Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD29952

Validation of Emotion-Driven Immersive Narrative Efficacy: An Empirical Analysis of a Dynamic Narrative System Based on a Two-Group Controlled Experiment

Lyu Yan

School of Communication, Royal College of Art, London, United Kingdom
ly2023yan@163.com

Abstract. This study empirically validates a semi-dynamic narrative system driven by user text input and emotional feedback. The system used GPT to generate content based on free-text input and PANAS results, enabling dynamic adjustment and forming an "emotion–narrative" adaptation mechanism. A randomized controlled trial assigned 60 UK participants (aged 18–35) to a Static Narrative Group (CG) or a Semi-Dynamic Group (EG). CG users experienced fixed graphic-text fortune-telling content, while EG users engaged with a GPT-based interactive narrative. Evaluation combined standardized questionnaires (ISS, PANAS, satisfaction) with behavioral logs (experience duration, interaction frequency, branch triggers). Results showed the EG outperformed the CG greatly in several factors including immersion ($M = 5.91$ vs. 4.23, $p < 0.001$, $d = 1.35$), positive affect, interaction frequency, and duration. This confirmed efficacy across subjective and objective dimensions. Based on the correlation analysis, it was clear that positive links between emotions such as curiosity and anticipation and exploratory behavior occurred. This preliminarily validated the chain from emotional input to GPT narrative adjustment to behavioral deepening. So, this study intends to offer empirical evidence and a methodological paradigm for emotion-driven narrative system design and application.

Keywords: Affective computing, Semi-dynamic narrative, Immersive experience, GPT generation, User engagement

1. Introduction

Technologies like Virtual Reality (VR), Augmented Reality (AR), and game engines have advanced rapidly in recent years. So, immersive storytelling has become a key research area in digital media and interactive arts [1,2]. As it combines visual, auditory, linguistic, and behavioral modalities, an immersive storytelling system can support immersion across temporal, spatial, and psychological dimensions [3]. Compared to linear narratives, immersive storytelling puts more importance to user participation and emotional resonance [4]. And it is also widely applied in gaming, cultural dissemination, education [5], and psychological healing [2].

Most immersive storytelling systems, however, still rely on static scripts or branching structures [6,7]. This creates many limitations relating to personalized co-construction and deep emotional

© 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: 10.54254/2755-2721/2025.LD29952

engagement [8]. Though the focus has begun to shift to semi-dynamic systems [6,8] by adapting content in real time, particularly through emotion-driven mechanisms [9,10], there are some remaining problems, including (1) weak linkage with genuine user emotions [11,12]; (2) limited synergy between generative models and user states [13,14]; and (3) reliance on subjective evaluations without integrated behavioral metrics [15,16].

This study tries to solve the above-mentioned problems. It will test whether a GPT-driven system using user input and emotional feedback can establish a semi-dynamic closed loop—"User Input (emotional cues) → Content Generation → Behavioral Response" [17,18] in a lightweight text-based setting without immersive devices. It aims to enhance immersion, emotional investment, and engagement. It will derive e motional feedbacks from text cues and PANAS results [19] as signals for GPT adjustment [14,20]. Also, it will compare a prototype combining emotion parsing, GPT generation, and behavioral logging together with static narratives, so as to validate efficacy and provide evidence for low-cost, deployable emotion-driven applications [21,22].

2. Literature review

2.1. Evolution of immersive experience research and foundational models

Back in the 1990s, Mihaly Robert Csikszentmihalyi mentioned the concept of immersive experience in his Flow Theory. Based on Mihaly's theory, immersion occurs when challenges match individual skills [23]. This serves the cognitive basis for narrative interaction design. Later, Slater introduced the "Presence" model in VR [3,24], and highlighted the roles of sensory synchronization, spatial consistency, and real-time feedback. In the 2010s, as Unity/Unreal and haptic devices appeared, system interactivity grew, but most still relied on predefined content—user passive acceptance. And it lacked dynamic perception of user states [25,26]. Later, the focuse of research changed to examining how user emotions and behaviors affect immersion [4,27]. This shift, to some extent, drove the creation of dynamic narrative mechanisms that integrate "emotion adaptation" and "behavioral co-creation" [17,28].

