Adaptive Content Pacing in a Virtual Reality Museum: A Preliminary Evaluation of LLM-Driven Personalization

Naomi Unkelos-Shpigel[1,*], Wadea Tanus[1] and Joseph Damouni[1]

[1]Braude College of Engineering, Karmiel, Israel

Abstract

We present a closed-loop adaptive VR museum system that personalizes exhibit descriptions in real time based on implicit behavioral sensing and LLM-driven content generation. This experience paper reflects on the design rationale, implementation trade-offs, and qualitative findings from deploying a closed-loop adaptive VR museum at the Berat Ethnographic Museum ($N = 16$). While behavioral outcomes are reported in the companion preprint [3], we focus here on what we learned building and running the system: which design decisions proved consequential, where the system failed in pilot testing, and what qualitative patterns distinguished adaptive from static visitors. These practitioner insights complement the technical report and target designers of adaptive cultural heritage experiences.

Keywords

virtual reality museum, adaptive content, engagement-aware computing, implicit sensing, large language model, cultural heritage

1. Introduction

This paper accompanies a technical report [3] on a closed-loop adaptive VR museum system deployed in a high-fidelity reconstruction of the Berat Ethnographic Museum. Rather than repeating the full system description, we focus on what the deployment revealed: the design decisions that proved consequential, the failure modes encountered in pilot testing, and the qualitative patterns that static metrics alone do not capture. Readers seeking the complete engagement modeling architecture, sensor fusion details, and behavioral outcome tables are referred to [3].

This work presents a work-in-progress exploring a closed-loop adaptive VR museum framework. Our central question is: How can engagement-aware content adaptation enhance the VR museum visit experience compared to static delivery? We address two sub-questions: (RQ1) Can implicit behavioral signals support real-time, attention-aware adaptation within consumer VR hardware constraints? (RQ2) How does adaptation influence exploration behavior, content consumption, and motivation to engage with exhibits?

2. Related work

2.1. Adaptive Interfaces in Cultural Heritage

Despite broad interest in museum personalization, most existing approaches rely on pre-visit profiling or explicit visitor input [7, 11]. These methods assume stable user intent and fail to accommodate the rapid fluctuations in attention that characterize immersive VR exploration. The case for real-time implicit adaptation is therefore strong, yet VR-specific implementations remain scarce [9, 15, 17]. This gap motivated the design choices described below.

Workshop on Advanced Visual Interfaces for Cultural Heritage (AVICH 2026), June 08, 2026, co-located with the 18th International Conference on Advanced Visual Interfaces, Venice, Italy.

*Corresponding author.
naomius@braude.ac.il (N. Unkelos-Shpigel); Wadea.Tanus@e.braude.ac.il (W. Tanus); Joseph.Damouni@e.braude.ac.il (J. Damouni)
0009-0004-6634-5236 (N. Unkelos-Shpigel)

For implicit sensing, head angular velocity below 30°/s reliably predicts sustained visual attention [16]; gaze dwell of approximately 1 second distinguishes intentional focus from transient glances [6, 10]; and locomotion velocity inversely relates to attentional availability [8].

2.2. Large Language Models in Cultural Heritage

LLM integration in extended reality has been explored [18], but most deployments prioritize offline generation over fully real-time adaptation due to latency concerns [1]. Our system differs by targeting content depth pacing through implicit cues alone, without calibration, explicit input, or external instruments.

3. System Design

The system architecture is described in full in [3]; here we summarize only the elements essential for interpreting the design rationale and qualitative findings below. It implements a continuous closed-loop cycle: behavioral sensing, engagement inference, adaptive content selection, and visitor re-interaction [7], with 50 ms latency.

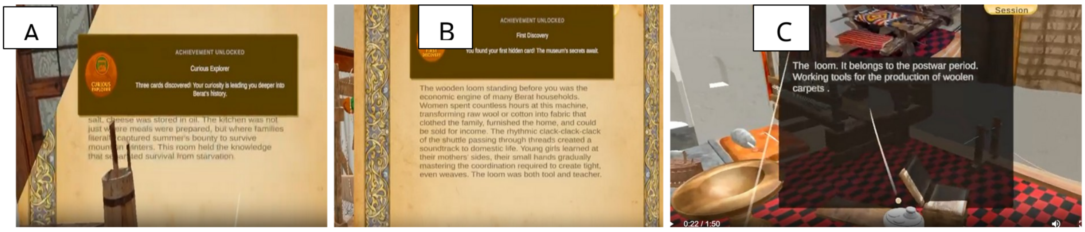

Figure 1: Berat VR selected screens. A: Neutral engagement state – concise exhibit description displayed on the loom. B: Highly Engaged state – rich scholarly description with gamification achievement overlay. C: Engaged state – standard-depth description with curiosity-unlocked card notification.

