# Cognitive-Adaptive Multimodal Interaction in Extended Reality: A Survey of Attention-Aware, Privacy-Preserving, and Intelligent Human–Computer Interaction Frameworks

*Bhuvana Kumar V [1, 2], Debajyoty Banik [3]*
[1] Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia;
[2] Department of CSE, N.B.K.R Institute of Science & Technology, Vidyanagar, India;
[3] Department of CSE, Anurag University, Hyderabad, India
[1, 2] pdf.Bhuvana@lincoln.edu.my, [3] debajyoty.banik@gmail.com

---

**Abstract:** Extended Reality (XR) has enabled immersive Human Computer Interaction (HCI) and consists of Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). The recent advancement of eye tracking, multimodal sensing, and AI has created a greater potential for interactive systems. Though, the available XR systems still rely on rigid interaction models and do not adapt to the user's dynamic cognitive states. Here, we provide a comprehensive review of cognitive adaptive multimodal interaction systems in XR, focusing on attention-aware, privacy-preserving, and intelligent HCI. The systems rely on and fusion multiple data streams and measurements, such as gaze, head pose, gestures, and other contextual data, to assess the users cognitive load and interaction. The survey focuses on the assessment of cognitive load and the design of adaptive interfaces and multimodal machine learning for personalization and context in XR. The review includes emerging privacy-preserving methods for XR as well as the security and privacy concerns for the persistent acquisition of behavioral and biometric data. Based on this literature, we identify the future directions for cognitive digital twins, child-locked, explainable adaptive systems, and privacy-by-design XR frameworks. The survey encourages the innovation of cognitive adaptive XR systems that are intelligent, scalable, secure and privacy-preserving.
**Keywords**: Extended Reality (XR), cognitive-adaptive interaction, multimodal interaction, eye tracking, attention-aware systems, cognitive load assessment, human–computer interaction, multimodal machine learning, privacy-preserving XR, human-centered AI.

---

**Introduction**

Extended Reality (XR), a suite of technologies such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), synthesizes technology, information systems, computing, and design, making it particularly useful in a variety of domains, including, but not limited to, education and training, healthcare, entertainment, and industry [1]. The rapid proliferation of XR technology has resulted in the increased demand for sophisticated yet intelligent interaction mechanisms.

Due to the increased personalization and capability XR provided for user interface design, coupled with the automation of the interface, the design of intelligent immersive environments incorporated gaze behavior, gaze patterns, and eye movement [2] and [3].

The design of XR systems [4] built on the principles of Human-Computer Interaction (HCI), relied on the quality of the interaction and combined gaze, gesture, speech, and other contextual information for the design of intelligent interactive environments and systems [6].

XR systems, can, in effect, collect and aggregate sensitive biometric and behavioral data.[7] The evolution of the metaverse and XR systems has also lead to increased risks and exposure, including, but not limited to, the harvesting of behavioral data and profiling users.[8]. Along with XR systems' usability and effectiveness, cognitive work- and performance- burdens should also be assessed and maintained due to the strain on a participant's cognitive resources.[9]. Furthermore, Cognitive Load Theory [10] should be considered and applied for maintaining the designs of XR systems and environments in a user-centered manner.

Cognitive-adaptive multimodal interaction frameworks within XR may include privacy-preserving techniques and intelligent HCI coupled with attention-aware interaction and cognitive workload assessment techniques. This survey aims to capture the state of research of these frameworks within XR and their potential future avenues for exploration.

**Fundamentals of Extended Reality and Human–Computer Interaction**

XR (Extended Reality) combines VR (Virtual Reality), AR (Augmented Reality), and MR (Mixed Reality) [1]. XR systems integrate physical and digital systems and offer new ways of combining the two for the user [1].

- VR allows the user to experience a wholly digital environment.
- AR allows the user to experience the digital world overlayed on the physical world.
- MR allows the user to experience both the virtual and the physical worlds.

XR HCI (Extended Reality Human-Computer Interaction) has for the most part concentrated on the creation of systems incorporating gaze tracking, gesture recognition, and the interpretation of verbal and body movements [4]. Such systems rely on multimodal interaction, which is the integration of many forms of input which is straightforward for the user to interpret [5]. Due to the ongoing evolution of XR systems, multimodal machine learning has been introduced, creating systems which interact in a much more flexible and intelligent manner [6].

The majority of XR systems require users to adopt a static form of interaction. Because of systems' reliance on such forms of interaction, the systems have been characterized as lacking flexibility and adaptability. Many systems also have not incorporated features that account for user attention, which has created a subpar experience for users and a system interaction that has been largely ineffective [2]. This has been reported to create an information overload which hinders cognitive function and system performance [9]. Since XR environments capture biometric and behavioral data in real time, there is increased concern for the user's privacy and data security [7].

