

This paper presents VirtualHealEd, an AI-powered XR simulation designed for healthcare communication and interaction training. The novelty of this work lies in its integration of a hybrid intelligence architecture that constrains generative AI through safety-aware design mechanisms, while preserving conversational flexibility within XR-based clinical scenarios, operationalising hybrid intelligence principles in an immersive learning context [7]. Rather than focusing solely on system implementation, the paper emphasises (1) the design rationale behind a hybrid AI–XR architecture, (2) safety-driven decisions that constrain AI behaviour in critical contexts, and (3) early observations from multimodal data, including physiological indicators of stress. In doing so, the study bridges healthcare education research with emerging technical and ethical frameworks in AI and XR systems design. The goal is to contribute design insights for the responsible and effective use of hybrid intelligence in healthcare education. VirtualHealEd has been approved by Kaunas Regional Biomedical Research Ethics Committee (Nr. 2025-BE10–0011).

2 Related Work

XR-based simulations are increasingly used in healthcare education to support clinical skills, teamwork, and communication training. Prior studies have shown that immersive simulations can enhance learner engagement and improve transfer of skills to clinical practice [1, 3]. Virtual patients, in particular, have been used to train communication skills in nursing and medicine [2].

Recent advances in conversational AI have enabled more dynamic virtual patient interactions. Large language models allow simulations to respond flexibly to learner input, supporting unscripted dialogue and personalised learning paths. However, concerns have been raised regarding the reliability, explainability, and safety of AI-generated content in medical contexts [4, 5, 8].

In parallel, research on stress and cognitive load in simulation-based learning has highlighted the importance of physiological measures such as heart rate variability and EEG-derived metrics. These measures provide insights into learner states that may not be captured through performance metrics alone. Integrating such data into XR environments remains challenging due to sensor noise, synchronisation issues, and individual variability.

While prior work has explored XR, conversational AI, and stress measurement separately, fewer studies have addressed how these components can be combined within a unified, safety-aware hybrid intelligence framework. This paper contributes to this gap by presenting design choices and early insights from an integrated AI–XR healthcare education platform.

3 Hybrid AI–XR Design Rationale

3.1 Human-Centred Hybrid Intelligence

The design of VirtualHealEd is grounded in the principles of hybrid intelligence, where human pedagogical intent and system constraints guide the use of AI. Rather than maximising AI autonomy, the system emphasises transparency, controllability, and alignment with educational objectives. AI is used to enhance realism and feedback, while critical decisions remain bounded by deterministic logic.

This approach reflects a core assumption: in healthcare education, realism must never come at the expense of safety or correctness. Consequently, AI-generated content is selectively constrained based on the educational and ethical implications of each interaction.

3.2 Two-Mode Communication Training Design

VirtualHealEd implements two complementary learning modes that support progressive skill development. The VR was developed for anaphylactic shock management, clinical reasoning and decision making in urgent situations, with the focus of communication among professionals and patients (NPC) themselves, including students in a clinical practice context. The first mode focuses on AI-assisted scripted communication practice[5]. Learners read from a teleprompter while local speech recognition transcribes their utterances in real time. Recognised phrases are compared against an expected script, and successfully completed segments are visually marked. This mode supports the acquisition of correct medical terminology and structured communication sequences, particularly for novice learners.

The second mode enables unscripted conversational interaction with virtual patients and healthcare professionals. In this mode, learners engage in natural dialogue, asking questions and responding to dynamic character behaviour (Fig. 1). This design supports higher-level communication skills, such as empathy, clarification, and adaptive questioning, which are difficult to train through scripted approaches alone.

Together, these modes provide pedagogical scaffolding, allowing learners to transition from structured practice to realistic, open-ended interaction.