3.1. Engagement Modeling

Four signals are fused into a composite engagement score (see [3] for full specification): head stability, gaze dwell, locomotion velocity, and reading dwell. The final output is one of four states: Highly Engaged, Engaged, Neutral, or Disengaged.

Several threshold choices proved consequential in piloting. A smoothing coefficient of $\alpha = 0.35$ was selected after $\alpha = 0.15$ caused rapid oscillation between states visible to observers, while $\alpha = 0.60$ introduced a 4–5 second lag that users described as “the text not keeping up.” Rule-based classification was preferred over a trained model because our pilot sample of six participants was too small to train reliably, and the rules’ interpretability allowed us to diagnose a critical early bug: the locomotion gate was initially set at 0.8 m/s, which mis-classified slow walkers as Disengaged. Raising it to 1.2 m/s resolved this. Figure 2 elaborates the model flow:

3.2. Adaptive Content Pipeline

GPT-4o is called asynchronously on engagement state transitions, with a two-part prompt that enforces length constraints and adjusts stylistic depth (full prompt in [3]). Responses are cached by (exhibit ID + engagement level). Two issues emerged in piloting: (1) GPT-4o occasionally violated word-count constraints for Disengaged outputs, requiring an explicit “You must stay within X words” enforcement instruction. (2) Network latency occasionally caused a blank panel for 1–2 seconds before adapted text appeared; future versions should pre-cache the Neutral tier at exhibit entry.

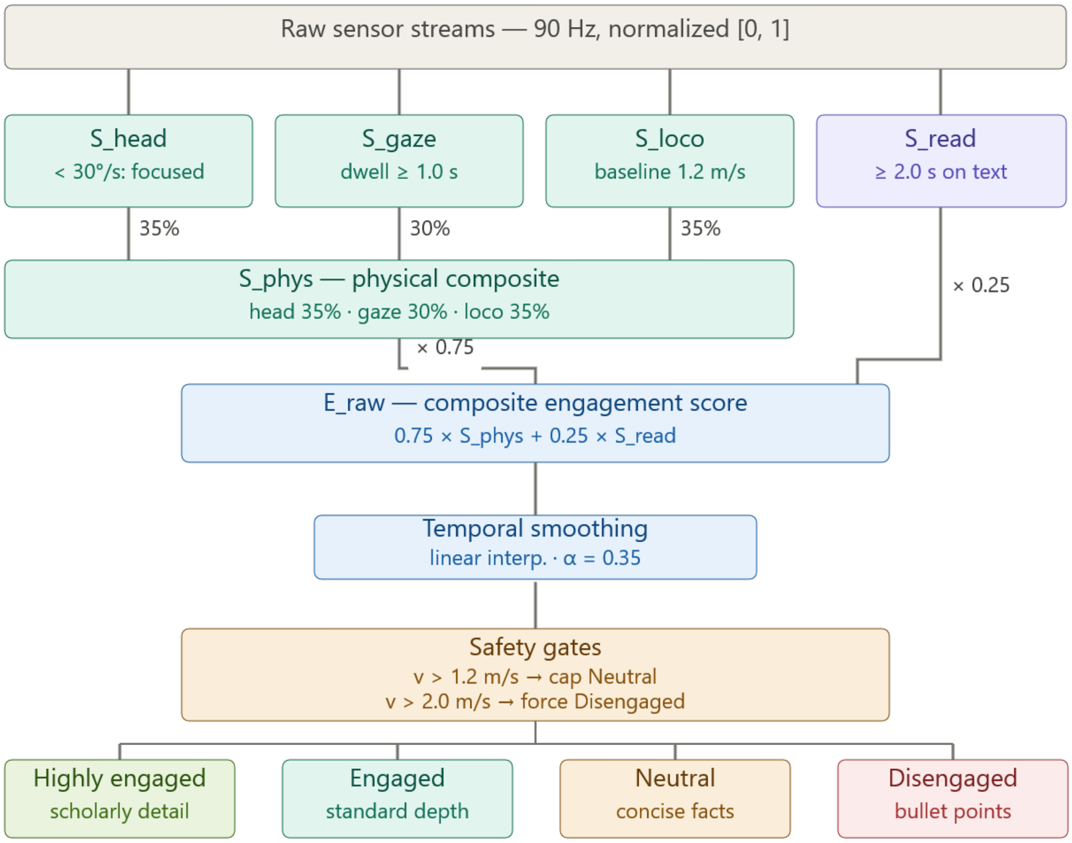

Figure 2: Engagement model flow [3]

4. Method

4.1. Study Design

Sixteen participants (N = 8 per condition) completed a 15-minute free-exploration session and post-experience questionnaires; full demographic details and group balance tests are reported in [3]. The sole manipulation was engagement-driven content adaptation versus static medium-length descriptions.