**Multimodal Interaction Frameworks in XR**

The integration of various modes of communication in XR systems can provide users the ability to interact in a way that feels more natural. Eye tracking is a method of gaze estimation that is useful for analyzing the movement of the eyes and monitoring visual attention. This provides systems better insight to users and their unique interactions [2]. Gaze interaction is only a single part of the XR system. XR systems also utilize the movement of the user's head, hand gestures, and voice to improve the system and provide users with a better experience [5].

The latest technology in the field of multimodal machine learning allows for the integration of different forms of data using feature-level fusion, decision-level fusion, and deep learning. These tools advance the modeling of users and the improvement of awareness, context, and interaction. These advancements in technology help the XR system become far more advanced and capable of real-time responsiveness to the user's needs. [6].

*Table 1. Comparison of Interaction Modalities in XR*

| **Modality** | **Advantages** | **Limitations** |
|---|---|---|
| Eye Tracking | Attention monitoring | Calibration |
| Hand Gestures | Natural interaction | Occlusion |
| Voice | Hands-free control | Noise |
| EEG | Cognitive monitoring | Expensive |

**Cognitive State Modeling in XR**

Cognitive State Modeling is essential for Extended Reality (XR) systems. It allows for understanding of the mental and behavioral states of users. One of the components of such systems is attention modeling, often based on eye-tracking. Attention modeling systems analyze the direction of gaze and where the user visually concentrates. This type of modeling can be used to realize attention adaptive interfaces. These interfaces allow dynamic manipulation of content and possible interaction strategies based on the user's attention and engagement levels [2]. Analyzing eye movements can also constitute an understanding of an individual's behavioral patterns and can lead to the design of intelligent adaptive XR interfaces [3].

Cognitive load evaluation is also important and characterizes the effort an individual is spending to complete a task in an immersive environment. An excessive mental workload can deteriorate the experience for the user and the quality of decisions and tasks that are completed. Consequently, it is necessary to evaluate cognitive load for the design of immersive XR systems [9]. Cognitive Load Theory deals with the effort an individual is spending, the complexity of the information being provided, and the level of fatigue that the individual is experiencing. This theory is most concerned with reducing cognitive fatigue and information overload [10].

The NASA Task Load Index (NASA-TLX) is considered to be an exemplary task load evaluation system. It subdivides cognitive load into six categories: mental, physical, and temporal demand, as well as perceived performance, effort and frustration. NASA-TLX evaluates the perceived task load and is broadly applied to evaluate XR systems [9].

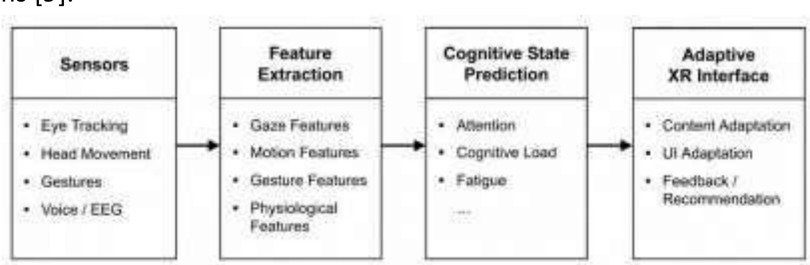


*Figure 1. Cognitive State Inference Pipeline*

**Cognitive-Adaptive Interaction Frameworks**

Cognitive-adaptive interaction frameworks for XR systems recognize user needs, preferences, and the context, and modify behavior accordingly. Adaptive User Interfaces (AUIs) modify user interfaces (UIs) to adapt layouts and content, and modify the level of interaction to manage cognitive load and optimize the usability of the systems. More advanced UIs further modify systems and suggestions based on user behavior and preferences [10].