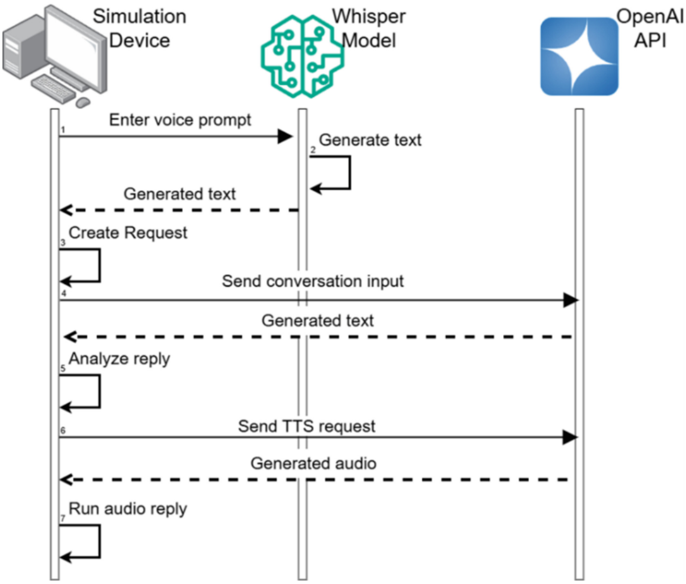

Fig. 1. Simplified AI conversation sequence diagram.

3.3 Safety-Critical Hybrid Architecture

A defining feature of the system is its hybrid AI architecture. While conversational AI is used for patient and nurse interactions, scenarios involving life-critical medical instructions—such as medication dosages or emergency protocols—are handled through deterministic dialogue paths with pre-recorded audio.

This design decision addresses the well-documented risk of AI hallucinations. Even minor inaccuracies in medication information could reinforce unsafe practices. By isolating AI-generated dialogue from safety-critical content, the system ensures absolute correctness while preserving conversational flow.

The architecture separates speech recognition, language generation, response analysis, and speech synthesis into modular components. This modularity supports future upgrades and allows AI components to be replaced or constrained without redesigning the entire system.

4 Stress-Aware Data Collection and Early Observations

A total of 52 fourth-year healthcare students participated in the study, the majority of whom were female (96.2%), with a mean age of 23.1 years (SD = 5.28; median = 22; range 21–51). More than half of the participants (55.8%) reported having work experience in healthcare settings ranging from 2 to 72 months. Among those who specified their experience, the average duration was 16.8 months (SD = 15.61), with a median of 12 months.

To support stress-aware analysis of AI-mediated XR learning, VirtualHealEd implements a multimodal data processing pipeline, illustrated in Fig. 2. The process begins with the learner engaging in an XR-based clinical scenario (scripted or AI-powered), during which multimodal data are captured, including wearable physiological signals (EEG, ECG, PPG), audio for speech recognition, VR and screen recordings, optional room video, and self-report questionnaires. These heterogeneous data streams are ingested through a validation and synchronisation layer that ensures data integrity, metadata extraction, and temporal alignment across devices and modalities. Signal processing is then applied to each stream, including filtering and feature extraction for cardiovascular and neurophysiological signals, speech and task-related metrics, and optional video-based event analysis. The processed data are subsequently analysed using descriptive statistics, robust inferential tests, and offline machine learning models to explore stress and workload patterns at both state and subject levels. Finally, results are visualised through dashboards and research reports, with a forward-looking design that supports future real-time adaptation of XR scenarios and feedback based on learner state.

4.1 Multimodal Data Sources

In addition to conversational interaction, VirtualHealEd collects multiple data streams to explore learner workload and stress during immersive XR training. These include speech-to-text transcripts, task and interaction events, and physiological signals recorded via wearable sensors. Neurophysiological and cardiovascular data—specifically EEG, ECG, and PPG—are collected during XR-based clinical simulations and synchronised with interaction timestamps and scenario states.

At the current stage, physiological signals are processed offline using standard preprocessing, filtering, and feature extraction procedures, followed by within-subject normalisation to account for individual baseline differences. Extracted features include heart rate and heart rate variability measures (e.g., RMSSD, SDNN), EEG spectral power indices, and speech- and task-related temporal metrics. The primary aim of this multimodal data collection is to examine how different interaction modes—resting baseline, scripted non-AI interaction, and AI-powered conversational XR—relate to physiological responses indicative of cognitive workload and stress.

Fig. 2. Multimodal stress-aware learning pipeline in VirtualHealEd.

