

Adaptive Virtual Reality Museum: A Closed-Loop Framework for Engagement-Aware Cultural Heritage

Joseph Damouni¹, Wadia Tanus¹ and Naomi Unkelos-Shpigel^{1,*}

¹Braude College of Engineering, Karmiel, Israel

Abstract

Static information presentation in VR cultural heritage often causes cognitive overload or under-stimulation. We introduce a closed-loop adaptive interface that tailors content depth to real-time visitor behavior through implicit multimodal sensing. Our approach continuously monitors gaze dwell, head kinematics, and locomotion to infer engagement via a transparent rule-based classifier, which drives a Large Language Model to dynamically modulate explanation complexity without interrupting exploration.

We implemented a proof-of-concept in the Berat Ethnographic Museum and conducted a preliminary evaluation ($N = 16$) comparing adaptive versus static content. Results indicate that adaptive participants demonstrated 2–3× increases in reading engagement and exploration time while maintaining high usability ($SUS = 84.3$). Technical validation confirmed sub-millisecond engagement inference latency on consumer VR hardware.

These preliminary findings warrant larger-scale investigation and raise questions about engagement validation, AI transparency, and generative models in heritage contexts. We present this work-in-progress to spark discussion about implicit AI-driven adaptation in immersive cultural experiences.

Keywords

Virtual Reality, Cultural Heritage, Adaptive Interfaces, User Engagement

1. Introduction

In recent years, virtual reality (VR) has emerged as a powerful medium for cultural heritage presentation, enabling immersive reconstructions of historical spaces that are otherwise geographically inaccessible, physically fragile, or no longer extant [1]. By supporting spatial exploration and embodied interaction, VR museums promise richer engagement than traditional screen-based digital exhibits [2].

Despite these advantages, most existing VR museum systems adopt a static content delivery paradigm inherited from physical exhibitions. Exhibit descriptions are typically authored in advance and presented identically to all visitors, regardless of differences in attention span or moment-to-moment engagement [3]. Consequently, users frequently experience cognitive misalignment: cognitive overload when complex explanations are presented during rapid movement or under-stimulation when brief descriptions fail to satisfy focused attention [4].

While modern standalone VR headsets provide continuous streams of behavioral data (e.g., head motion, gaze direction, locomotion patterns, and hand movement) that offer the potential to infer attention without interrupting the user experience, in most VR museum applications these signals remain underutilized [5]. Recent advances in large language models have further expanded the design space by enabling generation of textual explanations with controllable length and complexity, though their integration into latency-sensitive interactive systems remains underexplored.

In this paper, we present a multimodal closed-loop adaptive framework for VR museum experiences that personalizes exhibit explanations in real-time based on inferred visitor engagement. Our approach monitors implicit behavioral signals available on a consumer VR headset, including gaze dwell, head kinematics, locomotion velocity, and hand proximity. These signals are processed by a transparent, rule-based engagement classifier that estimates the user's current engagement state without requiring calibration or explicit input. The inferred engagement level drives an adaptive content mechanism

Joint Proceedings of the ACM Intelligent User Interfaces (IUI) Workshops 2026, March 23-26, 2026, Paphos, Cyprus

*Corresponding author.

✉ joseph.damouni@gmail.com (J. Damouni); wtanoos905@gmail.com (W. Tanus); naomius@braude.ac.il (N. Unkelos-Shpigel)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

that leverages a large language model to modulate the depth of textual explanations[6]. This work investigates the question: *Does engagement-driven content adaptation in a VR museum improve the user experience compared to a static VR museum?*

We focus on two specific sub-questions:

- **RQ1:** To what extent can implicit behavioral signals support real-time attention-aware adaptation?
- **RQ2:** Compared to a static control, how does adaptation influence exploration behavior, content consumption, and the motivation to engage with exhibits?

We implemented the system within a high-fidelity virtual reconstruction of the Berat Ethnographic Museum[7] and conducted a between-subjects user study ($N = 16$) to evaluate these questions.

In summary, this paper makes the following contributions:

1. A novel multimodal closed-loop framework that fuses implicit behavioral signals to infer user engagement in real-time.
2. A proof-of-concept implementation showing a content adaptation pipeline that integrates Large Language Models (LLMs) into VR to dynamically adjust text complexity based on user attention.
3. Initial empirical insights into the impact of implicit, pacing-aware personalization on behavioral engagement and content consumption in virtual museums.