*Table 2. Adaptation Strategies in Cognitive-Adaptive XR Systems*

| Adaptation Strategy | Objective |
|---|---|
| UI Adaptation | Reduce workload |
| Information Filtering | Reduce overload |
| Dynamic Feedback | Improve usability |
| Personalization | Enhance engagement |

**Privacy-Preserving XR Systems**

Privacy issues arise from combining immersive technology and extensive data harvesting from users. User profiling occurs when eye-tracking technology identifies user interests and thought patterns. This information is vulnerable to exploitation and abuse. Also, analyzing interaction and movement patterns to determine behavior is vulnerable to abuse when profiling user activities

*Table 3. Privacy Threats and Mitigation Strategies*

| Privacy Threat | Impact | Mitigation |
|---|---|---|
| Eye Tracking Leakage | Behavioral profiling | Encryption |
| Motion Data Leakage | Identity exposure | Anonymization |
| Cognitive Data Exposure | Privacy violation | Federated Learning |

**Research Challenges and Open Issues**

*Table 4. Research Challenges and Open Issues in Cognitive-Adaptive XR Systems*

| Research Challenge | Description | Impact on XR Systems |
|---|---|---|
| Real-Time Cognitive Inference | Prediction of user attentiveness and cognitive load with minimal latency. | Delayed adaptation and decreased responsiveness |
| Multimodal Data Fusion | Combining multimodal data, including gaze, gestures, and voice | Reduced prediction accuracy |
| Computational Complexity | Increased demands on processing and memory for multimodal analysis | Higher latency consumed resources more |
| Privacy Preservation | Sensitive behavioral and biometric data privacy protection | Threats to privacy and security |
| User Trust | Guaranteeing openness, dependability, and equity | Reduced user acceptance and engagement |
| Explainability | Facilitating Interpretability of AI-Based Reasoning and Modifications | A lack of transparency and accountability |

**Future Research Directions**

The fusion of cognition, XR, and adaptive capability systems will produce emerging methods of private personal data embedding. Digital Twins will represent user behavior and cognition. Brain-Computer Integrated XR will allow users to connect and interact with one another via neural links. Explainable Adaptive XR will use transparency of AI to cultivate trust with its users. During model training, Federated Cognitive Learning will provide measures to preserve privacy. Combined with Privacy-by-Design XR Systems, it will adopt a model of security within the design of the interactions. Human-Centered AI integrates a model of equity and user wellbeing.

## Conclusions

Experts envision an extension of virtual reality technologies as an upgrade to traditional user-computer interaction. Suggested XR systems allow for more dynamic user engagement using not just traditional interfaces, but an upgrade to interfaces that incorporate elements such as eye tracking, gesture and voice recognition, and situational context. Technologies that integrate eye tracking and multi-modal machine learning can assist in recognizing users' cognitive states through attention and behavior tracking and can measure cognitive load. Design of systems that use XR intelligently and adaptively will aid in the presentation of information in an accommodating manner to minimize cognitive burden. Unfortunately, privacy and security remain principal concerns with constant harvesting of sensitive biometric and behavioral data. Therefore, XR systems will likely incorporate attention awareness and intelligent systems focused on the preservation of user privacy and security. The IMMERSE project will implement secure, trustworthy, and focused systems and people-centered, designed and implemented immersive environments.

## References


1. Billinghurst M, Clark A, Lee G (2015), "A Survey of Augmented Reality". Foundations and Trends in Human-Computer Interaction, Vol. 8 No. 2-3 pp. 73–272, doi: https://doi.org/10.1561/1100000049.
2. Clay, V., König, P., & König, S. (2019). Eye Tracking in Virtual Reality. Journal of Eye Movement Research, 12(1), 1-18. https://doi.org/10.16910/jemr.12.1.3
3. Duchowski, A.T. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 455–470 (2002). https://doi.org/10.3758/BF03195475
4. Jacko, J.A. (Ed.). (2012). Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition (3rd ed.). CRC Press. https://doi.org/10.1201/b11963
5. Alejandro Jaimes, Nicu Sebe, "Multimodal human–computer interaction: A survey", Computer Vision and Image Understanding, Volume 108, Issues 1–2, 2007, Pages 116-134, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2006.10.019.
6. T. Baltrušaitis, C. Ahuja and L. -P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423-443, 1 Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.
7. Mahdi Alkaeed, Adnan Qayyum, Junaid Qadir, "Privacy preservation in Artificial Intelligence and Extended Reality (AI-XR) metaverses: A survey", Journal of Network and Computer Applications, Volume 231, 2024, 103989, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2024.103989.
8. Y. Huang, Y. J. Li and Z. Cai, "Security and Privacy in Metaverse: A Comprehensive Survey," in *Big Data Mining and Analytics*, vol. 6, no. 2, pp. 234-247, June 2023, doi: 10.26599/BDMA.2022.9020047.
9. Sandra G. Hart, Lowell E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research", Editor(s): Peter A. Hancock, Najmedin Meshkati, Advances in Psychology, North-Holland, Volume 52, 1988, Pages 139-183, ISSN 0166-4115, ISBN 9780444703880, https://doi.org/10.1016/S0166-4115(08)62386-9.
10. Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist, 38(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8