Integrating these heterogeneous data sources presents several challenges, including sensor noise and motion artefacts in immersive XR environments, temporal synchronisation across devices, and substantial inter-individual variability in physiological responses. These challenges motivate the use of within-subject analytical approaches and cautious interpretation of group-level trends.

4.2 Early Observations

Preliminary physiological analyses reveal systematic differences between learning conditions, with AI-driven conversational XR interaction emerging as the most demanding state.

Table 1. Within-participant differences in PPG-derived heart rate and HRV across states: Friedman test statistics and effect sizes (Kendall's W).

Parameter	Friedman's chi-square	p value	Kendall's W
HR (PPG)	35.70	**<0.001**	0.54
RMSSD (PPG)	6.73	**0.03**	0.10
SDNN (PPG)	2.62	0.27	0.04

Repeated-measures comparisons across resting, scripted, and AI-powered conditions indicate a significant increase in heart rate during the AI condition ($p < 0.05$), accompanied by a significant reduction in heart rate variability (RMSSD, $p < 0.05$), reflecting increased sympathetic activation and reduced parasympathetic regulation.

In contrast, EEG-derived measures, including relative theta power, were analysed using Friedman test, with statistical significance defined as $p < 0.05$, and did not show statistically significant differences across conditions. Visual inspection (see Fig. 3) suggested substantial inter-individual variability, with widely scattered values across participants, which limits the interpretation of group-level EEG trends in the present sample.

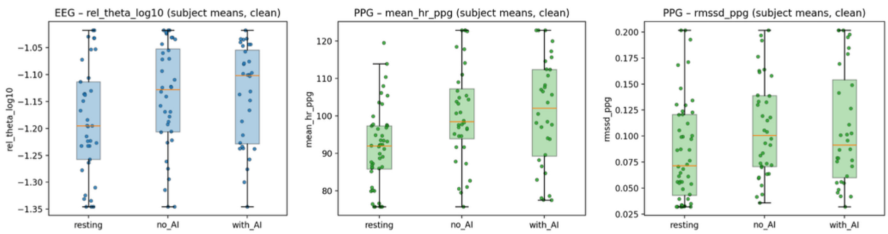

Fig. 3. Subject-level distributions of physiological markers across learning conditions. Boxplots show mean relative EEG theta power (log10-transformed), mean heart rate derived from PPG, and RMSSD derived from PPG for resting baseline, scripted non-AI interaction and AI-powered conversational XR conditions. Individual data points represent subject means. The AI-powered condition exhibits higher heart rate, reduced heart rate variability, and elevated theta power relative to resting and scripted conditions, indicating increased cognitive and physiological workload, with substantial inter-individual variability.

$\text{Load} = z(HR) - z(\log(\text{RMSSD})) - z(\log(\text{SDNN}))$

Cardiovascular responses were compared across conditions using PPG-derived heart rate and HRV indices(with z denoting standardisation) Load index was constructed from HR, RMSSD, and SDNN to reflect complementary aspects of cardiovascular regulation rather than to retain only individually significant markers. HR captures overall cardiovascular activation, RMSSD primarily reflects short-term parasympathetic modulation, and SDNN reflects overall beat-to-beat variability across the recording window [9]. Although SDNN did not show a significant condition effect (see Table 1) in the present sample, it was retained in the composite as a theory-informed component of global autonomic variability and to avoid over-representing a single HRV facet. This is particularly relevant in exploratory multimodal studies with substantial inter-individual variability, where some markers may contribute more clearly at the composite or subject-specific level than in group-level univariate tests. In future analyses with larger samples, the relative contribution of SDNN will be re-evaluated and may become more informative as statistical power increases. As shown in Fig. 3, the AI-powered condition was associated with higher heart rate and lower RMSSD than the resting and scripted conditions, although substantial inter-individual variability remained evident.

Effect magnitude for the Friedman test was quantified using Kendall's W, where values closer to 0 indicate weak agreement (small effects) and values closer to 1 indicate strong agreement (large effects) across repeated measures. Heart rate showed a large effect ($W \approx 0.54$), whereas RMSSD showed a small effect ($W \approx 0.10$) and SDNN showed a negligible effect ($W \approx 0.04$) (see Table 1).

