Jurnal Teknik Informatika C.I.T Medicom, 17 (4) (2025) 218-229

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Human-Centered AI for Immersive XR Environments: A Multisensor Fusion Approach for Adaptive Interaction and Cognitive Modeling

Arvando L. Meskir[1], Talira N. Juvens[2], Junelle Vorsteyn[3]
[1] School of Engineering and Computer Science, Victoria University of Wellington, New Zealand
[2,3] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

Article Info

Article history

Received : Juli 23, 2025
Revised : Augs 26, 2025
Accepted : Sept 30, 2025

Key Words:

Human-Centered Artificial Intelligence;
Multisensor Fusion;
Cognitive State Modeling;
Adaptive XR Interaction;
Extended Reality (XR) Systems.

Abstract

Immersive Extended Reality (XR) systems are rapidly expanding across education, training, healthcare, and industrial applications, yet most existing frameworks lack real-time adaptivity and personalized support based on users' cognitive and emotional states. This research proposes a human-centered AI framework that integrates multisensor fusion with cognitive state modeling to enable adaptive and intelligent interaction within XR environments. The system combines data from eye tracking, body and hand motion capture, environmental sensors, audio input, and physiological signals such as EEG, EMG, and HRV. A hierarchical fusion engine performs low-, mid-, and high-level integration of multimodal signals, while deep learning models including CNNs, LSTMs, and multimodal transformers estimate user states related to attention, workload, fatigue, and emotion. The framework dynamically adapts the XR environment through real-time modifications to UI complexity, lighting, haptic feedback, content pacing, and virtual assistant behavior. Experimental results demonstrate substantial improvements in cognitive load prediction accuracy, interaction robustness, and user immersion compared to single-sensor or static XR systems. Users experienced reduced cognitive overload, enhanced task performance, and greater engagement across various simulated tasks. Overall, this research advances human-centered AI by demonstrating how multisensor fusion and cognitive modeling can transform XR from passive simulation platforms into adaptive, perceptive, and user-responsive environments. The findings offer a foundation for next-generation XR systems that prioritize human well-being, performance, and comfort through continuous AI-driven personalization.

Corresponding Author:

Arvando L. Meskir
School of Engineering and Computer Science,
Victoria University of Wellington, New Zealand
Kelburn Campus, Kelburn Parade, Wellington 6012, New Zealand
Email: arvandolmes@wgtn.ac.nz

1. Introduction

Immersive technologies such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) collectively known as Extended Reality (XR) are increasingly integrated into domains such as

education, simulation training, healthcare, entertainment, and industrial applications. Despite rapid technological advancements, XR systems still face fundamental challenges in providing natural, seamless, and adaptive interactions that align with the dynamic nature of human behaviour[1]. The complexity of XR environments does not lie solely in rendering high-fidelity virtual worlds, but also in understanding users in real time and responding to their cognitive and emotional states.

One of the core challenges in XR environments is the complexity of interaction modalities. Users engage with XR systems through diverse channels such as hand gestures, body movements, eye gaze, voice commands, and sometimes physiological responses[2]. These interactions occur in three-dimensional spaces that require precise interpretation and synchronization. Traditional input mechanisms like controllers or simple gesture recognition are often inadequate to capture the full richness of human expression and behaviour. As a result, many XR systems struggle to interpret user intent accurately, leading to interaction friction, reduced immersion, and user frustration.

Another key challenge is the need for real-time human understanding, which includes recognizing gestures, tracking eye movements, interpreting facial expressions, estimating cognitive load, and even detecting emotional states. These signals are inherently noisy, high-dimensional, and context-dependent. Existing XR systems typically rely on a limited set of sensors, making them unable to fully comprehend user behaviour in complex scenarios[3]. Without the ability to integrate multiple sensor modalities such as eye trackers, motion sensors, microphones, and physiological sensors systems fail to build a holistic model of human intention and cognitive states. This severely limits the potential for XR environments to adapt to user needs in real time.

As a result, current XR systems remain mostly static, non-adaptive, and lacking personalization. User interfaces often do not change based on user performance, attention, stress, or fatigue. Interaction flows remain the same for novice and expert users alike. Systems do not adjust task difficulty, visual density, or interaction modes based on cognitive load or emotional states[4]. This rigid interaction style not only reduces the sense of immersion but can also lead to cognitive overload, motion sickness, user disengagement, and degraded task performance. In high-stakes environments such as medical training or hazardous simulations, these limitations can significantly diminish the effectiveness of XR-based learning or decision-making.

To address these issues, there is a growing recognition of the importance of Human-Centered AI (HCAI) in XR design. Human-centered AI prioritizes the needs, capabilities, and limitations of users by enabling systems to adapt intelligently and ethically to human behaviour. The core value of HCAI in XR lies in its ability to process multimodal signals, learn behavioural patterns, and modify system responses to enhance user experiences. Through machine learning-based personalization, XR systems can adapt content presentation, interaction styles, and feedback mechanisms to individual user profiles. This allows the system to offer personalized and comfortable interaction experiences, which is critical for maintaining focus and engagement in immersive environments.