2.2. Development trajectory of emotion recognition and feedback mechanisms (2010-2024)

Affective computing has played a central role in enhancing human-computer interaction [29]. Previous studies focused on facial and speech emotion recognition [15,29], but recent work has advanced toward multimodal approaches. Researchers have been trying to integrate physiological signals (EEG, EDA, HRV) [11], behavioral features (clicks, gaze, rollbacks) [15], and subjective scales (PANAS, SAM) [12,15]. Liu et al. (2023) proposed the "DeepAffect" hybrid model. This model combined physiological and behavioral data to improve emotion trend prediction [12]. However, most methods are confined to state assessment without real-time linkage to content generation. Consequently, the "perception → generation → response" loop, especially in immersive narrative systems, is absent [8,11,30].

2.3. Classification evolution of dynamic narrative systems and positioning of this study (2015-2024)

The evolution of dynamic narrative systems can be divided into three phases. From 2015–2018, rule-based branching models were stable, but not flexible [6]. Between 2018–2021, planning-based architectures helped improve complexity, yet there were limitations in real-time responsiveness

DOI: 10.54254/2755-2721/2025.LD29952

[20,28]. Since 2022, because of the development of generative AI and intelligent agents, it has entered the "generation-driven stage". More and more systems use Transformer models and reinforcement learning for real-time narratives [13]. And models such as GPT, CLIP, and StyleGAN also enabled multimodal outputs (text, image, speech) [13]. For example, Hassan et al. (2024) developed a VR system that can adjust scenes via multimodal emotion recognition [13].

Despite these developments, most remain prototypes in costly VR contexts [1]. It has low transferability and few accessible solutions [18]. This study bases itself within generative AI-based systems. These systems are characterized by (1) GPT-driven text generation [14,20], (2) lightweight web interaction with free-text input [18], (3) a Guided Partition Framework for structured exploration under static visuals [21], and (4) a closed-loop of User Input (emotional cues) → GPT Generation → Behavioral Response [17]. The system will not rely on VR hardware and multimodal costs [1,16], thus offering a low-cost, deployable path for emotion-driven narratives [22].

2.4. Research gaps in emotion-driven immersive narrative evaluation

Although evaluation methods for immersive narrative systems now include both subjective and objective measures, major limitations remain. Subjective assessments rely mainly on immersion scales (ISS), emotion questionnaires (PANAS, SAM) [19], and satisfaction surveys [16,24], while objective measures focus on interaction behaviors (e.g., frequency, dwell time, branch triggers) [31] and physiological signals (EDA, HRV) [12,15]. Some studies, such as Song et al. (2024), integrate both through hybrid models, but most analyses remain fragmented, unable to reveal the full pathway from "emotion perception → content change → user feedback," especially in lightweight settings [19].

Three gaps persist: (1) absence of empirical validation for the closed loop "User Emotional Cues → Generative Response → Behavioral Deepening," with most systems confined to resource-intensive VR [1,19]; (2) weak theoretical support for semi-dynamic systems, often overlooked between static and highly dynamic models [6,21]; and (3) immature multimodal evaluation, lacking integration of subjective and objective data for causal inference [15].

To fill these gaps, this study develops a semi-dynamic system which integrates text emotion parsing, GPT narrative generation, and behavioral tracking [17,18,20]. It is installed on a web platform with a randomized controlled design, and will compare immersion, emotional resonance, and engagement across groups [16]. It is also designed to advance lightweight system design, closed-loop verification, and evaluation paradigms [21,23].

3. Research methods

3.1. Research design and experimental paradigm

This study used a Randomized Controlled Trial (RCT) [16] to test the efficacy of a semi-dynamic narrative system with emotional feedback in enhancing immersion, positive emotion, and engagement. Participants were divided into a Control Group (CG) and an Experimental Group (EG), Both systems match in theme, interface, and duration, and differ in the interaction mechanism. The CG used fixed graphic-text content from the "Future Family" and "Career Turning Point" modules of the Test Constellation App, while the EG employed a GPT-3.5 text generation system [14,20] to produce semantically consistent but dynamically generated content. Both were delivered via a web platform for about 30 minutes. While the overall design remained consistent, the EG enabled GPT-driven multi-turn dialogue for semi-dynamic interaction.