2. Background and Related Work

2.1. Adaptive Interfaces in VR

Adaptive interfaces are systems that modify their presentation, navigation, or interaction logic in response to user signals. In the context of VR this often involves tailoring the environment to reduce cybersickness or manage cognitive load. Prior research in Human-Computer Interaction (HCI) has investigated adaptation in training, rehabilitation, and gaming [8]. Within cultural heritage applications, personalization has primarily focused on pre-visit profiling or manually selected content paths. However, such approaches assume stable user intent[9]. Design thinking driven AR heritage systems have also been used to enhance visitor engagement in cultural and memorial contexts [10]. Recent research reports that static content presentation in immersive media can induce elevated extraneous cognitive load, suggesting that explicit selection mechanisms alone are insufficient to address dynamic attentional demands [11]. While recent work has explored multimodal intent modeling [12] and LLM integration in XR [13], our work differs by specifically targeting the pacing of information delivery using a closed-loop system, rather than general intent detection. Furthermore, unlike prior personalization approaches that rely on pre-visit profiling or explicit narrative selection [14, 15], our framework adapts content depth solely through implicit, real-time behavioral cues, reducing reliance on manual calibration or menu-based interactions.

2.2. Implicit Behavioral Signals

User engagement is commonly described as a multidimensional construct encompassing attention, interest, and involvement. In the absence of explicit feedback (like buttons or ratings), systems must rely on implicit signals to estimate this state. In addition, head kinematics provide a reliable proxy for attentional stability in VR [16], showing that low head angular velocity is strongly associated with sustained visual attention. Regarding gaze-based interaction, studies indicate that dwell durations on the order of one second are sufficient to distinguish intentional focus from transient glances in gaze-driven interaction pipelines [17]. Similarly, locomotion behavior has been shown to influence cognitive availability. Specifically, increased movement is associated with reduced attentional resources during VR walking tasks[18].

2.3. Generative Content Adaptation

Generative content adaptation refers to the use of AI to dynamically create or modify media at runtime. Recent advances in Large Language Models (LLMs) have enabled the generation of fluent, context-aware textual explanations. However, integrating LLMs into immersive VR environments introduces challenges related to latency and predictability. Research examining text consumption in VR indicates that effective reading requires perceptual stability to maintain readability. Due to these latency and reliability constraints, many prior applications of LLMs in immersive cultural heritage settings have prioritized offline generation or pre-scripted retrieval over fully real-time adaptation[19].

3. Method

3.1. Engagement Modeling

Our engagement modeling architecture transforms raw sensor data into a high-level cognitive state estimation through a three-stage pipeline. In this work, we operationalize ‘engagement’ specifically as a proxy for attentional stability and reading readiness, rather than a broader emotional or cognitive construct. Figure 1 summarizes the architecture, which is explained in subsequent sections.