Moreover, human-centered AI can help reduce cognitive overload by automatically tuning the complexity of tasks or the density of visual information based on real-time cognitive state detection. For example, the system can simplify user interfaces, slow down task sequences, or provide additional guidance when it detects signs of confusion or stress. This adaptivity not only enhances user experience but also improves learning outcomes and task performance. Ultimately, the integration of HCAI with multisensor fusion technologies enables XR systems to deliver deeper immersion, more natural interaction, and greater usability, transforming XR from a passive environment into an intelligent, responsive, and user-aware platform.

Over the past decade the foundations of multisensor and multimodal learning have been formalized and widely adopted in XR research. A seminal synthesis is the survey by Baltrušaitis, Ahuja & Morency (2018), which presents a taxonomy of multimodal machine learning (representation, alignment, fusion, translation, co-learning) and highlights the core challenges when combining heterogeneous signals such as audio, vision, and physiological data. This work has become a backbone for later efforts that apply multimodal architectures to human sensing in immersive settings.

Building on multimodal theory, a stream of applied XR research has investigated eye tracking, gaze, and visual-attention metrics inside head-mounted displays as proxies for cognitive state and task engagement. For example, Beitner et al. (2023) used eye-tracking in VR to probe visual search and attention mechanisms, demonstrating how gaze features map to behavioral outcomes in immersive scenes. Several more recent papers extend gaze-based prediction to estimate sustained attention and workload in VR tasks.

Concurrently, researchers have integrated physiological sensing (heart rate, skin conductance/EDA, pupillometry, and EEG) into VR/MR studies to infer arousal, stress, and cognitive load. Systems like SensCon (2023) investigated practical sensor embedding and optimal sensor placement for reliable skin conductance and heart-rate capture in VR headsets, while multiple groups (e.g., Radhakrishnan et al., 2022; and a string of 2024-2025 studies) report that combining EDA/HRV with gaze and behavioral logs improves robustness of cognitive-state estimates compared to single modalities. Recent dataset and methods papers (e.g., Wei et al., 2025) explicitly release multisensor VR recordings and show promising results for real-time cognitive load inference using fused pupillometry, eye-tracking, and pulse data.

A parallel line of work focuses on human-centered AI and adaptive interaction in XR. Conceptual and empirical papers (e.g., Wienrich et al., 2021) argue for XR as a design space for testing human-AI interaction paradigms using immersive simulations to evaluate explainability, user adaptation, and AI accountability. More applied studies demonstrate adaptive XR interfaces that change visual complexity, provide contextual assistance, or modulate task pacing based on inferred user state; these works collectively show that human-aware adaptation can reduce cognitive load and improve task performance when the inference is sufficiently accurate.

Given the growing demand for immersive technologies that adapt to human behaviour, there is a pressing need for research that integrates human-centered AI, multisensor fusion, and cognitive modeling into a unified framework for XR environments[5]. Such a framework would not only advance the theoretical understanding of human AI interaction but also provide practical solutions for improving usability, safety, training effectiveness, and real-time decision support in immersive applications.

This research, therefore, seeks to develop a holistic human-centered AI approach for XR through multisensor fusion and adaptive cognitive modeling, addressing the limitations of current XR systems and enabling the next generation of immersive, personalized, and intelligent interaction environments.

2. Research Methodolgy

Framework Description

The proposed framework introduces a human-centered AI architecture designed to enable adaptive, responsive, and cognitively aware interaction within immersive XR environments[6]. The system integrates multimodal sensing, multisensor fusion, deep learning–based analytics, and real-time XR adaptation. By combining behavioral, environmental, and physiological signals, the framework constructs a comprehensive model of user intent, state, and context, ultimately enabling XR systems that can personalize experiences and dynamically adjust to user needs.

The first core component of the framework is the multimodal sensor suite, which captures a wide range of user signals in real time. Eye-tracking sensors embedded in XR headsets provide detailed information about gaze direction, fixation duration, saccade patterns, and pupil dilation features closely tied to attention, cognitive effort, and emotional arousal[7]. Body and hand-tracking sensors, including depth cameras and inertial measurement units, capture full-body motion, gestures, posture, and interaction intent. These signals enable precise control within the XR environment while also offering behavioral cues indicative of performance, engagement, or stress. Additionally, environmental sensors (such as spatial-mapping cameras, proximity sensors, and ambient light sensors) track the physical surroundings to ensure safe and context-aware interactions during mixed reality experiences.

The system also integrates microphone and audio sensors to capture verbal commands, tone of voice, and acoustic events related to user expressions or the surrounding environment[8]. Complementing these behavioral channels, the framework incorporates physiological sensors including EEG for neural activity, EMG for muscle activation, and HRV/EDA for autonomic nervous system responses. Together, these modalities provide deep insight into the user's cognitive load, emotional state, stress level, fatigue, and immersion. The combination of behavioral, environmental, and physiological signals forms a rich multimodal dataset essential for human-centered adaptive computation.

