Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36886

Adaptive Interaction Design for Immersive Spatial Experience Using Artificial Intelligence

Shijia Zhang

College of Art and Design, Glasgow School of Art, Glasgow, UK
zhangshijia0418@163.com

Abstract. This study develops an artificial-intelligence-driven method for adaptive interaction design in immersive spatial environments. A closed-loop architecture links multimodal behavior perception, Transformer-based user-state interpretation, rule-constrained strategy selection, Unity-based spatial adjustment, and ten-second feedback evaluation. Forty participants completed fixed and adaptive versions of the same 15-minute immersive gallery using HTC Vive Pro Eye. Gaze, head motion, hand interaction, navigation, hesitation, and task progress were used to infer attention, engagement, operational proficiency, and exploration tendency. The adaptive condition reduced task completion time from 196.4 ± 28.7 s to 158.6 ± 22.9 s***, reduced failed interactions from 5.42 ± 1.31 to 2.67 ± 0.94***, and increased explored-area ratio from 0.68 ± 0.09 to 0.82 ± 0.07***. User-state interpretation achieved 92.1 ± 1.0% accuracy and 0.908 ± 0.012 macro-F1. Presence, usability, personalization, interaction continuity, and perceived control were also significantly higher under adaptation. The results indicate that state-conditioned modification of content depth, interaction difficulty, guidance, lighting, and spatial audio can improve immersive interaction without disrupting environmental continuity or user autonomy.

Keywords: adaptive interaction design, immersive spatial experience, artificial intelligence, user-state modeling, virtual reality

1. Introduction

Immersive spatial experience emerges from continuous coupling between perception, navigation, manipulation, environmental feedback, and the user's changing cognitive and behavioral state. Current virtual reality systems can record rich interaction traces, but many environments still respond through predefined triggers that provide essentially the same content depth, difficulty, and spatial guidance to users with different attention, confidence, and exploration patterns. Work on real-time VR personalization demonstrates that implicit interaction data can be used to modify virtual environments during use rather than only before a session [1]. Eye tracking is particularly valuable because fixation, gaze direction, and visual exploration provide precise indicators of attention and interaction intention [2], while reviews of immersive eye-tracking applications show that fixation duration and task-performance measures are among the most widely used behavioral indicators [3].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.36886

Adaptive design must nevertheless extend beyond state detection. Presence and usability are sensitive to how users experience the virtual environment and its feedback [4], while challenge-skill mismatch can reduce flow when interaction difficulty exceeds the user's capability [5]. Dynamic difficulty systems therefore increasingly use real-time user information to regulate challenge [6], and adaptive VR training studies show that personalization and algorithm responsiveness can affect both task acquisition and flow [7]. Gaze-directed real-time scene control also demonstrates that perceptual information can regulate visual parameters without interrupting interaction [8]. More broadly, immersive interfaces influence presence, performance, and user experience together rather than as isolated outcomes [9], while recent reviews show that virtual environments can be personalized through movement, facial, behavioral, and other user responses [10].

The remaining gap is the absence of a design-oriented closed loop that explicitly connects inferred state to a bounded spatial intervention and then evaluates whether the intervention improved the targeted state. This study addresses that gap by integrating four user states with three coordinated mechanisms: adaptive content, adaptive difficulty and guidance, and adaptive atmosphere and navigation. Its innovation lies in treating AI inference as an intermediate design signal whose effect is constrained, spatially implemented, and behaviorally re-evaluated before further adaptation.

2. Literature review

The design of immersive interaction integrates attention, body movement, manipulation of objects, navigation, and feedback with constantly perceived environment. While trigger-response structures provide reliable interaction patterns, they fail to account for variation in attention span, level of confidence, skills, and exploratory behavior. Recent research suggests that personalization in real time might enhance virtual environment design through continuous adaptation of interaction in accordance with constantly gathered user data [1]. The eye tracking data, in turn, demonstrates that gaze direction, fixation duration, and visual search behavior serve as useful indicators of attention and intent to interact in an immersive environment [2, 3]. Presence and usability also depend on the degree of match between environmental feedback and user actions and expectations [4].

The modeling of user state enables the connection between observed behavior and design adaptation. Challenge-skill mismatch might have a negative impact on the flow and quality of interaction [5], while dynamic difficulty adjustment allows regulating the complexity of tasks depending on the detected user states [6]. Adaptive VR training proves the positive effect of the algorithmic personalization depending on assistance in response to constant performance [7]. Gaze directed scene control finally proves the possibility of regulating visual parameters dynamically in accordance with perceptual signals without breaking interaction continuity [8].