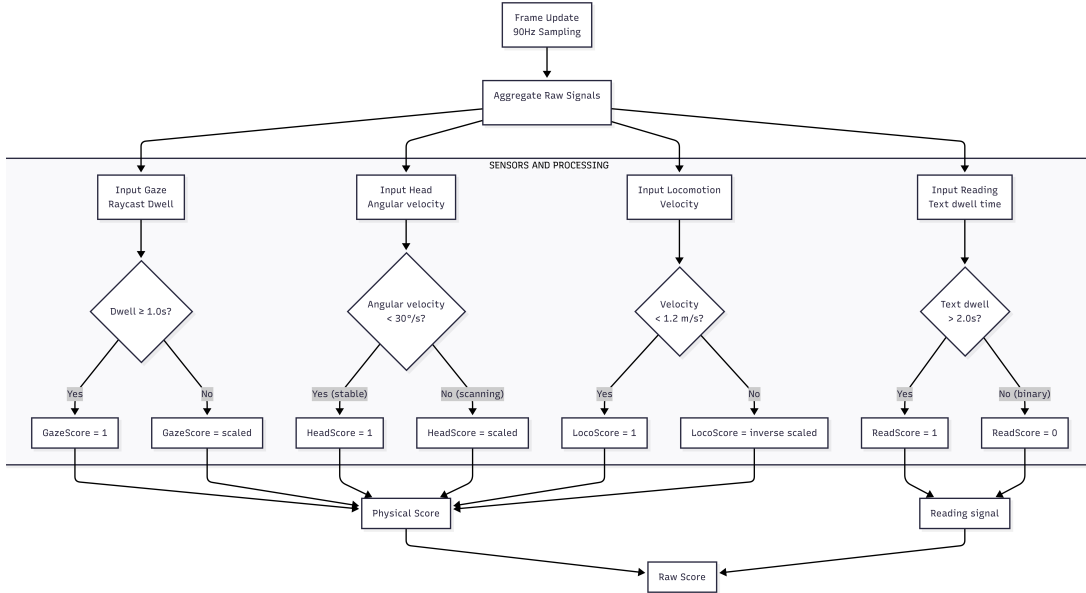


Figure 1: Signal Pre-processing and Sensor Fusion. Normalized behavioral signals are aggregated into a composite engagement score E_{raw} .

3.2. Signal Pre-processing

Raw sensor data from the VR headset is sampled at 90Hz. To ensure robust classification, all inputs are normalized to a unit scale [0, 1].

- **Head Stability (S_{head}):** Guided by research linking minimal head movement to sustained visual attention[16], we established a conservative threshold of $30^\circ/\text{s}$ to differentiate between environmental scanning and focused fixation.
- **Gaze Dwell (S_{gaze}):** Prior work on gaze-based interaction suggests that dwell durations in the range of one second effectively distinguish intentional focus from transient glances [17]. Following this guidance, we adopt a **1.0-second dwell threshold** for gaze raycast stability in our system.

- **Locomotion Velocity (S_{loco}):** Research on VR locomotion indicates that users instinctively slow down when their cognitive load or engagement increases[18]. Based on typical VR walking speeds, we establish **1.2 m/s** (approximating standard real-world walking speed) as a normalization baseline for comfortable locomotion; velocities exceeding this limit are inversely mapped to the engagement score.

3.3. Sensor Fusion and Weighting

Normalized signals are aggregated into a composite engagement score (E_{raw}) using a weighted fusion strategy ($W_{phys} = 0.75, W_{read} = 0.25$).

$$E_{raw} = (W_{phys} \times S_{phys}) + (W_{read} \times S_{read}) \quad (1)$$

- **Physical Score (S_{phys}):** Weighted average of Head Stability (35%), Gaze (30%), and Locomotion (35%).
- **Reading Context (S_{read}):** Binary signal triggered when the user dwells on text for $> 2.0s$ [11].

To model affective inertia [20], the fused score is temporally smoothed using linear interpolation ($\alpha = 0.35$).

3.4. Safety Gates

To reduce false positives during high-entropy states, the system implements conservative logic gates motivated by prior findings that locomotion increases cognitive load and reduces attentional availability in VR [18].

1. **Locomotion Gate ($v > 1.2 \text{ m/s}$):** During sustained locomotion beyond typical walking speed, users are unlikely to maintain stable visual attention to dense exhibit text. Consistent with previous research on locomotion as demanding cognitive resources and reducing processing capacity, we conservatively cap engagement inference at a *Neutral* state.
2. **Run Gate ($v > 2.0 \text{ m/s}$):** During high-speed locomotion exceeding typical walking velocities, users are unlikely to sustain focused content consumption. Drawing on evidence of velocity-dependent cognitive demands during VR locomotion, we conservatively treat running behavior as incompatible with focused reading and force the state to *Disengaged*.

Figure 2 illustrates the temporal stabilization and safety gating applied to the raw engagement estimate. The continuously fused engagement score (E_{raw}) is first temporally smoothed using linear interpolation to model affective inertia and to prevent rapid oscillations between engagement states. The resulting smoothed score is then evaluated against velocity-based safety gates. During moderate locomotion ($v > 1.2 \text{ m/s}$), engagement inference is conservatively capped at a *Neutral* state, reflecting reduced attentional capacity while walking. During high-speed locomotion ($v > 2.0 \text{ m/s}$), the system overrides the inferred score and forces a *Disengaged* state. Only after temporal smoothing and safety gating is the engagement estimate mapped to discrete engagement states, which subsequently drive the adaptive content mechanism.

3.5. Design Goals

Our system supports real-time adaptive content in VR cultural heritage environments while remaining interpretable and deployable on consumer hardware. Four design goals guided the architecture:

1. Adaptation relies solely on implicit behavioral signals from standard VR sensors, eliminating explicit input, calibration, or external instruments that could disrupt exploration.
2. Engagement inference uses transparent, rule-based classification with defined thresholds rather than opaque machine learning, ensuring interpretability and reproducibility.