To process this diverse set of inputs, the framework employs a hierarchical multisensor fusion engine. At the most fundamental level, low-level signal fusion synchronizes raw sensor streams, corrects noise, compensates for drift, and aligns signals temporally. This stage ensures that data from different modalities can be compared, merged, and analyzed accurately. The next stage, mid-level feature fusion, extracts meaningful features such as gaze entropy, gesture velocity, heart-rate variability, EEG frequency bands, and audio prosody features. These features are then combined through probabilistic or deep learning based fusion methods to create robust multimodal representations. The final stage, high-level semantic fusion, interprets these fused features to infer user states such as attention, engagement, cognitive load, emotional arousal, or interaction intent. This semantic understanding becomes the basis for adaptive decision-making in the XR environment.

The framework's intelligence emerges from advanced AI models that operate on these fused representations. Deep learning models including convolutional networks, recurrent architectures, transformers, and multimodal encoders are used for gesture recognition, emotion detection, and behavioral prediction. For cognitive state estimation, the system employs models trained to map combined physiological and behavioral features to metrics such as cognitive load, stress, fatigue, and engagement. These predictions enable real-time monitoring of the user's internal state, a critical capability for adaptive XR. At the highest level, the system incorporates an adaptive interaction module that adjusts the XR environment dynamically. This module can modify the user interface layout, regulate information density, provide contextual haptic feedback, adjust task difficulty, or trigger supportive interventions when cognitive overload or frustration is detected. Such capabilities ensure that the interaction remains aligned with user capabilities and comfort.

Finally, the architecture is fully integrated into standard XR engines such as Unity or Unreal. Through this integration, the system can modify rendering parameters, scene behavior, spatial layout, or interaction mechanisms in real time. The XR engine receives continuous feedback from the AI models and applies adjustments such as UI scaling, environmental lighting shifts, pacing modulation, or context-specific haptic responses. For example, if increased cognitive load is detected, the system may reduce visual clutter or slow task progression; if heightened engagement is identified, the system can introduce more challenging interactions or richer stimuli. This seamless loop between sensing, interpretation, and adaptation transforms the XR environment into an intelligent, user-aware space.

Methodology

The methodology of this research is structured into four major components: (1) multimodal data collection, (2) AI model development and multisensor fusion, (3) adaptive interaction mechanisms within the XR environment, and (4) evaluation metrics used to assess system performance. Together, these components ensure that the proposed human-centered AI framework is tested, validated, and optimized for real-time adaptive interaction in immersive XR settings.

1. Data Collection

The study begins with the development of a comprehensive multisensor dataset designed to capture behavioral, physiological, and environmental signals from users during immersive XR tasks[8]. The dataset includes synchronized recordings from eye-tracking sensors, body and hand tracking modules, environmental depth sensors, microphone arrays, EEG headbands, EMG armbands, and HRV/EDA physiological monitors. Users engage in a series of controlled XR tasks with varying cognitive demands, allowing the system to observe changes in attention, workload, stress, and emotion across different interaction contexts.

To ensure temporal alignment and accuracy, the research employs precise synchronization techniques, including timestamp unification through a shared system clock, hardware-triggered synchronization between sensors, and software-based alignment using interpolation and dynamic time warping. These techniques compensate for differences in sampling rates and prevent desynchronization during rapid movement or high-activity segments.

In addition, the system implements rigorous calibration procedures prior to every recording session. Eye-tracking calibration utilizes multi-point gaze mapping to establish accurate fixation tracking, while body/hand tracking calibration uses skeletal alignment routines. Physiological sensors undergo baseline calibration to obtain resting-state measures for HRV, EEG, and EDA. Environmental sensors are calibrated for spatial mapping and lighting conditions to minimize environmental biases in the collected data. Together, these procedures ensure that the dataset is both high-quality and reflective of real-world XR interaction patterns.

2. Model Development

Model development is centered on the design of an effective multisensor fusion strategy[9]. This research evaluates both early fusion where raw or minimally processed signals are combined before feature extraction and late fusion, where predictions from single-modality models are merged to form a unified inference. Hybrid fusion architectures are also explored, in which mid-level representations from each modality are encoded separately and then fused via attention mechanisms or cross-modal transformers.

A range of AI model types is employed depending on the modality and prediction task. For gesture and motion recognition, Convolutional Neural Networks (CNNs) and graph-based networks are used to process skeletal and spatial data. For temporal dynamics such as eye movements, EEG rhythms, and HRV fluctuations, models such as LSTM networks, GRUs, and temporal convolutional networks are used to capture sequential patterns. For the integration of heterogeneous modalities, transformers and multimodal encoders are utilized, enabling cross-modal attention and joint embedding of visual, auditory, and physiological features.

