Designing Intent-Aware AI Systems for Personalized Applications Using Multimodal

Kholod Alsufiani

Computer Sciences Program, Turabah University College, Taif University, Taif, Saudi Arabia

Abstract
Human-Centered Artificial Intelligence (HCAI) emphasizes aligning intelligent system behavior with human goals, cognitive states, and contextual needs. Although prior research has explored adaptive and affect-aware systems, most existing approaches remain reactive and rely on isolated interaction signals. This paper proposes a framework for intent-aware personalization in human-centered AI, grounded in multimodal cognitive interaction signals such as gaze, affect, physiological responses, and paralinguistic audio cues. The framework theorizes how multimodal cognitive signal integration enables accurate intent inference, which in turn drives adaptive personalization mechanisms that enhance engagement, reduce cognitive load, and improve trust. A set of research propositions is presented to guide future empirical validation. The proposed framework provides a theoretical foundation for designing

Keywords:
Human-Centered AI; Intent Inference; Multimodal Interaction; Adaptive Personalization; Conceptual Framework

1. Introduction

Artificial Intelligence (AI) systems have evolved from static automation tools into interactive entities embedded in learning, decision support, and collaborative environments. As AI systems increasingly interact with humans, aligning system behavior with human goals and cognitive states has become a critical challenge. This shift has led to the emergence of **Human-Centered Artificial Intelligence (HCAI)**, which emphasizes human needs, transparency, and adaptability in intelligent system design [1].

Despite advances in adaptive interfaces and affect-aware computing, many AI systems still rely on **reactive personalization strategies**, adjusting system behavior only after explicit feedback or performance degradation occurs [2]. Such approaches are insufficient in cognitively demanding scenarios, where user intent may shift rapidly due to changes in attention, emotion, or workload. Consequently, understanding and anticipating **user intent** has emerged as a central research challenge in HCAI [3].

Previous studies have explored intent inference using unimodal interaction signals such as gaze, speech, or physiological data [4], [5]. However, unimodal approaches often provide incomplete or ambiguous representations of user state. Recent advances in multimodal interaction suggest that integrating heterogeneous cognitive signals can yield richer and more robust representations of human behavior [6]. Nevertheless, the theoretical relationships between multimodal eyesignals, intent inference, and adaptive personalization remain underdeveloped. To address this gap, this paper proposes a **framework for intent-aware personalization in human-centered AI**. Rather than presenting a system implementation or experimental evaluation, this work theorizes how multimodal cognitive interaction signals can be integrated to infer user intent and drive adaptive personalization mechanisms that enhance engagement, reduce cognitive load, and improve trust.

2. Literature Review

a) Human-Centered Artificial Intelligence

HCAI focuses on designing intelligent systems that respect human cognitive processes, values, and goals. Research has shown that human-centered design improves usability, trust, and long-term engagement in intelligent interfaces [7]. In interactive and immersive environments, aligning system behavior with human expectations has been shown to enhance collaboration effectiveness and situational awareness [8].

Manuscript received January 5, 2026
Manuscript revised January 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.1.25

However, many HCAI approaches emphasize explainability or interface usability while offering limited guidance on how systems should dynamically adapt to evolving user intent. As a result, personalization mechanisms often remain static or rule-based, limiting their effectiveness in complex interaction contexts [9].

b) Multimodal Cognitive Interaction Signals

Multimodal cognitive interaction signals—including gaze, facial expressions, speech dynamics, and physiological responses—are widely used to infer user attention, emotion, and cognitive workload [10]. Gaze tracking provides insight into visual attention and task focus, while physiological signals such as heart rate variability and galvanic skin response correlate with mental effort and stress [11].

Although each modality contributes valuable information, unimodal analysis suffers from ambiguity and noise. To address these limitations, multimodal fusion techniques have been proposed to integrate heterogeneous signals into unified representations [12]. Such approaches have demonstrated improved robustness and predictive performance across diverse interaction environments, including virtual and augmented reality systems [13].

c) Intent Inference and Adaptive Personalization

Intent inference refers to a system's ability to recognize what users aim to accomplish during interaction. Prior work has investigated intent modeling using multimodal behavioral and task-related signals in adaptive and intelligent systems [14]. Adaptive personalization systems leverage inferred user state to adjust content delivery, feedback, and interaction strategies, enhancing engagement when personalization is contextually appropriate [15].

However, excessive or poorly calibrated adaptation may increase cognitive load or disrupt user flow, particularly in multimodal environments [16]. Importantly, existing models rarely conceptualize intent inference as a central mediating construct linking multimodal signals and adaptive personalization.