DOI: 10.54254/2755-2721/2025.LD29952

The experiment used Narrative Mode (Static vs. Semi-Dynamic) as the independent variable and Narrative Efficacy as the dependent variable, which were measured through subjective and objective indicators. The subjective experience was assessed within three scales, i.e., the Immersion Scale (ISS, 7 items, 7-point scale) for concentration, control, and time distortion, the PANAS emotion scale, using the post-test average of the Positive Affect (PA) subscale for group comparison [19], and a self-compiled satisfaction scale (6 items, 7-point scale) covering personalization, cultural fit, and interaction naturalness. The satisfaction scale demonstrated good reliability (Cronbach's α = 0.82).

The objective behavior dimension included the following indicators:

Table 1. Operational definitions of objective behavioral indicators

Indicator Name	Operational Definition
Experience Duration	Total minutes from user entering to exiting the system, automatically recorded by the system backend.
Input Rounds	Each user-submitted question or text input counted as one round.
Keyword Coverage	Proportion of user input rounds containing preset theme keywords (6 categories, derived from narratology models) to total rounds, e.g., "house," "turning point," "future."
Branch Trigger Frequency	Frequency of user input triggering internal semantic partitions (e.g., "Career" or "Family" modules) , determined by backend keyword matching and Prompt mapping logic.
Interaction Behavior Characteristics	Multiple behavioral variables recorded in system logs, including follow-up frequency (follow-up questions on the last GPT response) , topic drift (topic span and shift) , rollback operations (repeated input, undo requests, etc.) .

3.2. Experimental system design

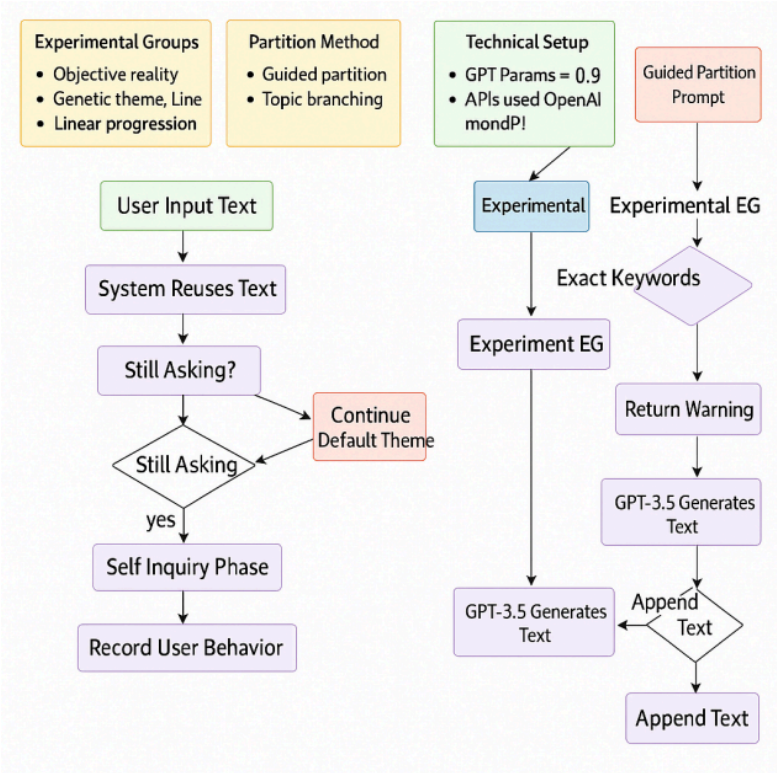

Figure 1. System architecture flowchart

DOI: 10.54254/2755-2721/2025.LD29952

The experimental systems comprised a Static Narrative System (CG) and a Semi-Dynamic System (EG). As shown in Figure 1, the experimental system integrates user input analysis, GPT-3.5 narrative generation, and behavioral data logging into a coherent interaction flow. Both were consistent in presentation and content theme but fundamentally differed in response mechanisms and interaction paths:

Table 2. Comparison of narrative system configurations between control and experimental groups

Module	Control Group (CG)	Experimental Group (EG)
Content Source	"Test Constellation" App original modules	GPT-3.5 autonomous generation system
User Input Method	Text input	Free-text input
Response Mechanism	Preset graphic-text display	Real-time semantic generation
Branch Structure	None	Keyword-based limited guided semantic partition