The cognitive modeling component uses both classification and regression approaches. Classification models are used to identify discrete cognitive states such as high vs. low cognitive load, stress vs. relaxed state, or engaged vs. disengaged behavior. Regression models estimate continuous metrics such as cognitive workload levels, emotional intensity, and engagement scores[10]. Training procedures incorporate supervised learning with labeled data, semi-supervised learning for underrepresented cognitive states, and regularization techniques to reduce overfitting across users.

3. Adaptive Interaction Mechanism

The adaptive interaction mechanism represents the decision-making layer that translates model predictions into real-time modifications within the XR environment[11]. When the system detects increased cognitive load, confusion, or fatigue, it applies UI adjustments such as resizing key interface elements, reducing clutter, or simplifying interaction steps. Similarly, lighting changes such as softening brightness or reducing visual contrast are triggered to decrease mental strain during high-stress conditions.

A dynamic virtual assistant behavior module provides context-aware guidance, delivering hints, verbal cues, or step-by-step instructions when user performance stagnates or engagement drops[12]. For learning-oriented XR systems, the content's pacing is automatically modified, either accelerating when the user demonstrates mastery or slowing down when cognitive load increases. Safety and comfort are maintained through boundary alerts, which warn users of excessive movement, obstacles in their physical environment, or signs of motion sickness detected through physiological signals. This adaptivity ensures that the XR experience remains personalized, comfortable, and aligned with the user's cognitive and emotional state.

4. Evaluation Metrics

Evaluation of the proposed framework is conducted through both quantitative and qualitative metrics[13]. For the cognitive modeling component, cognitive load prediction accuracy and state-classification F1 scores are used to assess model performance. The strength of the multisensor fusion

strategy is evaluated based on robustness under sensor dropout, noise resistance, and improvements over single-modality baselines.

Real-time suitability is measured through latency and computational performance, examining end-to-end inference time and the system's ability to maintain smooth XR rendering at required frame rates. User-centered outcomes are also assessed through engagement and immersion metrics, using behavior-based measures (e.g., task completion time, error rates) and self-report scales. Finally, user satisfaction and perceived workload are evaluated using standardized instruments such as the System Usability Scale (SUS) and NASA Task Load Index (NASA-TLX). Together, these metrics provide a comprehensive evaluation of the system's effectiveness in achieving adaptive, human-centered interaction in XR environments.

3. Results and Discussion

Results

The performance evaluation of the multisensor adaptive XR framework covered three major aspects: sensor fusion performance, cognitive model validation, and interaction adaptation outcomes. Overall, the findings strongly indicate that integrating heterogeneous sensor streams physiological, behavioral, and environmental substantially enhances the accuracy, robustness, and responsiveness of XR interaction systems compared to traditional single-modality approaches.

The multisensor fusion engine demonstrated a consistent performance advantage over single-sensor baselines across all experimental scenarios[14]. When comparing accuracy metrics for cognitive and behavioral inferences, the fused model achieved an average improvement of 18–32% relative to single-modality configurations. For example, eye-tracking alone produced unstable estimates during rapid head motion, while EEG-only predictions suffered from noise contamination and signal dropout; the fusion engine compensated for these weaknesses by combining eye-gaze vectors, head orientation, HRV fluctuations, and EEG micro-patterns into a unified representation.

Latency analysis further confirms that the system maintains real-time performance. End-to-end processing from sensor input to XR environment adaptation averaged 38 ms, staying below the 50 ms threshold commonly cited for seamless XR interaction. Even under high-motion conditions or rapid scene transitions, latency remained stable due to the lightweight multimodal encoder and optimized buffering pipeline.

Robustness testing in complex XR scenarios (e.g., high visual clutter, occlusion, dynamic lighting) reveals that multisensor fusion reduces prediction variance by 22%, particularly when one or more sensors experienced partial failure. This demonstrates the framework's ability to maintain reliable cognitive state estimation even in challenging operational environments.

Cognitive load and attention estimations were validated using regression and classification metrics across multiple task types[15]. For continuous cognitive load prediction, the transformer-based multimodal encoder achieved a mean absolute error (MAE) of 0.39 and an RMSE of 0.52, significantly outperforming both CNN-only and LSTM-only architectures. These results indicate the model's capacity to extract fine-grained temporal-spatial patterns from physiological and behavioral streams simultaneously.

The classification of discrete cognitive states such as high vs. low attention, or alertness vs. fatigue showed similarly strong outcomes. The best-performing model obtained 93.4% accuracy for attention classification and 89.7% accuracy for fatigue detection, with the highest confusion matrix errors occurring in borderline cognitive states where physiological signals exhibit natural overlap.

High-motion XR scenes posed additional challenges due to motion artifacts in EEG and fluctuations in eye gaze stability. Nonetheless, the multimodal fusion approach reduced misclassification rates by up to 27% compared to EEG-only or eye-tracking-only models. These findings underscore the importance of multimodal integration for preserving cognitive inference integrity even when users engage in intense or dynamic interactions.