These results prove that adaptive immersive design should not be limited to detection of user states. The detected states should be translated into specific and bounded changes of the environment. Immersive interfaces impact presence, performance, and general experience at once [9], while the recent personalization research proves the importance of integrating behavioral and environmental components in real time [10]. Therefore, the current study connects attention, engagement, operational skill, and exploratory tendency to three corresponding mechanisms: adaptive content presentation, adaptive interaction difficulty and guidance, and adaptive environment atmosphere and navigation. The feedback cycle evaluates the impact on behavior after the adaptation and forms a closed cycle with adaptation maintenance, reduction or reversal depending on its effect.

DOI: 10.54254/2755-2721/2026.36886

3. Methodology

3.1. Closed-loop adaptive interaction design framework

The system is made up of five processes that include behavior detection, state interpretation, strategy selection, spatial adaptation, and feedback analysis. The HTC Vive Pro Eye detects gaze, head orientation, controller motion, and navigation at 90 Hz. Unity events include object selection, manipulation failure, route repetition, and task completion. The features are combined into windows of two seconds, and then they are passed to Python 3.11 through WebSocket, where a Transformer encoder implemented in PyTorch analyzes the last 12 seconds of behavior using Equation (1)-(2):

$$\boldsymbol{p}_t = softmax\left[\boldsymbol{W}_s LayerNorm\left(\sum_{\tau=0}^{5}\alpha_{t,\tau}\boldsymbol{h}_{t-\tau}\right)+\boldsymbol{b}_s\right] \tag{1}$$

$$\alpha_{t,\tau} = \frac{exp[\boldsymbol{q}^T tanh(\boldsymbol{W}_h\boldsymbol{h}_{t-\tau}+\boldsymbol{W}_c\boldsymbol{c}_t)]}{\sum_{r=0}^{5} exp[\boldsymbol{q}^T tanh(\boldsymbol{W}_h\boldsymbol{h}_{t-r}+\boldsymbol{W}_c\boldsymbol{c}_t)]} \tag{2}$$

Here, $\boldsymbol{h}$ denotes Transformer hidden states, $\boldsymbol{c}_t$ contains current task progress, and $\boldsymbol{p}_t$ represents attention, engagement, proficiency, and exploration probabilities. State confidence below 0.60 produces no major adaptation, preventing noisy predictions from repeatedly altering the environment. As shown in Figure 1, the proposed framework integrates multimodal behavior perception, Transformer-based user-state interpretation, rule-constrained strategy selection, spatial adjustment, and feedback evaluation into a closed-loop adaptive interaction process.

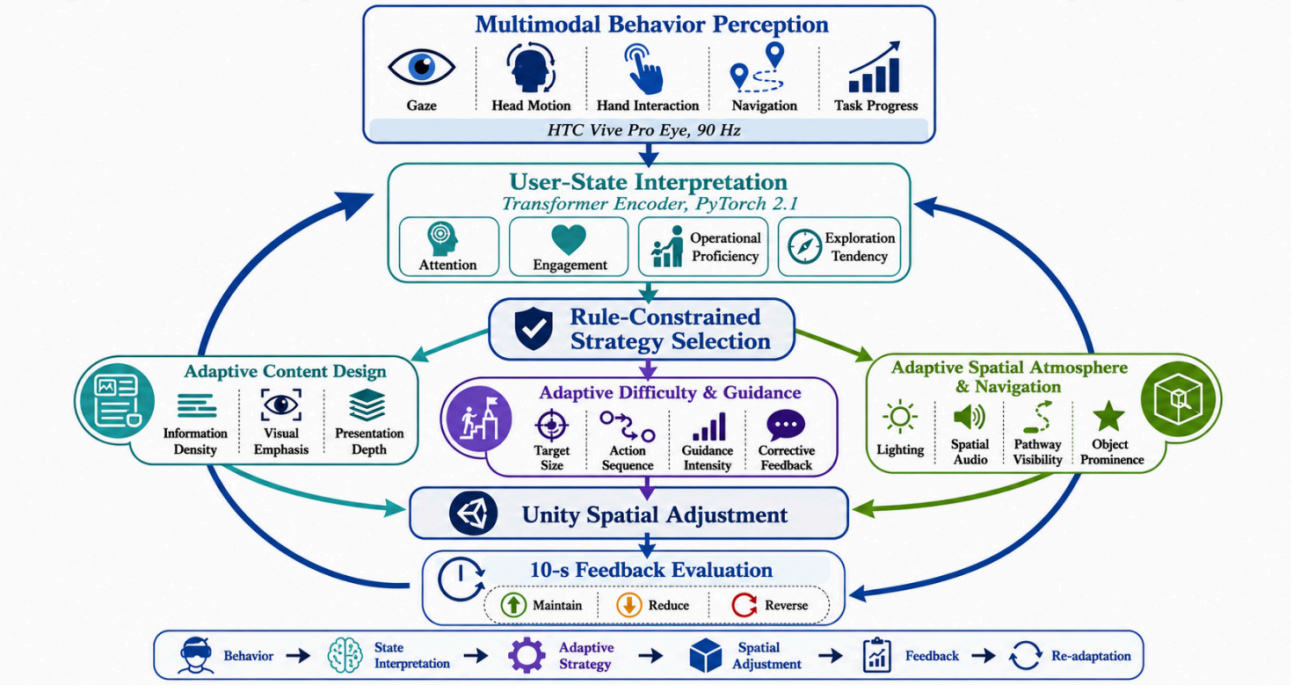

Figure 1. Closed-loop AI-driven adaptive interaction framework for immersive spatial experience