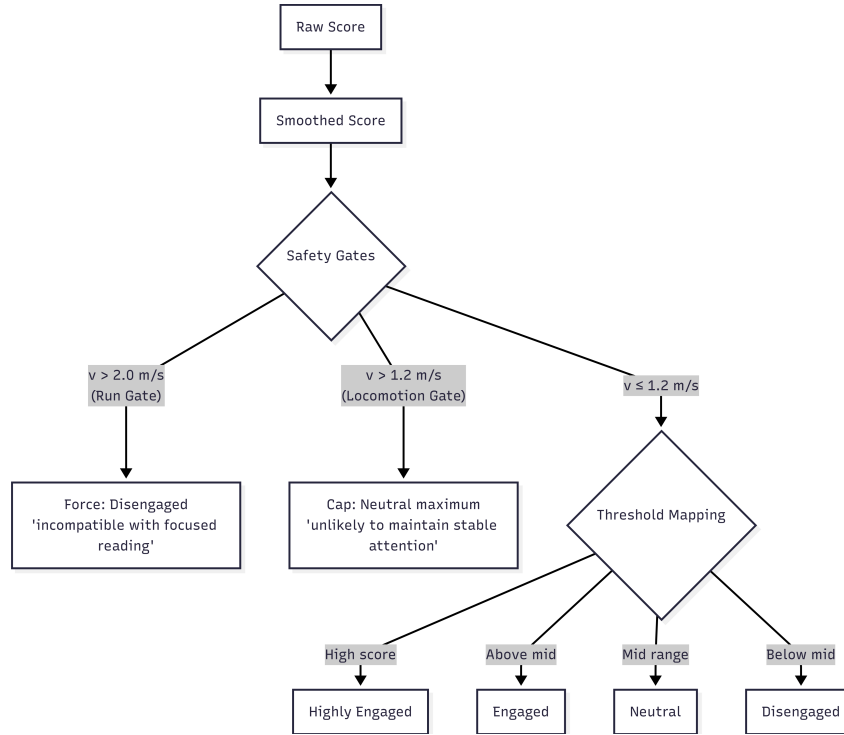


Figure 2: Temporal Smoothing and Safety Gates. The smoothed score passes through velocity-based safety gates before mapping to discrete engagement states.

3. Content adaptation operates under low-latency constraints compatible with VR frame rates, with sensing and inference executing in milliseconds to enable responsive adaptation.
4. The system supports scalable content variation by leveraging generative language models, enabling adaptation across arbitrary exhibits without per-item scripting.

3.6. Experimental Setup

To assess whether real-time engagement-driven content adaptation influences user behavior and content consumption in immersive cultural heritage environments, we conducted a controlled between-subjects user study comparing an adaptive VR museum system with a static baseline.

3.7. Participants

Sixteen participants (8 per condition), aged 13–34 years, 15 men and 1 woman, were recruited through university mailing lists and social media. Table 1 summarizes participant characteristics by condition. Mann–Whitney U tests confirmed that groups were balanced with respect to prior museum interest ($U = 30.0, p = .872$) and museum visit likelihood ($U = 28.5, p = .743$).

4. System Implementation

The system follows a continuous closed-loop cycle: behavioral sensing, engagement inference, adaptive content selection, and user re-interaction. At each frame, multimodal behavioral signals from the VR headset are processed to estimate engagement state.

The system was developed using **Unity 6000.2.8f1** and deployed on a **Meta Quest 3**. All core interaction loops—behavioral sensing, signal aggregation, and engagement inference—execute locally on the Snapdragon XR2 Gen 2 processor for low-latency performance. Figure 3 illustrates the study setup, virtual environment, and key interface components.