3. Framework Development

The proposed framework conceptualizes human–AI interaction as a dynamic process driven by multimodal cognitive interaction signals. Human cognitive signals form the input layer, which is transformed through multimodal cognitive signal integration into a unified representation. This representation enables intent inference, conceptualized as the central mediator between human state and system behavior. Inferred intent subsequently drives adaptive personalization mechanisms, resulting in improved human-centered outcomes such as engagement, reduced cognitive load, and trust.

3.1 Design Principles for Intent-Aware Human-Centered AI

To further clarify the relationships between the core constructs of the proposed framework, Figure 1 presents a conceptual instantiation of intent-aware human-centered AI. Rather than depicting a concrete system implementation, the figure provides an illustrative architectural view that translates the theoretical components of the framework into a coherent interaction pipeline. This visualization aims to support conceptual understanding of how multimodal cognitive interaction signals may be processed, integrated, and leveraged to drive intent-aware adaptive personalization in future systems.

As illustrated in Figure 1, the framework begins with a sensor input layer that captures heterogeneous cognitive interaction signals, including visual behavior, audio cues, physiological responses, and gaze patterns. These signals are temporally aligned and integrated through a multimodal cognitive signal fusion process, enabling the construction of a unified representation of the user's cognitive state.

This integrated representation supports intent inference, which is conceptualized as the central mediating mechanism between human cognitive signals and adaptive system behavior. Inferred user intent subsequently informs the adaptive personalization core, enabling real-time adjustment of system responses, interaction strategies, and content presentation.

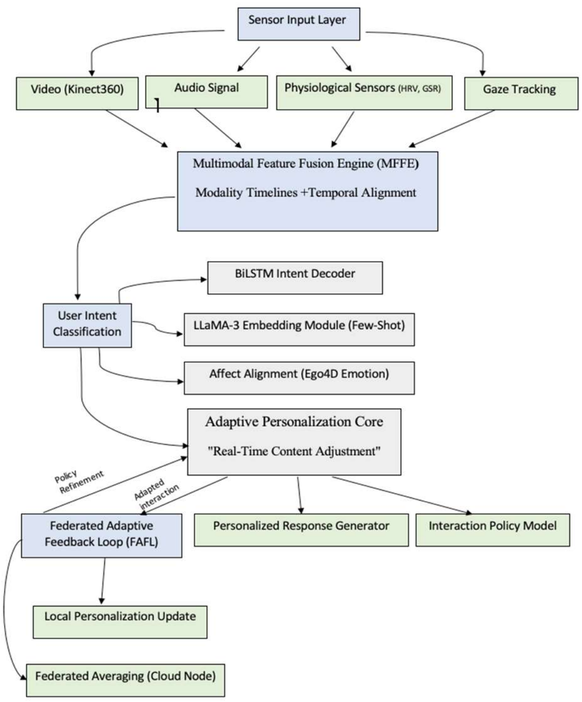

Figure 1. Conceptual Architecture of an Intent-Aware Human-Centered AI System

The framework further incorporates an adaptive feedback loop that allows personalization strategies to evolve over time, while conceptually supporting privacy-preserving learning through decentralized or federated update mechanisms. Importantly, the figure serves as a conceptual illustration of the theoretical relationships proposed in this work, rather than a prescriptive or fully specified system architecture. While specific computational components are shown for illustrative purposes, the framework remains algorithm-agnostic and is intended to guide future empirical implementations rather than constrain them.

To translate the proposed conceptual framework into actionable guidance for system design, this work derives a set of design principles grounded in prior literature on multimodal interaction, intent inference, and adaptive personalization. These principles articulate how the core components of the framework should be operationalized to support reliable intent-aware human-centered AI systems.

DP1. Multimodal cognitive interaction signals should be jointly leveraged to support reliable intent inference.

Prior studies have shown that unimodal signals such as gaze, speech, or physiological responses provide incomplete and sometimes ambiguous indicators of user state [4], [5], [10]. Integrating multiple cognitive modalities enables a richer and more robust representation of human behavior, which supports more accurate inference of user intent [6], [12].

DP2. Multimodal cognitive signal integration should function as a mediating layer between raw cognitive signals and intent inference.

Multimodal integration techniques reduce noise and ambiguity by aligning heterogeneous signals into a shared representational space [6], [12]. This integration enables higher-level reasoning about user state that cannot be achieved through isolated signal processing, thereby mediating the relationship between raw cognitive signals and accurate intent inference [1].

DP3. Adaptive personalization strategies should be driven by inferred user intent rather than surface-level interaction patterns.