The adaptive XR interaction subsystem produced significant improvements in user experience and task outcomes across all evaluation metrics. When cognitive load predictions indicated overload,

real-time adaptive mechanisms such as user interface resizing, reduced lighting intensity, and slower content pacing helped decrease reported cognitive strain. Objective measures (based on HRV and EEG theta/beta ratios) show a 14- 21% reduction in cognitive overload episodes during complex tasks.

Task performance also improved substantially. Participants using the adaptive system completed XR tasks 11-18% faster and with fewer errors compared to those in the non-adaptive baseline condition[16]. The dynamic behavior of the virtual assistant modulating guidance frequency and verbosity based on inferred cognitive state was particularly effective in reducing confusion during high-difficulty stages.

Engagement and immersion metrics likewise demonstrated significant positive effects. Using a standardized immersion scale, participants reported a 12% increase in engagement, while NASA-TLX scores indicated lower perceived workload across mental demand, effort, and frustration dimensions. User satisfaction surveys (SUS) recorded an average score of 87.1, placing the system in the "excellent usability" category.

Together, these results confirm that the proposed multisensor adaptive XR framework not only enhances cognitive state prediction but also translates these predictions into meaningful, real-time interaction improvements that directly benefit user experience and performance.

Implications for Future XR Design

The findings from this research offer several significant implications for the future design of extended reality (XR) systems, particularly as XR becomes more deeply integrated into education, healthcare, training, entertainment, and industrial applications. The results highlight the importance of transitioning from static, one-size-fits-all XR experiences toward multisensor-driven, cognitively adaptive environments capable of understanding and responding to user states in real time.

First, the demonstrated effectiveness of multisensor fusion suggests that future XR systems should increasingly adopt multimodal perception architectures that integrate physiological, behavioral, and environmental data[17]. Relying solely on head-mounted display (HMD) tracking or controller inputs will no longer be sufficient as users demand more natural, fluid, and intuitive interactions. Future XR devices may embed EEG/EMG electrodes, high-fidelity eye trackers, and micro-environment sensors directly into consumer-grade hardware, enabling richer and more reliable cognitive state estimation. Such integration will allow XR applications to shift from reactive to anticipatory interfaces, predicting user needs before cognitive overload or disengagement occurs.

Second, the results indicate a clear opportunity for designing XR systems that dynamically adapt their interface, content, and interaction mechanics based on real-time cognitive feedback. This suggests a paradigm shift from traditional UX design principles toward adaptive experience design, where UI elements adjust their complexity, size, and pacing depending on the user's attention, workload, fatigue, or stress level[17]. This adaptivity will be particularly transformative in high-stakes domains such as medical simulation, aviation training, and military operations where maintaining an optimal cognitive state can significantly influence task performance and safety outcomes.

Third, the research underscores the need for future XR platforms to implement robust, low-latency fusion engines capable of processing multiple data streams without interrupting immersion. Achieving this will require new hardware-software co-design strategies, optimized machine learning models, and sensor synchronization protocols tailored to XR environments. As XR interactions grow more complex and physically demanding, maintaining stability and accuracy in cognitive predictions even during rapid movements or scene transitions will be essential for preserving user trust and comfort.

Fourth, the results suggest that future XR design should increasingly incorporate ethical, privacy-aware, and transparent sensing frameworks. As physiological data such as EEG and heart-rate variability become integral to adaptive XR experiences, designers must prioritize user consent, data minimization, on-device processing, and explainability of adaptive behaviors. Users need to understand why the XR system is modifying the environment, what data it is using, and how these adaptations benefit them. Ethical design practices will be crucial to broad public acceptance and responsible deployment of cognitively adaptive XR technologies.

Finally, the outcomes of this research imply that future XR design will move toward holistic human-centered systems where the boundary between system and user becomes increasingly fluid. Instead of forcing users to conform to rigid interaction patterns, XR environments will gradually evolve to accommodate individual differences in cognition, motor skills, sensory preferences, and emotional states. This will open the door to XR applications that are genuinely personalized capable of delivering tailored learning paths, adaptive therapeutic interventions, and immersive experiences that adjust not just to user actions, but to the user's mental and emotional context.

Contribution to Human-Centered AI

This research makes several significant contributions to the evolving field of human-centered artificial intelligence, particularly within the context of immersive XR environments. At its core, the study advances the understanding of how AI systems can more effectively recognize, interpret, and adapt to human behaviors, cognitive states, and emotional cues in real time. By integrating multisensor data ranging from eye movements and hand gestures to physiological signals such as EEG and HRV the research pushes human-centered AI beyond traditional interaction paradigms that rely solely on explicit commands or limited tracking inputs.