Table 1

Participant Characteristics by Condition

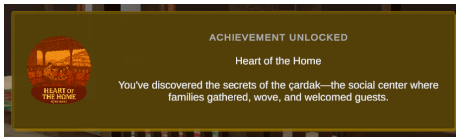
Characteristic	Adaptive (N=8)	Control (N=8)
<i>Prior VR Experience</i>		
Yes (once or twice)	5 (62.5%)	4 (50.0%)
No	3 (37.5%)	4 (50.0%)
<i>Controller Gaming Frequency</i>		
Frequently (weekly/daily)	4 (50.0%)	1 (12.5%)
Occasionally	2 (25.0%)	3 (37.5%)
Rarely (monthly or less)	2 (25.0%)	2 (25.0%)
Never	0 (0.0%)	2 (25.0%)
<i>Baseline Museum Attitudes, M (SD)</i>		
Interest in museums	3.12 (1.25)	3.25 (1.49)
Likelihood to visit	3.75 (1.04)	3.88 (1.36)



(a) Participant during study



(b) Virtual museum environment



(c) Achievements interface



(d) Object information panel

Figure 3: System implementation overview: (a) participant wearing the Meta Quest 3 headset, (b) the virtual museum environment, (c) the achievements and gamification interface, and (d) an example object information panel showing adaptive content.

4.0.1. Sensing and Inference

Signal acquisition occurs within Unity's update and physics loops. Gaze is approximated via forward raycasting against exhibit colliders, while head kinematics and locomotion velocity derive from pose deltas of the CenterEyeAnchor and CameraRig transforms. The engagement inference engine applies temporal smoothing and safety gates before updating the global state machine.

4.0.2. Asynchronous Content Pipeline

Content adaptation is decoupled from the rendering thread to prevent stalls. Upon engagement state transitions, the system issues asynchronous requests to the LLM (GPT-4o). To ensure consistent output, we utilize a two-part prompting strategy: a System Instruction that emphasizes adherence to length constraints (e.g., ‘You must stay within the requested word count’), and a dynamic User Prompt that adjusts the stylistic depth. As detailed in our supplementary material, this ranges from requesting a ‘scholarly style with rich detail and anecdotes’ for Highly Engaged users, down to ‘2-3 short bullet points under 10 words’ for Disengaged users. Generated explanations are cached using a composite key (exhibit ID + engagement level), enabling instant retrieval during revisits and reducing network overhead.

To concretely illustrate how engagement-driven prompting affects content presentation, Table 2 provides an example exhibit (a copper brazier) showing the contrast between static content and dynamically adapted outputs generated for different engagement states. These examples are illustrative and reflect the prompt strategies described above rather than fixed, hard-coded responses.

Table 2

Comparison of Content Adaptation (Example Exhibit: Copper Brazier)

Engagement Level	Prompt Strategy (Illustrative)	Generated Content Example	Words
Static / Neutral	“Provide essential historical facts in a clear and concise manner.”	This copper brazier with cover dates to the Renaissance period. It served as a portable heater for wealthy families during colder months. The intricate floral engravings signify the owner’s high social status.	32
Disengaged	“Format as 2–3 short bullet points, each under 10 words.”	Renaissance-era portable heater. Used by wealthy families. Floral engravings indicate high status.	13
Highly Engaged	“Scholarly style with rich detail; include contextual anecdotes.”	This brazier is a masterwork of Renaissance metalworking. Beyond its function as a portable heater, the floral motifs on the lid were reserved for elite families, indicating use in formal guest spaces to impress visitors.	46

5. Evaluation

5.1. Study Design

We employed a between-subjects design with two experimental conditions (N=16):

- **Adaptive (N=8):** The system continuously inferred user engagement from behavioral signals and dynamically adapted content depth in real time.
- **Control (N=8):** The system presented static, medium-length content for all exhibits without behavioral sensing or adaptation.

Both conditions used identical virtual environments, exhibit assets, interaction techniques, and navigation mechanics. The only difference between conditions was the presence of engagement-driven content adaptation.

5.2. Procedure

After providing informed consent, participants completed a brief demographic questionnaire and received standardized instructions on how to navigate and interact within the VR museum using a Meta Quest 3 headset. Participants were instructed to explore the museum freely, read exhibit descriptions as desired, and collect information cards distributed throughout the environment.

Each participant completed a single exploration session lasting up to 15 minutes. Following the VR experience, participants immediately completed post-experience questionnaires measuring usability, perceived presence, motivation to read, and interest in visiting the physical museum. Participants also provided optional qualitative feedback. The total session duration was approximately 20–30 minutes per participant.

5.3. Measures

5.3.1. Objective Behavioral Measures

The system automatically logged:

- **Session Duration:** Total time spent in the VR experience.
- **Reading Behavior:** Number of reading events, total reading view time, and estimated number of words exposed.
- **Interaction Outcomes:** Number of cards collected (maximum 18) and badges unlocked.
- **Engagement States:** Distribution of time across engagement classifications.