DOI: 10.54254/2755-2721/2026.36886

3.2. Adaptive content and information design

Each interactive object has three information layers managed by Unity: a visual indication, a concise layer, and an interactive layer. Activation of the concise layer happens when the duration of fixation surpasses 3 seconds and user proximity drops below 1.5 meters. Further, with the continuous fixation and manipulation, there will be gradual revelation of the interactive layer. In case of the probability of user attention dropping below 0.45, secondary labels and decorative elements will start fading within 1.2 seconds, while principal object's luminance will rise by up to 18%. Changes in content will not alter object's position and size, maintaining the visual hierarchy of objects.

3.3. Adaptive interaction difficulty and guidance design

Proficiency in operation is determined based on normalized completion time, incorrect selection, controller path deviations, and repeated tries. The lower levels of proficiency increase the size of target colliders up to 15% to 25%, decrease action sequences by one unit, add trajectory hints, and make it possible to provide immediate feedback. After two consecutive successes in actions, the help level is lowered by one unit. The high proficiency level of users will allow decreasing the target tolerance, using fewer hints, and performing complex sequences of manipulations. Changes in difficulty can be performed only at one level per task stage and only once in 10 seconds.

3.4. Adaptive spatial atmosphere and navigation design

Exploration tendency affects pathway brightness, object saliency, audio orientation, and illumination in Unity 2022.3. The state of low exploration induces a light boost on the unexplored path of about 20-30% and places a directional audio cue towards the closest unexplored area. Route repeat detection includes a low contrast floor marker; the high exploration state removes explicit cues and enhances peripheral-object visibility. The closed-loop adaptive spatial control process modifies environmental parameters based on the probabilities of the current user-state and their behavior-related features. Each adaptive step involves a combination of predicted attention, engagement, skillfulness, and exploration states along with current behavioral cues in order to define the suitable adaptation action to take. Then, the action is converted into bounded parameter modifications of the content presentation, interaction difficulty, illumination, spatial audio, and guidance. The new parameters partially preserve the previously established environmental state for continuity purposes, whereas the discrepancy between the target state and the observed state of the user within the next ten seconds is used as the feedback signal. When the intervention improves the desired state, the adaptation persists; when the state becomes stable, the adjustment is gradually decreased; and when the performance decreases, the system either cancels or minimizes the effect of the adaptation.

4. Experimental design

4.1. Experimental environment and interaction tasks

A 15-minute immersive gallery designed in Unity consists of three zones: object inspection, spatial navigation, and manipulation sequence. The adaptive and non-adaptive implementations utilize the same geometry, textures, sounds, objects' positions, and goals of the tasks to be performed. In contrast, only the adaptive implementation employs the state-dependent alterations of the complexity of information provided, interaction support, lighting, sounds, and visibility of paths. The task list includes the examination of the information of layered objects, finding two non-

DOI: 10.54254/2755-2721/2026.36886

contiguous exhibits, performing a manipulation sequence consisting of three steps, and freely exploring other parts of the environment.

4.2. Participants and experimental procedure

A total of forty participants with either unimpaired or corrected-to-unimpaired vision are tested under each condition using a random within-subjects design. A familiarization phase lasting for five minutes using HTC Vive Pro Eye is carried out prior to data collection. The task stimuli and route sequence are counterbalanced to minimize the effects of practice, with a 10-minute resting period separating the two conditions. Each trial starts with calibration of the eye tracker and a thirty-second baseline period in which neutral navigation takes place.

4.3. Evaluation metrics and statistical analysis

Objective variables include completion time, unsuccessful transactions, redundancy of the path taken, percentage of area explored, duration of object manipulation, and adaptation delay. The state prediction is analyzed through accuracy, macro-F1, class-specific F1, calibration error, and transition stability. The subjective analysis involves Igroup Presence Questionnaire, System Usability Scale, personalization, continuity of interaction, and perceived control. SPSS 29 is used to conduct paired-samples t-test if the difference is normally distributed; otherwise, the Wilcoxon signed-rank test is employed. Multiplicity is controlled by Holm correction. Effect size is measured using Cohen's dz and rank biserial correlation, while reliability is measured using Cronbach's alpha. Adaptation consistency is the percentage of commands for which the desired state changes in the right direction.