Research on adaptive interfaces and intelligent systems indicates that personalization is more effective when grounded in an understanding of user goals instead of observable behaviors alone [2], [14]. Accurate intent inference allows systems to anticipate user needs and tailor responses accordingly, improving alignment between system behavior and user objectives [15].

DP4. Intent-aware adaptive personalization should aim to minimize users' cognitive load during interaction.

Studies on cognitive load and adaptive interaction suggest that poorly timed or irrelevant system adaptations can increase mental effort [11], [16]. In contrast, intent-aware personalization enables systems to regulate interaction complexity and feedback based on inferred user intent, thereby reducing unnecessary cognitive burden [4].

DP5. Intent-aware personalization should enhance user engagement and trust by maintaining relevance and predictability of system behavior.

Prior work has demonstrated that adaptive systems increase engagement and trust when system responses are perceived as relevant and predictable [7], [8]. Personalization strategies informed by user intent strengthen this perception by aligning system behavior with user expectations and interaction goals [15].

DP6. Intent-aware personalization mechanisms should be particularly emphasized in high cognitive workload contexts.

In cognitively demanding environments such as learning and immersive interaction, users are especially sensitive to misalignment between system behavior and intent [11], [13]. Empirical evidence indicates that adaptive systems yield greater benefits under high workload conditions, amplifying the effectiveness of intent-aware personalization [4], [16].

3.2 Operationalization and Measurement Guidance

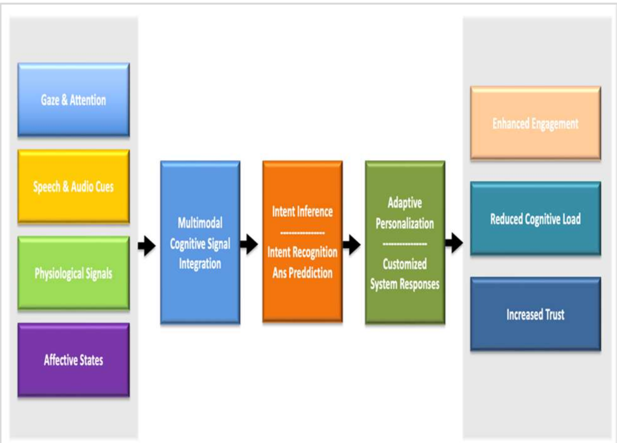

Figure 2. Conceptual Framework for Intent-Aware Human-Centered AI

To support future empirical validation of the proposed conceptual framework (Figure 2), this section outlines how each construct can be operationalized and measured in practice. Human cognitive interaction signals can be quantified using standard indicators such as gaze fixation patterns, paralinguistic speech features, physiological workload measures, and affective state estimations [4], [5], [10], [11]. These signals provide complementary perspectives on user attention, emotion, and cognitive effort.

Multimodal cognitive signal integration can be operationalized through the quality of fusion achieved across modalities, measured via performance gains over unimodal baselines, robustness under modality dropout, and real-time processing latency [6], [12]. Intent inference can be evaluated using predictive accuracy, temporal stability of inferred intent, and time-to-correct-intent metrics [14].

Adaptive personalization may be measured by observable system adaptation behaviors, including adaptation frequency, latency, and perceived appropriateness of responses relative to inferred intent [2], [3], [7]. Finally, human-centered outcomes can be assessed using engagement metrics, standardized cognitive load instruments (e.g., NASA-TLX), trust scales, and task effectiveness measures [7], [8], [11], [16].

Together, these operationalization guidelines provide a practical pathway for translating the proposed conceptual framework into empirical studies, while preserving its theoretical orientation.

4. Conclusion

This paper presented a framework for intent-aware personalization in human-centered AI, grounded in multimodal cognitive interaction signals. By positioning intent inference as the central mediator between human cognitive states and adaptive system behavior, the framework advances existing approaches beyond reactive personalization. The proposed research propositions and operationalization guidance provide a foundation for future empirical validation and system design. Overall, this work contributes a theoretically grounded perspective to the development of anticipatory, human-aligned AI systems

References

[1] Z. Stefanidi, G. Margetis, S. Ntoa, and G. Papagiannakis, "Real-time adaptation of context-aware intelligent user interfaces for enhanced situational awareness,"*IEEE Access*, vol. 10, pp. 23367–23393, 2022,doi: 10.1109/ACCESS.2022.3152973.

[2] R. Gong and M. Hua,"Designing multimodal user interfaces for hybrid collaboration: A user-centered approach,"in *Proceedings of the International Conference on Human-Computer Interaction*,pp. 67–82, 2023,doi: 10.1007/978-3-031-35647-9_5.