Post-hoc comparisons indicate higher load values during AI-powered interaction relative to both resting and scripted conditions, with small-to-moderate within-subject effect sizes. These trends align with participant self-reports, although considerable inter-individual variability is observed. To summarise these cardiovascular changes at the subject level, a composite physiological load index was calculated (Fig. 4).

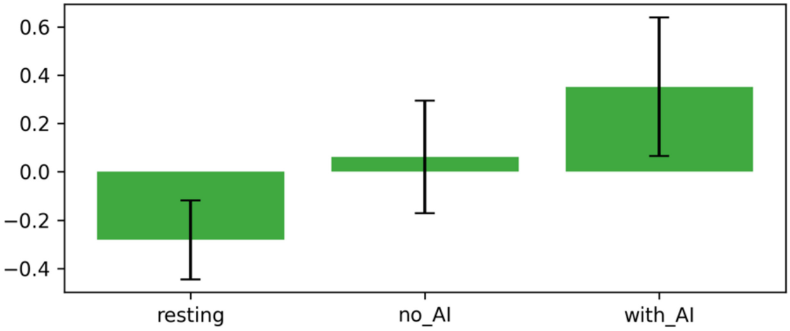

Fig. 4. Composite physiological load index across resting, scripted non-AI, and AI-powered XR conditions. Bars represent condition means and error bars indicate inter-individual variability. The right panel presents the corresponding confusion matrix for three stress classes (no stress, low stress, high stress), illustrating above-chance classification performance with notable overlap between adjacent stress levels.

Complementing the physiological analyses, offline machine learning models trained on multimodal physiological features demonstrate the feasibility of classifying stress states above chance level. A tuned Random Forest classifier identifies cardiovascular features, particularly PPG-derived heart rate and variability measures, as the most informative predictors. However, confusion matrix analysis reveals substantial overlap between low- and high-stress classes, indicating graded physiological transitions rather than clearly separable stress states. The feasibility of stress-state classification based on multimodal physiological features is illustrated in Fig. 5.

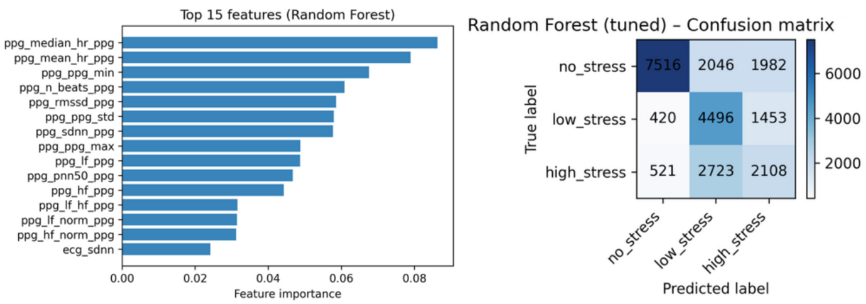

Fig. 5. Offline machine learning results for physiological stress classification. The left panel shows the top 15 feature importances from a tuned Random Forest model, highlighting the dominance of PPG-derived heart rate and heart rate variability features. The right figure demonstrates the performance of the Random Forest classifier within three states (resting, low_stress, high_stress) with up to 70% accuracy.

Overall, these findings should be interpreted as exploratory. The observed variability underscores the limitations of group-level averages and highlights the importance of within-subject analysis in stress-aware learning research. Rather than supporting causal claims regarding learning effectiveness, the results motivate future work toward real-time, individualised adaptation of XR learning environments based on multimodal stress indicators.

5 Design Implications for AI-Driven Healthcare Education

Based on the design experience and early observations, several implications emerge for the development of AI-driven XR healthcare education systems. First, AI realism must be balanced with pedagogical and ethical safety. Fully generative systems may enhance immersion but require strict constraints in medical contexts. Second, hybrid architectures that combine AI-driven flexibility with deterministic safety mechanisms provide a practical pathway for responsible deployment in healthcare education. Third, stress and workload are integral aspects of immersive learning and should be considered design parameters rather than unintended side effects. Multimodal sensing offers valuable insights but must be interpreted cautiously.