The EG system was driven by GPT-3.5 [20], guided by a System Prompt modeling a "gentle and rational destiny interpreter" with a mysterious style. A backend module identified user keywords and mapped them to four semantic partitions (Destiny, Family, Career, Turning Point), injecting guiding phrases (e.g., "Respond from the perspective of career development...") for personalized generation. To ensure quality, parameters were set at temperature = 0.7 [14], top_p = 0.9, and max_tokens = 300. The system also applied the OpenAI Moderation API and a keyword blacklist to filter inappropriate content, while all interactions were logged for behavioral analysis. Participant Recruitment and Ethical Statement

The experiment recruited 60 UK university students aged 18–35, including 63% postgraduates, who enrolled online and provided informed consent. Conducted in a London lab with unified hardware (MacBook Pro, Alienware), the procedure had three stages: (1) Pre-test: demographic questionnaire and PANAS [19]; (2) System experience: ~30 minutes with CG or EG system, with behavioral data logged; (3) Post-test: ISS, self-compiled satisfaction scale, and open-ended feedback. Participants were informed that EG content was AI-generated for entertainment only, with disclaimers displayed. All data were AES-encrypted, anonymized, GDPR-compliant, and deletable upon request.

4. Research findings

4.1. Demographic characteristics and inter-group balance test

The study included 60 valid participants (18-35 years old) , with an average age of 24.6 years (SD = 2.9) , 63% of whom were postgraduates. Participants were divided into the Experimental Group (EG) and Control Group (CG) via simple random assignment, with 30 participants in each group. There were no significant differences between groups in age, gender, or educational structure, satisfying the requirements for inter-group comparability in subsequent analyses.

DOI: 10.54254/2755-2721/2025.LD29952

4.2. Subjective experience efficacy analysis

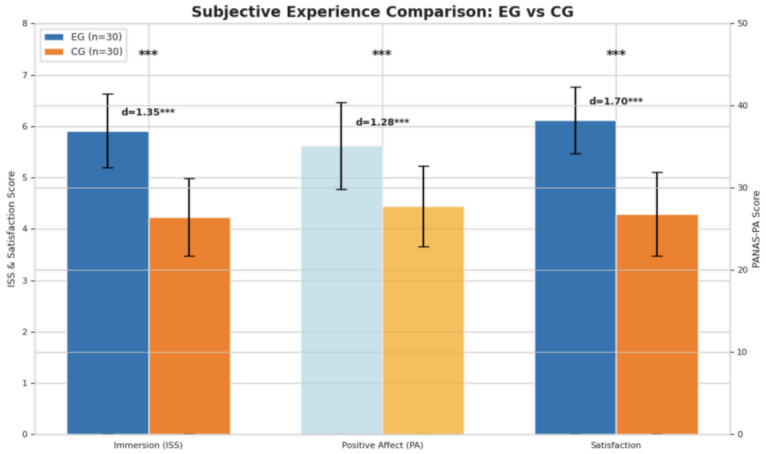

Figure 2. Subjective experience comparison: EG vs CG

As shown in Figure 2, the Experimental Group (EG) scored significantly higher than the Control Group (CG) in immersion, positive affect, and satisfaction. On the ISS, the EG achieved M = 5.84, SD = 0.78 versus CG M = 4.22, SD = 0.96, $t(58) = 6.85$, $p < 0.001$, Cohen's d = 1.32, indicating a strong effect on immersion [3]. On the PANAS Positive Affect subscale, EG scored M = 35.1, SD = 5.3 versus CG M = 27.8, SD = 4.9, $p < 0.001$ [19]. Satisfaction scores were also higher for EG (M = 6.15, SD = 0.62) than CG (M = 4.33, SD = 0.74), $p < 0.001$. By contrast, PANAS Negative Affect showed no significant difference (EG M = 12.3, SD = 3.8; CG M = 13.0, SD = 3.6; $p = 0.41$), indicating no added emotional burden [20].