5.3.2. Subjective Measures

Usability was assessed using the System Usability Scale (SUS), producing scores from 0 to 100. Additional Likert-scale items (1–5) measured perceived presence, motivation to read exhibit text, navigation ease, and content appropriateness. Participants also reported whether the VR experience increased their interest in visiting the physical museum.

5.4. Results

Given the modest sample size ($N = 16$), Mann–Whitney U tests were used with Cohen’s d effect sizes reported. Table 3 presents behavioral outcomes and Table 4 presents engagement state distribution. Overall system usability was high, with a mean SUS score of 84.3 across both conditions.

Table 3
Behavioral Outcomes by Condition

Metric	Adaptive M (SD)	Control M (SD)	p	d
Session Duration (s)	889.4 (422.0)	451.4 (157.3)	.003**	1.38
Reading View Time (s)	473.4 (294.5)	164.1 (121.1)	.015*	1.37
Reading Events	67.2 (19.9)	50.1 (34.7)	.050*	0.61
Words Exposed	1515.5 (616.8)	745.1 (511.3)	.007**	1.36
Reading Motivation (1–5)	4.00 (0.76)	3.25 (0.89)	.083	0.91
Cards Collected	8.6 (5.4)	4.8 (3.7)	.140	0.84

* $p < .05$, ** $p < .01$; Mann–Whitney U tests

Subjective Measures. Perceived presence trended higher in the Adaptive condition (Adaptive: $M = 4.12$, $SD = 0.35$; Control: $M = 3.50$, $SD = 1.20$; $U = 42.5$, $p = .201$), suggesting that adaptive text generation did not disrupt immersion and may have enhanced it. Interest in visiting the physical museum also trended higher in the Adaptive condition (Adaptive: $M = 3.75$, $SD = 0.89$; Control: $M = 3.25$, $SD = 1.16$; $U = 40.0$, $p = .414$), though neither difference reached statistical significance

Table 4

Engagement State Distribution (% of Session Time)

Condition	Highly Eng.	Engaged	Neutral	Disengaged	Highly Dis.
Adaptive	24.3%	47.4%	22.2%	5.9%	0.3%
Control	13.6%	48.0%	28.4%	9.6%	0.3%

given the sample size. Due to the exploratory nature of the study, these self-report findings should be interpreted cautiously.

5.5. Discussion

Our preliminary results suggest that LLM-driven content pacing, guided by implicit multimodal sensing, can increase behavioral indicators of engagement in immersive cultural heritage experiences.

RQ1: To what extent can implicit behavioral signals support real-time attention-aware adaptation?

The results indicate that the selected implicit behavioral signals were sufficient to support stable, real-time adaptation of content depth without disrupting the user experience. While these signals do not provide direct access to users’ internal cognitive states, their combination enabled the system to consistently differentiate between scanning behavior and sustained reading readiness, allowing timely transitions between engagement states. The absence of usability degradation ($SUS > 84$) and the smooth operation of the adaptive pipeline suggest that attention-aware adaptation based on implicit signals is feasible within the latency and interaction constraints of consumer VR systems. These findings position implicit multimodal sensing as a practical foundation for attention-aware content pacing rather than a definitive measure of cognitive engagement.

RQ2: How does adaptation influence exploration behavior, content consumption, and motivation to engage?

The adaptive condition demonstrated higher positive engagement, with 71.7% of session time in combined *Highly Engaged* and *Engaged* states compared to 61.6% in the control condition. Notably, *Highly Engaged* time nearly doubled (24.3% vs. 13.6%), while combined negative engagement (*Disengaged* + *Highly Disengaged*) decreased by 37% (6.2% vs. 9.9%). The adaptive condition also reduced neutral or passive time from 28.4% to 22.2%, suggesting that users spent more time actively engaged rather than passively browsing.

These patterns indicate that the LLM-mediated adaptive scaffolding helped maintain flow by providing appropriately challenging content, reducing cognitive friction and disengagement while keeping users within productive learning zones [11, 20]. This observation aligns with prior work showing that adaptive and affect-aware systems can positively influence engagement during complex learning and exploratory tasks without disrupting user experience [8, 3].