4.2. Usability and Subjective Measures

Behavioral outcomes (session duration, reading view time, words exposed) are reported in full in [3]. Here we focus on the subjective and qualitative dimensions that the preprint covers only briefly. The Adaptive condition yielded a mean SUS of 84.7 (SD = 10.9) versus 78.1 (SD = 14.6) for Control (U = 42.5, p = .292, d = 0.51). The overall mean of 81.4 falls in the 'good' range, and the absence of significant degradation confirms that real-time adaptation does not impair usability. *Table 2, referenced here for subjective outcomes across ten experiential dimensions, was not present in the source manuscript and should be inserted by the authors.* Reading motivation reached trending significance (p = .083, d = 0.91). Large effect sizes also emerged for navigation effort (d = 0.91) and perceived presence (d = 0.71), though neither reached significance at this sample size. All eight Adaptive participants reported increased interest in visiting the physical museum after the experience, compared to five of eight Control participants. We note that engagement classification was not validated against external ground-truth measures such as physiological signals or think-aloud protocols; this remains a priority for future work.

Table 1
Content adaptation by engagement state – example exhibit (Copper Brazier)

State	Prompt Strategy	Generated Example (Copper Brazier)	Design Note
Disengaged	2–3 bullet points, ≤10 words each	Renaissance-era portable heater. Used by wealthy families. Floral engravings indicate high status.	Bullet format minimizes reading load; early pilot used a single sentence, but users skipped it entirely.
Neutral	Essential facts, concise	This copper brazier dates to the Renaissance period. It served as a portable heater for wealthy families. The floral engravings signify high social status.	Serves as the baseline; mirrors what the Control received throughout.
Highly Engaged	Scholarly, rich detail and anecdotes	A masterwork of Renaissance metalworking, the floral motifs on the lid were reserved for elite families, indicating use in formal guest spaces to impress visitors.	Anecdote framing was preferred over pure facts in pilot feedback; increased dwell time measurably.

4.3. Qualitative Findings

Adaptive participants most frequently recalled specific exhibit content (the loom room, weaving room, brazier collection), while Control participants more often recalled the VR medium itself. Only 2 of 8 Adaptive participants noticed that descriptions varied across objects, indicating unobtrusive adaptation. One Adaptive participant noted: "once you get into exploring it becomes extremely engaging since it also ignites the spirit of a gamer" (P058). Control participants surfaced design problems directly addressable by LLM generation: one identified repetitive content structure across exhibits (P073), another reported confusion when some objects lacked information cards (P061). Six of eight Adaptive and seven of eight Control participants reported mild dizziness, reflecting a general locomotion VR challenge rather than an effect of the adaptive system.

The contrast in what participants recalled is notable: Adaptive visitors anchored memories to content ('the weaving room', 'the brazier with the floral pattern'), while Control visitors anchored to form ('the floating text', 'the button you press'). This suggests that when content depth matched attention, the medium became transparent. Conversely, the Control condition's static text made the delivery mechanism itself salient. The two Control participants who specifically noted repetitive structure (P073) and missing cards (P061) identified precisely the failure modes that LLM generation is designed to solve: structural monotony and coverage gaps.

5. Discussion and Conclusions

RQ1: The engagement classifier operated stably without degrading usability, and the low detection rate of content variation (2/8 Adaptive participants) confirms that adaptation remained perceptually transparent. This positions implicit multimodal sensing as a practical foundation.

RQ2: The most consistent signal was trending improvement in reading motivation ($d = 0.91$), consistent with the hypothesis that appropriately scaled descriptions sustain reading engagement. Adaptive participants recalled specific exhibit content rather than the medium, suggesting deeper cultural engagement. The Control condition further motivates LLM generation: repetitive static structure and inconsistent coverage were the primary frustrations, both of which generative adaptation directly addresses.