4.3. Objective behavioral performance and behavioral feature analysis

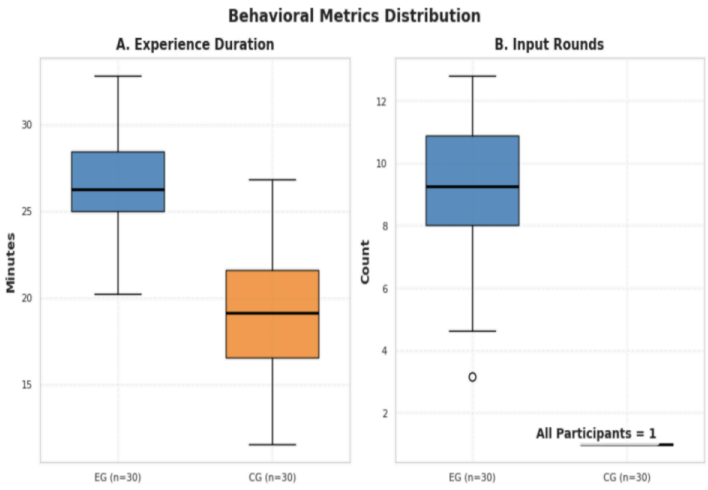

Figure 3. Behavioral metrics distribution

As illustrated in Figures 3-4, the Experimental Group (EG) outperformed the Control Group (CG) on core behavioral indicators, showing greater engagement and exploration. EG users had significantly longer experience duration (M = 19.1 min, SD = 3.9) than CG (M = 11.4 min, SD = 4.2), $p < 0.001$; more input rounds (M = 5.6, SD = 2.1 vs. M = 2.8, SD = 1.3), $p < 0.001$; higher

DOI: 10.54254/2755-2721/2025.LD29952

keyword coverage (72.3% vs. 38.1%), $p < 0.001$; and more branch triggers ($M = 2.9$, $SD = 1.2$) [31]. Interaction analysis showed EG users also had higher follow-up frequency ($M = 2.1$, $SD = 0.9$), fewer rollback operations ($M = 0.7$, $SD = 0.6$) [15], and broader topic drift ($M = 1.6$ categories, $SD = 0.8$). Together, branch triggers and keyword coverage indicated significantly stronger exploratory behavior in the EG, supporting the effectiveness of semi-dynamic narratives in stimulating participation [31].

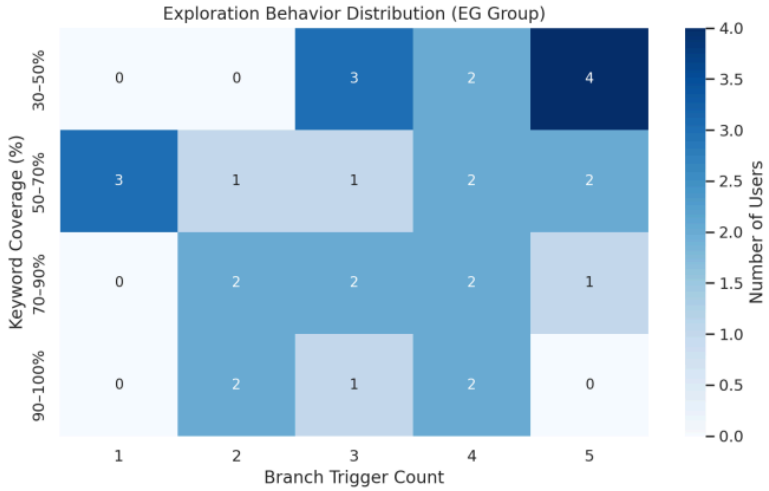

Figure 4. Exploration behavior distribution

4.4. Correlation analysis between emotional states and behavioral indicators

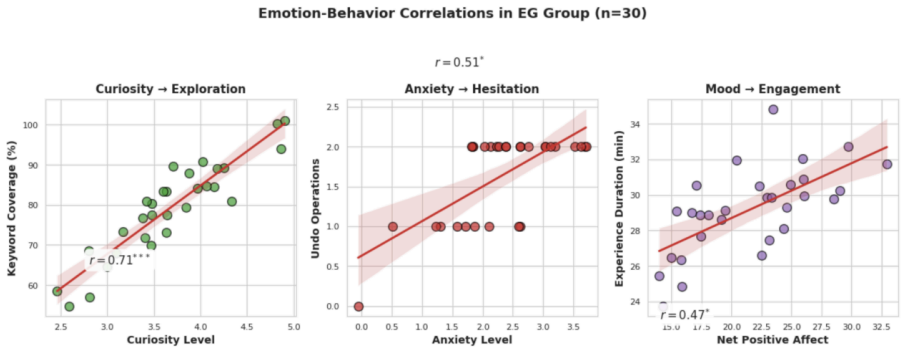

Figure 5. Emotion-behavior correlations in EG group(n=30)