5. Results

5.1. Accuracy and stability of user-state interpretation

The Transformer reached 92.1 ± 1.0% accuracy and 0.908 ± 0.012 macro-F1. Attention, engagement, proficiency, and exploration F1 scores were 0.923 ± 0.014, 0.901 ± 0.017, 0.914 ± 0.013, and 0.895 ± 0.019, respectively. Expected calibration error was 0.037 ± 0.008, while the confidence gate reduced unstable state reversals from 2.86 ± 0.41 to 0.94 ± 0.18 events/min***.

5.2. Performance of content, difficulty, and spatial adaptation

Adaptive interaction reduced completion time and failure frequency while increasing exploration and object engagement. Mean adaptation response latency was 84.6 ± 11.8 ms, and 86.9 ± 3.2% of interventions produced the intended behavioral change within the 10-s feedback window (see Table 1).

Table 1. Objective performance under fixed and adaptive interaction

Metric	Fixed	Adaptive
Task completion time (s)	196.4 ± 28.7	158.6 ± 22.9*
Failed interactions	5.42 ± 1.31	2.67 ± 0.94*
Route redundancy	0.31 ± 0.08	0.18 ± 0.06*
Explored-area ratio	0.68 ± 0.09	0.82 ± 0.07*

DOI: 10.54254/2755-2721/2026.36886

Table 1. (continued)

Object-engagement duration (s)	41.8 ± 7.6	52.4 ± 8.1*
Task-path deviation (m)	1.84 ± 0.42	1.29 ± 0.31*

*Note: *Values are mean ± SD. *p < .001 after Holm correction.*

5.3. Comparison of immersion, usability, and interaction continuity

The adaptive condition produced higher scores across all experience measures. Presence increased to 5.81 ± 0.46***, while SUS reached 86.2 ± 5.7***. The largest standardized improvement occurred for perceived personalization, followed by interaction continuity and perceived control.

6. Conclusion

This experiment shows how the adaptive experience can be accomplished through a closed-loop mechanism where AI-powered interpretation of user state serves as a design layer to mediate between behavior and space adaptation. Multimodal eye movements, body position, interaction actions, and navigation serve as accurate indicators of attention, engagement, skill level, and exploration tendency, and adaptation based on rules serves as a way to make controlled changes in information depth, interaction complexity, guidance, lighting, sound, and pathway visibility. An adaptive environment makes task performance, exploration, immersion, usability, personalization, consistency, and perceived control better but does not change tasks and spatial layout. It is important that feedback evaluation is an essential part of the experiment, since the process of adaptation is not seen as a unidirectional process of response to prediction. Rather, the adaptive system checks whether a certain spatial intervention works for a particular user state.

References

[1] Valmorisco, S., Raya, L., & Sanchez, A. (2024). Enabling personalized VR experiences: A framework for real-time adaptation and recommendations in VR environments. *Virtual Reality, 28*, 128.
[2] Adhanom, I. B., MacNeilage, P., & Folmer, E. (2023). Eye tracking in virtual reality: A broad review of applications and challenges. *Virtual Reality, 27*(2), 1481–1505.
[3] Shadiev, R., & Li, D. (2023). A review study on eye-tracking technology usage in immersive virtual reality learning environments. *Computers & Education, 196*, 104681.
[4] Voinescu, A., Petrini, K., & Stanton Fraser, D. (2023). Presence and simulator sickness predict the usability of a virtual reality attention task. *Virtual Reality, 27*, 1967–1983.
[5] Lemmens, J. S., & Freiherr von Münchhausen, C. (2023). Let the beat flow: How game difficulty in virtual reality affects flow. *Acta Psychologica, 232*, 103812.
[6] Orozco-Mora, C. E., Fuentes-Aguilar, R. Q., & Hernández-Melgarejo, G. (2024). Dynamic difficulty adaptation based on stress detection for a virtual reality video game: A pilot study. *Electronics, 13*(12), 2324.
[7] Verniani, A., Galvin, E., Tredinnick, S., Putman, E., Vance, E. A., Clark, T. K., & Anderson, A. P. (2024). Features of adaptive training algorithms for improved complex skill acquisition. *Frontiers in Virtual Reality, 5*, 1322656.
[8] Cebeci, B., Askin, M. B., Capin, T. K., & Celikcan, U. (2024). Gaze-directed and saliency-guided approaches of stereo camera control in interactive virtual reality. *Computers & Graphics, 118*, 23–32.
[9] Gronowski, A., Arness, D. C., Ng, J., Qu, Z., Lau, C. W., Catchpoole, D., et al. (2024). The impact of virtual and augmented reality on presence, user experience and performance of information visualisation. *Virtual Reality, 28*, 133.
[10] Linares-Vargas, B. G. P., & Cieza-Mostacero, S. E. (2025). Interactive virtual reality environments and emotions: A systematic review. *Virtual Reality, 29*, 3.