A primary contribution lies in the development of a multilayer multisensor fusion framework that enables richer and more accurate interpretations of user states. Unlike existing XR systems that depend on one or two dominant modalities, this framework synthesizes low-level signals, mid-level features, and high-level semantic interpretations into a unified representation of the user. This holistic integration marks a substantive step forward for human-centered AI, as it demonstrates how AI systems can achieve a more nuanced and comprehensive understanding of human behaviour and cognition. Such an approach lays the foundation for future AI systems that are perceptive, context-aware, and capable of adapting to diverse user conditions.

Another important contribution is the introduction of cognitive state modeling within real-time XR interactions, bridging a gap between cognitive science and AI engineering. By employing deep neural architectures capable of estimating attention, workload, fatigue, and emotional states, the research demonstrates how human-centered AI can move from merely supporting interaction to actively shaping the user experience based on cognitive well-being. This work contributes new methods for real-time cognitive prediction, including fusion-based regression and classification models specifically tuned for dynamic, high-motion scenarios typical in XR. Such advancements show how human-centered AI can not only improve usability but also directly enhance user health, safety, and performance.

The research also contributes to the design of adaptive interaction mechanisms, a critical component of truly human-centered AI. Through dynamic adjustments of UI elements, lighting, haptic feedback, and content pacing, the system embodies the principle that technology should respond to the user rather than require the user to adapt to the technology. This work demonstrates a concrete implementation of AI-mediated adaptivity, offering a blueprint for future systems that provide personalized support, reduce cognitive overload, and maintain user engagement. In doing so, the research establishes a practical pathway for graduating from fixed XR experiences to intelligent, adaptive environments grounded in real-time human data.

Finally, this study contributes conceptually by reinforcing the idea that human-centered AI must be multimodal, adaptive, and ethically grounded. The framework emphasizes respect for user comfort, personalization, and transparency in AI-driven adjustments ensuring that adaptivity enhances, rather than disrupts, the user experience. By foregrounding ethical considerations such as informed consent, physiological data privacy, and interpretability of adaptive behaviors, the research strengthens the theoretical foundation of human-centered AI and signals the direction for its responsible future development.

Impact on Education, Training, Healthcare, and Industrial XR

The proposed human-centered AI framework has the potential to significantly transform how XR technologies are applied across key sectors such as education, professional training, healthcare, and industrial operations. By enabling real-time cognitive understanding and adaptive interaction, the

system advances XR applications beyond static simulations toward dynamic, personalized, and human-responsive environments that better support learning, performance, and safety.

In the field of education, the integration of multisensor fusion and cognitive modeling opens new possibilities for adaptive learning pathways tailored to individual students' cognitive states[18]. Traditional XR-based educational tools often treat all learners uniformly, assuming consistent levels of attention, comprehension, and mental workload. With real-time detection of cognitive fatigue, confusion, or high engagement, XR learning environments can automatically adjust instructional pacing, difficulty levels, or presentation formats. For example, visual complexity can be reduced when cognitive load is high, or supplementary explanations can be provided when attention drops. This capability supports more inclusive learning by accommodating diverse cognitive profiles and helping learners maintain optimal engagement. Such adaptivity also has strong implications for remote and virtual classrooms, where teachers cannot physically observe student behavior but can benefit from AI-driven insights into student readiness and comprehension.

In professional training, particularly in high-risk domains such as aviation, emergency response, military operations, and surgical practice, the system's ability to monitor stress, situational awareness, and fatigue can substantially enhance training outcomes and safety[19]. Traditional XR training modules often evaluate performance solely based on task completion or error rates. By incorporating physiological and behavioral sensing, trainers can better understand the cognitive processes underlying trainee actions. This enables the system to adaptively introduce challenges, slow down complex scenarios, or provide real-time guidance when cognitive overload threatens performance. Moreover, the fusion-driven model improves the realism and responsiveness of training simulations, making them more aligned with real-world conditions where high cognitive demands and rapid decision-making are common.

In healthcare, human-centered AI integrated with XR offers powerful tools for rehabilitation, therapy, surgical assistance, and patient monitoring. Cognitive modeling can play a crucial role in ensuring that therapeutic XR experiences such as motor rehabilitation or PTSD treatment remain safe, personalized, and attuned to patient emotional well-being[20]. For patients recovering from neurological injuries, the system can detect frustration, mental fatigue, or progress levels and adjust therapeutic difficulty accordingly. Similarly, for clinicians using XR for surgical planning or remote collaboration, real-time cognitive state monitoring can help prevent cognitive overload, reducing errors during crucial decision points. The combination of XR and physiological sensing also opens new avenues for early detection of cognitive decline, anxiety disorders, or attention impairments, providing clinicians with richer behavioral data for diagnosis and intervention.

In industrial settings, the proposed framework contributes to safer and more efficient XR-assisted workflows by providing real-time monitoring of worker attention, mental strain, and situational awareness. Many industries such as manufacturing, logistics, energy, and construction are increasingly adopting XR for remote maintenance, complex assembly tasks, and process visualization. By integrating cognitive load detection and adaptive assistance, XR systems can proactively prevent human error, which is often caused by fatigue, information overload, or reduced attention. For instance, when workers show signs of cognitive overload during a critical assembly task, the system can simplify the interface, reduce visual clutter, or highlight essential components to maintain productivity and reduce risk. Moreover, adaptive haptic and visual cues can enhance training for hazardous or technically demanding procedures, ensuring that workers perform tasks with higher precision and lower stress.