[3] S. Mai, Y. Zeng, and H. Hu,"Multimodal information bottleneck: Learning minimal sufficient unimodal and multimodal representations,"*IEEE Transactions on Multimedia*, vol. 25, pp. 4121–4134, 2022,doi: 10.1109/TMM.2022.3167954.

[4] M. Yang, Y. Gao, L. Tang, J. Hou, and B. Hu,"Wearable eye-tracking system for synchronized multimodal data acquisition,"*IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, pp. 5146–5159, 2023,doi: 10.1109/TCSVT.2023.3246462.

[5] J.Li,"Recent advances in end-to-end automatic speech recognition," *APSIPA Transactions on Signal and Information Processing*, vol. 11, e8, 2022,doi: 10.1017/ATSIP.2022.8.

[6] D. Wang, T. Zhao, W. Yu, N. V. Chawla, and M. Jiang,"Deep multimodal complementarity learning,"*IEEE Transactions on Neural Networks and Learning Systems*, vol. 34,pp. 10213–10224, 2022,doi:10.1109/TNNLS.2022.3140982.

[7] T. Darin, R. Andrade, and J. Sánchez,"Usability evaluation of multimodal interactive virtual environments for learners who are blind: An empirical investigation,"*International Journal of Human-Computer Studies*, vol. 158, p. 102732, 2022, doi: 10.1016/j.ijhcs.2022.102732.

[8] J. K. Gibbs, M. Gillies, and X. Pan,"A comparison of the effects of haptic and visual feedback on presence in virtual reality,"*International Journal of Human-Computer Studies*, vol. 157, p. 102717, 2022,doi: 10.1016/j.ijhcs.2021.102717.

[9] G. Schiavo, O. Mich, M. Ferron, and N. Mana,"Trade-offs in the design of multimodal interaction for older adults,"*Behaviour & Information Technology*, vol. 41, no. 5, pp. 1035–1051, 2022, doi: 10.1080/0144929X.2021.1876139.

[10] S. Paplu, H. Ahmed, A. Ashok, S. Akkus, and K. Berns, "Multimodal perceptual cues for context-aware human–robot interaction," in *Proceedings of the IFToMM International Symposium on Science of Mechanisms and Machines (SYROM)*,pp. 283–294, 2022,doi: 10.1007/978-3-031-19485-9_28.

[11] M. Lagomarsino, M. Lorenzini, E. De Momi, and A. Ajoudani, "An online framework for cognitive load assessment in industrial tasks," *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102380, 2022, doi: 10.1016/j.rcim.2022.102380.

[12] L. Cao, H. Zhang, C. Peng, and J. T. Hansberger,"Real-time multimodal interaction in virtual reality: A case study with a large virtual interface,"*Multimedia Tools and Applications*, vol. 82, pp. 25427–25448, 2023,doi: 10.1007/s11042-023-15710-4.

[13] Y. Luo, F. Liu, Y. She, and B. Yang,"A context-aware mobile augmented reality pet interaction model to enhance user experience,"*Computer Animation and Virtual Worlds*, vol. 34, e2123, 2023,doi: 10.1002/cav.2123.

[14] L. Wachowiak, P. Tisnikar, G. Canal, A. Coles, M. Leonetti, and O. Celiktutan,"Predicting when and what to explain from multimodal eye tracking and task signals,"*IEEE Transactions on Affective Computing*, pp. 1–12, 2024,doi: 10.1109/TAFFC.2024.3361671.

[15] K. Liu, F. Xue, D. Guo, L. Wu, S. Li, and R. Hong, "MEGCF: Multimodal entity graph collaborative filtering for personalized recommendation,"*ACM Transactions on Information Systems*, vol. 41, no. 4, pp. 1–27, 2023,doi: 10.1145/3592392.

[16] Y. Huang, K. Yao, J. Li, D. Li, H. Jia, Y. Liu, C. K. Yiu, W. Park, and X. Yu,"Recent advances in multi-mode haptic feedback technologies towards wearable interfaces,"*Materials Today Physics*, vol. 22, p. 100602, 2022, doi: 10.1016/j.mtphys.2021.100602.

Kholod Alsufiani holds a BSc degree with honors in Computer Science from Taif University, followed by an MSc degree in Computer Science from Middlesex University in London, which she obtained in 2013. In 2020, she successfully completed her PhD in Computer Science from Middlesex University. Currently serving as an Assistant Professor at Turaba University College-Taif university . Kholod's research interests are centered around the fields of information visualization, human-centered artificial intelligence, human-centered design and artificial intelligence.