Practitioners considering similar systems should note three lessons from this deployment. First, the smoothing window is the most sensitive parameter: too short and the system flickers; too long and visitors perceive lag before the system 'notices' them. A 4-second rolling window worked for our exhibit pacing, but denser exhibit layouts may require shorter windows. Second, discoverability of interactive elements is a prerequisite for measuring adaptation effects: one Control participant spent three minutes before discovering object interactivity, which skewed their behavioral data. Onboarding that explicitly demonstrates one interaction reduces this noise. Third, caching adapted responses by (exhibit, engagement state) is essential for latency on network-limited deployments, but invalidates the cache if the LLM prompt is updated mid-study.

Limitations include a small and demographically narrow sample, absence of physiological ground truth for engagement validation, and a discoverability gap for interactive elements (one Control participant required three minutes to discover object interactivity). Future work will expand the participant pool, integrate longitudinal multi-session tracking, validate engagement classification against external ground-truth measures, and examine whether explicit override controls can improve perceived agency without sacrificing adaptation benefits. Extension to outdoor archaeological sites and art museums is planned after further validation.

Declaration on Generative AI

During the preparation of this work, the authors used Claude.ai in order to: Grammar and spelling check. Claude (Anthropic) assisted in manuscript structure, literature synthesis, and technical descriptions. After using these tools and services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content. All research design, data collection, analysis, and interpretation remain the authors' original contributions.

References

[1] B. L. Cho and K. S. Park, "A mobile augmented reality integrating KCHDM-based ontologies with LLMs for adaptive Q&A and knowledge testing in urban heritage," *Electronics*, vol. 15, no. 2, 336, 2026.

[2] V. Clay, P. König, and S. U. König, "Eye tracking in virtual reality: A broad review of applications and challenges," *Virtual Reality*, vol. 27, pp. 1481–1505, 2023.

[3] J. Damouni, W. Tanus, and N. Unkelos-Shpigel, "Adaptive Virtual Reality Museum: A Closed-Loop Framework for Engagement-Aware Cultural Heritage," *arXiv preprint arXiv:2603.13639*, 2026.

[4] E. Dim and T. Kuflik, "Automatic detection of social behavior of museum visitor pairs," *ACM TIIS*, vol. 4, no. 4, article 17, 2014.

[5] V. Y. Han et al., "A dynamic Bayesian network based framework for multimodal context-aware interactions," in *IUI 2025*, pp. 54–69, 2025.

[6] S. Imamura et al., "Advantage of gaze-only content browsing in VR using cumulative dwell time," in *SUI 2023*, 2023.

[7] R. Ivanov, "Advanced visitor profiling for personalized museum experiences using telemetry-driven smart badges," *Electronics*, vol. 13, no. 20, 3977, 2024.

[8] O. Janeh et al., "Infinity walk in VR: Effects of cognitive load on velocity during continuous long-distance walking," in *SAP 2019*, 2019.

[9] R. U. Khan et al., "From static displays to interactive AR: Evaluating the effectiveness of an AR app for Gen Z and Alpha," *JTAIT*, vol. 103, no. 10, 2025.

[10] M. Konstantakis and G. Caridakis, "A bibliometric analysis of immersive technology in museum exhibitions," *Frontiers in Virtual Reality*, vol. 4, 1240562, 2023.

[11] T. Kuflik et al., "A visitor's guide in an active museum," *JOCCH*, vol. 3, no. 3, article 11, 2011.

[12] J. Lanir et al., “The influence of a location-aware mobile guide on museum visitors’ behavior,” *Interacting with Computers*, vol. 25, no. 6, 2013.

[13] J. Lee et al., “How do we read texts in VR? Effects of text segment quantity and social distraction on text readability in virtual museum contexts,” *IEEE TVCG*, vol. 31, no. 11, 2025.

[14] A. Marto and A. Gonçalves, “A systematic review of virtual 3D reconstructions of cultural heritage in immersive VR,” *Multimedia Tools and Applications*, vol. 83, 2024.

[15] K. Sawyer, “Group Flow and Group Genius,” *NAMTA Journal*, vol. 40, no. 3, pp. 29–52, 2015.

[16] L. Sidenmark and H. Gellersen, “Eye & Head: Synergetic eye and head movement for gaze pointing and selection,” in *UIST 2019*, 2019.

[17] J. Sheidin, L. Ankava, N. Iyov, and N. Unkelos-Shpigel, “Bringing the sea back to life: an augmented reality voyage in a maritime heritage museum,” *Computing*, vol. 1, p. 2, 2024.

[18] Y. Tang et al., “LLM integration in extended reality: A comprehensive review,” in *CHI 2025*, 2025.

[19] N. Unkelos-Shpigel, “Heights of Glory: Design Thinking to Enhance Battle Monuments Visits via Augmented Reality,” in *AVICH@AVI*, 2024.