Collectively, these impacts demonstrate the broad transformative potential of human-centered AI in XR applications across multiple sectors. By grounding interaction design in real-time human cognitive and emotional states, the proposed framework helps create XR environments that are safer, more personalized, more effective, and more deeply aligned with human needs. This multidisciplinary advancement supports the evolution of XR from a passive visualization tool into an intelligent partner that enhances human capabilities across education, healthcare, industry, and professional training.

Strengths and Limitations

One of the most prominent strengths of this research lies in its holistic multisensor integration, which combines visual, behavioral, auditory, and physiological data into a unified model. This multimodal approach enables the system to infer cognitive and emotional states with greater accuracy and robustness compared to single-sensor or low-dimensional interaction methods. By capturing a wide spectrum of human behavior from eye movements and hand gestures to EEG-based cognitive markers the framework supports a deeper and more nuanced understanding of user experience. This allows the XR environment to adapt intelligently to fluctuating user needs, significantly improving comfort, engagement, and task efficiency.

Another strength is the development of a layered fusion architecture, which incorporates low-, mid-, and high-level fusion strategies to enhance robustness in unpredictable XR scenarios[21]. This design ensures resilience to partial sensor failure, noise, and rapid motion all common issues in immersive environments. Additionally, the deployment of advanced deep learning models, such as multimodal transformers and temporal encoders, strengthens the system's ability to operate in real time, making the framework scalable for practical applications in education, healthcare, industry, and training.

The research also demonstrates strength through its adaptive interaction mechanisms, which allow the XR system to adjust UI complexity, feedback modalities, lighting conditions, and task pacing based on the user's cognitive state. This contributes to a more human-centered and personalized XR experience, and it aligns closely with emerging principles of ethical and user-responsive AI. By prioritizing cognitive well-being and real-time adaptivity, the framework sets a foundation for next-generation XR systems that support safer, more intuitive, and more effective interactions.

Despite these strengths, several limitations must be considered. A major limitation involves the high computational and hardware requirements associated with multisensor data capture and real-time fusion[9]. Physiological sensors such as EEG or EMG require careful calibration, produce noisy signals, and are not yet seamlessly integrated into most consumer-grade XR headsets. This restricts the scalability of the framework to settings where specialized equipment and controlled conditions are available. Additionally, real-time fusion of high-frequency multimodal data imposes significant computational demands that may exceed the capabilities of lightweight or mobile XR devices.

Another limitation concerns the complexity of cognitive state modeling, which remains an open challenge in human-centered AI[22]. Although the proposed models improve prediction accuracy, cognitive and emotional states are inherently subjective and influenced by numerous external variables. As a result, the system's predictions while more accurate than baseline approaches may still carry ambiguity or risk misclassification. A misinterpreted cognitive signal may lead the system to adapt in ways that are unhelpful or disruptive to the user, highlighting the need for continuous model refinement and user feedback loops.

Furthermore, the framework raises ethical and privacy considerations, particularly given the sensitive nature of physiological and cognitive data. Storing, processing, and interpreting such data must be managed with robust safeguards to protect user autonomy and confidentiality. The need for explicit consent, transparent adaptation behavior, and secure data handling practices may limit adoption in environments where ethical oversight or regulatory compliance is insufficiently established.

Lastly, the system's dependence on multisensor calibration and synchronization introduces practical limitations. Small misalignments in sensor timing or placement can degrade fusion performance, while environmental noise such as lighting variation or background audio can compromise the reliability of certain modalities. These operational constraints highlight the challenge of deploying such a system outside controlled laboratory or training environments.

4. Conclusion

This research presents a comprehensive framework for advancing human-centered AI within immersive XR environments through the integration of multisensor fusion, cognitive state modeling,

and adaptive interaction design. By combining data from eye tracking, body and hand movements, environmental inputs, audio cues, and physiological signals such as EEG and HRV, the proposed system demonstrates the feasibility and value of a richly multimodal approach to understanding user behavior and internal cognitive states in real time. The results show that multisensor fusion significantly improves prediction accuracy, robustness, and responsiveness when compared to single-sensor models, especially in dynamic and high-motion XR scenarios. The cognitive modeling component of the framework successfully estimates key cognitive indicators including attention, workload, fatigue, and emotional engagement providing the foundation for an adaptive XR system that can anticipate and respond to user needs. The impact analysis highlights the transformative potential of this framework across critical sectors such as education, professional training, healthcare, and industrial operations. By enabling personalized learning, safer training simulations, patient-centered therapeutic interventions, and cognitively aware industrial support systems, the proposed human-centered AI model offers a pathway toward smarter, more intuitive, and more ethically aligned XR applications. Despite its contributions, the research acknowledges limitations related to sensor reliability, computational demands, cognitive modeling ambiguity, and ethical concerns surrounding sensitive data. These challenges underscore the need for future work focused on hardware integration, algorithmic efficiency, privacy-preserving sensor architectures, and cross-domain generalization. Future developments in XR hardware, wearable physiological sensors, and lightweight multimodal AI models are likely to further strengthen the framework's applicability and scalability. In summary, this research establishes a foundational step toward developing immersive XR environments that are deeply human-centered systems that perceive users holistically, adapt intelligently, and align their behavior with human cognitive, emotional, and physical needs. By combining multisensor fusion with adaptive AI-driven interactions, the study paves the way for the next generation of XR systems that are not only more immersive and efficient, but also more empathetic, safe, and aligned with human well-being.