Finally, language and cultural context significantly shape AI feasibility. Implementing conversational AI in less-resourced languages requires careful model selection and prompt engineering, influencing system performance and latency.

6 Limitations and Future Work

The current work has several limitations. Data collection is ongoing, and the physiological analyses should therefore be considered exploratory in nature. Stress indicators are analysed offline, and adaptive mechanisms are not yet implemented. Additionally, latency in AI response generation affects conversational flow.

Future work will focus on larger-scale studies, improved signal synchronisation, and the development of real-time adaptive mechanisms that adjust scenario complexity or feedback based on learner state. Advances in locally executable language models may further enhance privacy and performance.

7 Conclusion

This paper presented the design rationale and early insights from VirtualHealEd, a hybrid AI–XR simulation for healthcare communication training. By combining conversational AI with deterministic safety mechanisms and stress-aware data collection, the system illustrates how hybrid intelligence can support realistic yet responsible healthcare education. The findings highlight both the promise and the challenges of AI-driven immersive learning and provide design principles for future stress-aware, human-centred XR training environments.

Acknowledgements. Project"EnhancingHealthcareEducationwithHuman-CentredExtendedVirtualReality–VirtualHealEd"hasreceived-fundingfromtheResearchCouncilofLithuania (LMTLT) underGrantNo.S-MIP-24-132. We would like to acknowledge the support of The University of Oulu & The Academy of Finland Profi 7 (352788) for also supporting this study. Finally, we gratefully acknowledge the collaboration with FrostBit Software Lab at Lapland University of Applied Sciences, whose expertise in emerging technologies and extended reality contributed valuable insights and support to this work.

Disclosure of Interests ChatGPT (version 5.3, OpenAI) was used to support linguistic revision of the manuscript. The authors take full responsibility for the content.

References

1. Ropponen, P., Tomietto, M., Pramila-Savukoski, S., Kuivila, H., Koskenranta, M., Liaw, S.Y., et al.: Impacts of virtual reality simulation on nursing students' competence, confidence, and satisfaction: a systematic review and meta-analysis of randomised controlled trials. Nurse Educ. Today. **152**, 106756 (2025). https://doi.org/10.1016/j.nedt.2025.106756
2. Mikkonen, K. et al.: Multidimensional pedagogical framework for interprofessional education: blending classroom, high fidelity and extended reality simulation. Nurse Educ. Today. (2025). https://doi.org/10.1016/j.nedt.2025.106838
3. Radianti, J., Majchrzak, T.A., Fromm, J., Wohlgenannt, I.: A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda. Comput. Educ. **147**, 103778 (2020)
4. Graham, A., Hökkä, M., Tomietto, M., Mikkonen, K.: The use of digital gamification, extended reality, artificial intelligence, and other digital learning tools in palliative care education of undergraduate nurses: a systematic review of mixed-methods. Nurse Educ. Today. (2026). https://doi.org/10.1016/j.nedt.2026.106982
5. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT), pp. 610–623 (2021)
6. Kovalainen, T. et al.: Utilising artificial intelligence in developing education of health sciences higher education: an umbrella review of reviews. Nurse Educ. Today. **147**, 106600 (2025). https://doi.org/10.1016/j.nedt.2025.106600
7. Dellermann, D., Ebel, P., Söllner, M., Leimeister, J.M.: Hybrid intelligence. Bus. Inf. Syst. Eng. **61**, 637–643 (2019)
8. Mikkonen, K. et al.: How does human-centred extended reality support healthcare students' learning in clinical conditions? In: Särestöniemi, M. et al. (eds.) Digital Health and Wireless Solutions (NCDHWS 2024). Commun. Comput. Inf. Sci., vol. 2083, pp. 169–184. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-59080-1_13
9. Wang, B.X., Brennand, E.E., Le Page, P., Mitchell, A.R.J.: Heart rate variability in cardiovascular disease diagnosis, prognosis and management. Front. Cardiovasc. Med. **12**, 1680783 (2025). https://doi.org/10.3389/fcvm.2025.1680783