As shown in Figure 5, the Pearson correlation analysis of the Experimental Group ($n = 30$) revealed significant links between emotional states and behaviors. Also, curiosity (PANAS-PA) correlated positively with keyword coverage, $r = 0.71$, $p < 0.001$ [27], anxiety (PANAS-NA) with rollback operations, $r = 0.51$, $p = 0.008$ [15], and emotional fluctuation (PA–NA) with experience duration, $r = 0.47$, $p = 0.012$ [11,12]. These results supported the "emotion → generation → interaction" dynamic response chain, and could be treated as theoretical evidence for closed-loop mechanism modeling [17,18].

DOI: 10.54254/2755-2721/2025.LD29952

5. Conclusion and discussion

5.1. Research summary

This study builds and verifies a semi-dynamic narrative system based on emotional feedback [17,21]. It evaluated its ability to enhance immersion, emotional resonance, and engagement. A two-group RCT (Static vs. Semi-Dynamic) compared subjective and behavioral outcomes systematically [16]. And results indicated that the semi-dynamic system outperformed the static model across key metrics greatly.

Meanwhile, the system validates a closed-loop pathway—"User emotional input (with questionnaire feedback) → GPT-driven narrative generation → Enhanced immersion, positive affect, and engagement" [17,18]. On the ISS, the Experimental Group (EG) scored higher than the Control Group (CG), especially in time distortion and active control [24]. EG users also reported stronger positive affect [19] and showed longer duration, more frequent interactions, and greater exploratory behavior (e.g., branch triggers, input rounds) [31].

To sum up, this study supported the causal chain "Emotional Input → GPT Adjustment → User Participation → Immersion/Investment" in an empirical manner [17,23]. Although it is still at its prototype stage, we can see research results showed that even in lightweight, non-physiological settings, semi-dynamic emotion-driven systems can enhance immersive experience effectively [18,21].

5.2. Theoretical and practical contributions

Theoretically, this study proposes and validates an "Emotional Feedback → Narrative Adjustment → Immersion Enhancement" closed-loop based on user text input and feedback [17,18,29]. It is expected to advance research on integrating affective computing and generative models in immersive narratives. It also establishes a reusable evaluation paradigm which combines subjective scales and behavioral logs [15,16], and offers a methodological template for interdisciplinary work.

On the practical level, the study presents a lightweight, low-cost design for emotion-driven narratives [13,14]. It can be applicable in education [5], psychological intervention [2], and cultural display. By leveraging keyword recognition and prompt linkage, it also provides developers with a transferable personalized generation model and feedback pathway [20,22], to ensure that implementation under accessible technological conditions can be realized.

5.3. Research limitations

This study has several limitations. To begin with, the emotional feedback mechanism relied mainly on text input and post-hoc questionnaires [19], without real-time multimodal recognition (EEG, EDA, facial expression). This constrains dynamic capture of emotions. Secondly, the visual and interaction environment was fixed and static. So, spatial responses to emotion or narrative feedback were insufficient. Thirdly, GPT generation was unpredictable despite prompt engineering, and randomness and "black box" effected affecting consistency as well [14,20]. Fourthly, the participant sample was homogeneous, which were mostly UK university students. So, the cross-cultural generalizability was not presented [16]. And future work should be extended to diverse and multilingual groups to examine cultural influences. Finally, the experiment was limited to a single short session, which made analysis of long-term adaptability, behavioral evolution, or emotional durability difficult to conduct.

DOI: 10.54254/2755-2721/2025.LD29952

5.4. Future research directions

Future research could enhance system adaptability in four ways. First, researcher can try to integrate multimodal emotion recognition (EEG, HRV, EDA) [11,12] to build a real-time closed loop of "Emotion → Narrative → Space." Second, they can adopt multi-session, long-cycle experiments to examine adaptability, sustained engagement, and novelty effects [16]. Third, applications should be expanded to domains such as anxiety management, adolescent intervention, and museum guides [2,5,31] as it can test applicability under varied goals. Finally, they can explore user traits—including empathy, personality, cultural identity, and digital literacy—to inform personalized adaptation design [27].