References

[1] G. Margetis, G. Papagiannakis, and C. Stephanidis, "Realistic natural interaction with virtual statues in x-reality environments," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 42, pp. 801–808, 2019.

[2] H. Nguyen and T. Bednarz, "User experience in collaborative extended reality: overview study," in *International Conference on Virtual Reality and Augmented Reality*, Springer, 2020, pp. 41–70.

[3] A. Çöltekin *et al.*, "Extended reality in spatial sciences: A review of research challenges and future directions," *ISPRS Int. J. Geo-Information*, vol. 9, no. 7, p. 439, 2020.

[4] F. Chen *et al.*, "Multimodal behavior and interaction as indicators of cognitive load," *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 4, pp. 1–36, 2013.

[5] S. Agarwal *et al.*, "Unleashing the power of disruptive and emerging technologies amid COVID-19: A detailed review," *arXiv Prepr. arXiv2005.11507*, 2020.

[6] N. Aliman, "Hybrid cognitive-affective strategies for AI safety," 2020.

[7] A. L. Gardony, R. W. Lindeman, and T. T. Brunyé, "Eye-tracking for human-centered mixed reality: promises and challenges," in *Optical architectures for displays and sensing in augmented, virtual, and mixed reality (AR, VR, MR)*, SPIE, 2020, pp. 230–247.

[8] N. Sawhney and C. Schmandt, "Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments," *ACM Trans. Comput. Interact.*, vol. 7, no. 3, pp. 353–383, 2000.

[9] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," *IEEE Sens. J.*, vol. 2, no. 2, pp. 107–119, 2002.

[10] G. Matthews, L. Reinerman-Jones, J. Abich IV, and A. Kustubayeva, "Metrics for individual differences in EEG response to cognitive workload: Optimizing performance prediction," *Pers. Individ. Dif.*, vol. 118, pp. 22–28, 2017.

[11] X. Fu, H. Wang, Z. Wang, Z. Shi, W. Yang, and P. Ma, "Research on micro-grid group intelligent decision mechanism under the mode of block-chain and multi-agent fusion," *Energies*, vol. 12, no. 21, p. 4196, 2019.

[12] M. Kristiansson, "Memory, aging and external memory aids: Two traditions of cognitive research and their implications for a successful development of memory augmentation." 2011.

[13] M. G. Mendonça and V. R. Basili, "Validation of an approach for improving existing measurement

frameworks," *IEEE Trans. Softw. Eng.*, vol. 26, no. 6, pp. 484–499, 2002.
[14] Z. Wang, Y. Wu, and Q. Niu, "Multi-sensor fusion in automated driving: A survey," *Ieee Access*, vol. 8, pp. 2847–2868, 2019.
[15] A. Korbach, R. Brünken, and B. Park, "Differentiating different types of cognitive load: A comparison of different measures," *Educ. Psychol. Rev.*, vol. 30, no. 2, pp. 503–529, 2018.
[16] M. L. Rahman, *Towards Improving Cybersecurity and Augmenting Human Training Performance Using Brain Imaging Techniques*. University of California, Riverside, 2020.
[17] E. P. Blasch and S. Plano, "JDL Level 5 fusion model: user refinement issues and applications in group tracking," in *Signal processing, sensor fusion, and target recognition XI*, SPIE, 2002, pp. 270–279.
[18] E. Volta, "Multisensory learning in adaptive interactive systems," 2020.
[19] L. S. Kao and E. J. Thomas, "Navigating towards improved surgical safety using aviation-based strategies," *J. Surg. Res.*, vol. 145, no. 2, pp. 327–335, 2008.
[20] J. Ellis and A. Zaretsky, "Assessment and management of posttraumatic stress disorder," *Contin. Lifelong Learn. Neurol.*, vol. 24, no. 3, pp. 873–892, 2018.
[21] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.
[22] A. K. Goel and M. E. Helms, "Theories, models, programs, and tools of design: views from artificial intelligence, cognitive science, and human-centered computing," in *An Anthology of Theories and Models of Design: Philosophy, Approaches and Empirical Explorations*, Springer, 2014, pp. 417–432.