Trust, Transparency, and Multimodal Interaction

While participants were not informed texts were AI-generated, none questioned content authenticity, though this raises ethical questions about transparency in AI-mediated cultural experiences. Scalability beyond our constrained domain would require safeguards against hallucination, potentially through retrieval-augmented generation grounding LLM outputs in curated knowledge bases [19]. Our multimodal sensing approach demonstrates that effective human-AI collaboration need not rely on explicit commands [5]. The success of purely implicit signals suggests embodied behavior serves as a rich communication modality, enabling proactive AI support without breaking presence [18]. However, future work should investigate user awareness of AI adaptation and desires for explicit control.

5.6. Limitations

The modest sample size ($N = 16$) limits statistical power for detecting non-significant effects, even when effect sizes are moderate to large ($d \approx 0.87$).

Our engagement inference relies on a fixed 4-second rolling window to smooth signal noise. Informal pilot observations suggested that shorter windows ($< 2s$) introduced instability due to transient glances, while longer windows ($> 6s$) led to perceived system responsiveness lag, highlighting the need for future personalization of temporal thresholds.

Additionally, metrics such as words exposed are intrinsically influenced by the intervention itself, as the adaptive system presents longer text to users inferred as more engaged; this measure therefore reflects both system behavior and user choices rather than engagement alone.

We did not assess knowledge retention or validate engagement classification against external ground-truth measures, which remains an important direction for future multimodal engagement modeling in immersive museum experiences [21, 5].

The 15-minute exploration sessions capture initial interaction patterns but do not reflect long-term engagement or learning effects. Finally, privacy implications of continuous behavioral sensing and AI-mediated adaptation warrant further investigation [13].

5.7. Conclusions and Future Work

This work introduces a closed-loop framework in which a Large Language Model responds to implicit multimodal behavioral signals in a VR cultural heritage environment, leading to substantially increased exploration time, reading engagement, and word exposure while maintaining high usability. Our findings suggest that AI-integrated XR systems can leverage multimodal behavioral data as practical signals for attention-aware adaptation, allowing the LLM to act as a dynamic mediator that adapts authored content to user capacity by balancing curated historical accuracy with personalized pacing. Future work should investigate long-term effects, explicit validation of engagement proxies, and the integration of multimodal foundation models that enable agents to suggest exhibits, answer questions, or provide collaborative annotations based on observed user interests.

Acknowledgments

We thank all participants for their time and valuable feedback. The virtual reconstruction was informed by publicly available materials and virtual tour content provided on the official website of the Berat National Museum.

Declaration on Generative AI

During the preparation of this work, the author(s) used OpenAI GPT-4o in order to: Generate adaptive exhibit content at runtime. Further, the author(s) used AI-assisted tools in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] A. Marto, A. Gonçalves, A systematic review of virtual 3d reconstructions of cultural heritage in immersive virtual reality, *Multimedia Tools and Applications* 83 (2024) 55879–55949. doi:10.1007/s11042-024-18700-3.
- [2] M. Konstantakis, G. Caridakis, A bibliometric analysis of immersive technology in museum exhibitions: exploring user experience, *Frontiers in Virtual Reality* 4 (2023) 1240562. doi:10.3389/frvir.2023.1240562.
- [3] R. U. Khan, Y. B. Oon, A. B. Madihie, M. H. B. Barawi, I. J. Hutasuhut, H. N. Ranudinata, D. D. Kristanto, From static displays to interactive ar: Evaluating the effectiveness of an ar app for gen z and alpha in museum pusaka, *Journal of Theoretical and Applied Information Technology* 103 (2025) 4069–4079.