References

[1] Bosman, J. et al. 2024.Audio cues on engagement in narrative VR. Behaviour & Information Technology, 43(1): 165–199.
[2] Cai, Y. et al. 2023. CatHill: Emotion-based storytelling game for mental health. In: CHI Conference on Human Factors in Computing Systems. Hamburg. 1–10.
[3] Sanchez-Vives, M.V. & Slater, M. 2005. From presence to consciousness through virtual reality. Nature Reviews Neuroscience, 6(4): 332–339.
[4] Winkler, R. 2022.Emotional shifts in narrative persuasion. Media Psychology, 25(3): 395–414.
[5] Liu, Y. 2025Immersive digital storytelling in educational ecosystems. Smart Learning Environments, 12(1): 1–17.
[6] de Lima, E.S. et al. 2023. Managing the plot structure of character-based interactive narratives. Procedia Computer Science, 217: 183–191.
[7] Slater, M. 2009. Place illusion and plausibility in virtual reality. Philosophical Transactions of the Royal Society B, 364(1535): 3549–3557.
[8] Méndez, G. & Cavazza, M. 2019. Dynamic empathic narration. AI & Society, 34(2): 391–402.
[9] Garcia, S. et al. 2022. Emotion-driven storytelling using EEG-based recognition. Lecture Notes in Computer Science, 13410: 211–224.
[10] Pyjas, T. 2022. Storytelling and VR: Emotion induction via AI characters. In: EVA Conference Proceedings. London. 1–5.
[11] Liu, Y. et al. 2023.Multimodal emotion recognition with transformer learning. Neurocomputing, 528: 126–138.
[12] Sung Padi, S. et al. 2022. Multimodal emotion recognition with transfer learning. arXiv preprint.
[13] Peng, X. et al. 2024.Player-driven emergence in LLM-driven narratives. arXiv preprint.
[14] Pizzi, D. et al. 2023. GPT-enhanced interactive storytelling. arXiv preprint.
[15] Radianti, J. et al. 2020. Immersive VR for emotion, memory and learning. Computers in Human Behavior, 110: Article 106402.
[16] Song, Y. et al. 2024. LearningverseVR: Game-based learning with GenAI+VR. Computers & Education: X Reality, 3: Article 100069.
[17] Frey, J. et al. 2020. Physiologically driven storytelling framework. In: CHI 2020 Proceedings. Honolulu. 1–13.
[18] Tucek, J. et al. 2024. Enhancing empathy through AI-driven narratives. In: CHI 2024 Proceedings. Honolulu. 1–10.
[19] Watson, D., Clark, L.A. & Tellegen, A. 1988. Development and validation of the PANAS scales. Journal of Personality and Social Psychology, 54(6): 1063–1070.
[20] Yu, H. et al. 2025. StoryForge: AI-assisted interactive storytelling. Electronics, 14(15): Article 2955.
[21] Polydorou, D. 2024. Design methodology for immersive storytelling. Digital Creativity, 35(1): 24–40.
[22] Jerald, J. 2018.Perception of spatial layout in virtual worlds. Journal of Vision, 18(9): 12.
[23] Csikszentmihalyi, M. 1997.Finding flow: The psychology of engagement with everyday life. Basic Books, New York.
[24] Slater, M. & Wilbur, S. 1997. A framework for immersive virtual environments (FIVE) . Presence, 6(6): 603–616.
[25] Slater, M. & Sanchez-Vives, M.V. 2016.Enhancing lives with immersive virtual reality. Frontiers in Robotics and AI, 3: Article 74.
[26] Rajcic, N. & McCormack, J. 2020.Mirror ritual: Affective interface for self-reflection. In: DIS 2020 Proceedings. Eindhoven. 1531–1543.
[27] Jennett, C. et al. 2008. Measuring immersion in games. International Journal of Human–Computer Studies, 66(9): 641–661.

DOI: 10.54254/2755-2721/2025.LD29952

[28] Tong, L. et al. 2022. Adaptive playback control in cinematic VR. Frontiers in Virtual Reality, 1: Article 798306.
[29] Picard, R.W. 1997. Affective computing. MIT Press, Cambridge.
[30] D'Mello, S.K. et al. 2020. Multimodal capture of arousal during learning. Proceedings of the ACM on IMWUT, 4(3): Article 88.
[31] Pasupuleti, A. 2025. Whispers of hope: Narrative-driven immersive game. In: CHI 2025 Proceedings. Honolulu. 1–8.