- [4] M. Nasri, Towards intelligent vr training: A physiological adaptation framework for cognitive load and stress detection, *arXiv preprint arXiv:2504.06461* (2025) 419–423. doi:10.1145/3699682.3727575.
- [5] V. Clay, P. König, S. U. König, Eye tracking in virtual reality: A broad review of applications and challenges, *Virtual Reality* 27 (2023) 1481–1505. doi:10.1007/s10055-022-00738-z.
- [6] B. Steiner, I. Saada, N. Unkelos-Shpigel, Ar-institute: Enhancing sustainable manufacturing practices through gamified augmented reality, in: D. May, M. Auer (Eds.), *2025 Yearbook Emerging Technologies in Learning*, volume 59 of *Learning and Analytics in Intelligent Systems*, Springer, Cham, 2026. doi:10.1007/978-3-032-09058-4_17.
- [7] Berat National Museum, Ethnographic museum, <https://muzeumet-berat.al/en/ethnographic-museum/>, n.d. Accessed: 2026-01-16.
- [8] C. Yildirim, Transformative virtual reality games for adaptive sports training: Exploring the perceptions of individuals with mobility impairments, *International Journal of Human-Computer Interaction* 41 (2025) 1078–1090. URL: <https://doi.org/10.1080/10447318.2024.2311970>. doi:10.1080/10447318.2024.2311970. arXiv:<https://doi.org/10.1080/10447318.2024.2311970>.
- [9] R. Ivanov, Advanced visitor profiling for personalized museum experiences using telemetry-driven smart badges, *Electronics* 13 (2024) 3977. doi:10.3390/electronics13203977.
- [10] N. Unkelos-Shpigel, Heights of glory: Design thinking to enhance battle monuments visits via augmented reality, in: *Proceedings of AVICH 2024*, Genoa, Italy, 2024.
- [11] J. Lee, H. Jang, J. Lee, How do we read texts in vr?: The effects of text segment quantity and social distraction on text readability in virtual museum contexts, *IEEE Transactions on Visualization and Computer Graphics* 31 (2025) 9699–9708. doi:10.1109/TVCG.2025.3616803.
- [12] V. Y. Han, T. Wang, H. Cho, K. Todi, A. S. Fernandes, A. Levi, Z. Zhang, T. Grossman, A. Ion, T. R. Jonker, A dynamic bayesian network based framework for multimodal context-aware interactions, in: *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*, Association for Computing Machinery, New York, NY, USA, 2025, pp. 54–69. doi:10.1145/3708359.3712070.
- [13] Y. Tang, J. Situ, A. Y. Cui, M. Wu, Y. Huang, Llm integration in extended reality: A comprehensive review of current trends, challenges, and future perspectives, in: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA, 2025. doi:10.1145/3706598.3714224.
- [14] K. H. C. Lau, Designing and evaluating gen-ai for cultural resilience, in: *Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25)*, Association for Computing Machinery, New York, NY, USA, 2025, pp. 706–710. doi:10.1145/3716553.3750819.
- [15] T. Kuflik, O. Stock, M. Zancanaro, A. Gorfinkel, S. Jbara, S. Kats, J. Sheidin, N. Kashtan, A visitor's guide in an active museum: Presentations, communications, and reflection, *Journal on Computing and Cultural Heritage* 3 (2011). doi:10.1145/1921614.1921618.
- [16] L. Sidenmark, H. Gellersen, Eyehead: Synergetic eye and head movement for gaze pointing and selection, 2019, pp. 1161–1174. doi:10.1145/3332165.3347921.
- [17] S. Imamura, L. Jieun, J. Rekimoto, I. Makoto, Advantage of gaze-only content browsing in vr using cumulative dwell time compared to hand controller, in: *Proceedings of the 2023 ACM Symposium on Spatial User Interaction, SUI '23*, Association for Computing Machinery, New York, NY, USA, 2023. URL: <https://doi.org/10.1145/3607822.3614513>. doi:10.1145/3607822.3614513.
- [18] O. Janeh, N. Katzakis, J. Tong, F. Steinicke, Infinity walk in vr: Effects of cognitive load on velocity during continuous long-distance walking, in: *ACM Symposium on Applied Perception 2019, SAP '19*, Association for Computing Machinery, New York, NY, USA, 2019. URL: <https://doi.org/10.1145/3343036.3343119>. doi:10.1145/3343036.3343119.
- [19] Y. Cho, K. S. Park, A mobile augmented reality integrating KCHDM-based ontologies with LLMs for adaptive Q&A and knowledge testing in urban heritage, *Electronics* 15 (2026) 336. doi:10.3390/electronics15020336.
- [20] S. D'Mello, A. Graesser, Dynamics of affective states during complex learning, *Learn-*

ing and Instruction 22 (2012) 145–157. URL: <https://www.sciencedirect.com/science/article/pii/S0959475211000806>. doi:<https://doi.org/10.1016/j.learninstruc.2011.10.001>.

- [21] R. Sawyer, A. Smith, J. Rowe, R. Azevedo, J. Lester, Investigating visitor engagement in interactive science museum exhibits with multimodal bayesian hierarchical models, in: Proceedings of the 21st International Conference on Artificial Intelligence in Education (AIED), Springer, 2020, pp. 395–407. doi:[10.1007/978-3-030-52237-7_32](https://doi.org/10.1007/978-3-030-52237-7_32).