Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

Exploring Multimodal Digital Narratives and Cognitive Construction in Immersive Media

Huaijin Ren, Xiaoyan Du

How to cite: Ren H, Du X. Exploring Multimodal Digital Narratives and Cognitive Construction in Immersive Media. Textile & Leather Review. 2026; 9:1688-1705. https://doi.org/10.31881/TLR.2026.1688

How to link: https://doi.org/10.31881/TLR.2026.1688

Published: 25 April 2026

Exploring Multimodal Digital Narratives and Cognitive Construction in Immersive Media

Huaijin Ren[1,2*], Xiaoyan Du[1]

[1] School of Marxism, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

[2] School of Design, Xi'an Technological University, Xi'an 710021, Shaanxi, China

*renhuaijin@xatu.edu.cn

Article

https://doi.org/10.31881/TLR.2026.1688

Published 25 April 2026

ABSTRACT

Aiming at the problem that digital narration is not perfect for users' cognitive construction in immersive communication environment, particularly within application scenarios involving smart textiles and virtual fashion content, this paper puts forward a multimodal evolution mechanism of digital narration in immersive communication based on MICU. Firstly, a multimodal evolution path model is constructed and compiled as a systematic framework with four core variables: modal density (M)-interaction frequency (I)-contextual resonance (C)-Depth of Cognitive Construction (U). Then, a controlled user experiment was designed, and the interactive experience and cognitive feedback of 40 users were collected through three typical immersive media platforms: visual short film, multimodal interactive drama and AI-driven narrative generation system, which simulates digital interactive environments of textiles and garments with varying levels of complexity, and an analysis data set was constructed. Finally, multiple linear regression and structural equation are used to establish the path model of evolution mechanism and make empirical verification. The results show that M (β = 0.351, p = 0.001), I (β = 0.279, p = 0.009) and C (β = 0.414, p < 0.001) all have significant positive effects on U, and the contextual resonance is between M and U (indirect effect = 0.154, CI). 0.216]), and there is a significant hierarchical distinction in user cognitive construction. The average U value of the generation-oriented users (0.54) is significantly higher than that of the reproduction-oriented users (0.22). This study provides a quantitative model for optimizing the interaction design of smart textiles and virtual garments, enabling more effective stimulation of users' higher-order cognition and value creation.

KEYWORDS

AI, immersive communication, digital narrative, user cognitive construction, smart textiles

INTRODUCTION

With the rapid advancement of artificial intelligence technology, the widespread application of AI narrative engines, extended reality devices, and virtual interaction platforms (particularly in the fields of smart textile design, interactive garment presentation, virtual try-on, and personalized fashion content generation) is profoundly transforming the generative logic of digital content and users' perceptual patterns. The new generation of immersive communication spaces is jointly constituted by multimodal input, high-frequency interaction, and context-driven mechanisms. Users no longer passively receive content but actively participate in the co-construction of meaning through dynamic co-creation and reconfiguration [1-3]. However, existing research predominantly describes immersive communication at the level of technological form or perceptual experience, lacking a quantitative model that examines how modal and interaction-related input conditions shape users' cognitive construction outcomes [4,5]. Effectively correlating modal complexity, interactive behavior, and user cognitive output represents a pressing theoretical challenge requiring breakthroughs in contemporary digital narrative communication research. It is crucial for understanding how users acquire and process complex multimodal information (visual, tactile, and auditory) from smart wearable devices, multisensory fabrics, or immersive virtual fashion experiences, and ultimately generate high-value cognitive outcomes such as design innovation inspiration or deep brand identification.

In recent years, digital narratives have developed rapidly in terms of their generation mechanisms, dissemination structures, and user responses, with narrative structures progressively evolving towards computability, predictability, and co-constructability [6]. Norrie et al. regarded modal fusion intensity as a technical indicator of immersion, demonstrating its significance in immersion research [7]. However, this study remained confined to the level of 'perceptual responses' and failed to further track whether perceptual stimuli were transformed into structural cognitive processing. Díaz E P posits that theories such as Narrative Transportation and Flow experience provide explanatory foundations for cognitive engagement within interactive narrative contexts [8]. Yet this study relies on subjective scales to evaluate immersive experiences, lacking a cognitive transition hierarchy grounded in actual output content. Furthermore, while contextual resonance serves as a key factor influencing comprehension depth, its mediating role within pathway mechanisms remains structurally unvalidated. Consequently, this paper introduces 'contextual resonance' as an intermediary variable. Through structural equation modelling, it validates its positional mechanism between multimodal input and cognitive output, transforming 'subjective experience' into a 'pathway link'. The proposed 'MICU' mechanism framework achieves a systematic cognitive construction depth in immersive narrative dissemination, providing measurable and interpretable theoretical support for user construction patterns within digital communication environments, and provides an operational quantitative model for textile and apparel design, demonstrating how optimized configurations of multimodality (M) and interactivity (I) can effectively evoke users' emotional resonance (C) and higher-order cognition (U).

THE CONSTRUCTION OF COGNITIVE MECHANISMS FOR DIGITAL NARRATIVES

Construction of the Cognitive Mechanism Framework

Multimodal integration, interactive behavior, and situational engagement constitute pivotal factors influencing immersive experiences. Presently, there exists a lack of structured models integrating these elements within cognitive mechanisms, particularly within digital narrative contexts where their role in users' 'meaning-making' processes remains unsystematically validated [9]. Guided by the 'multimodal input-emotional mediation-depth of cognitive construction' framework, an operational and verifiable cognitive mechanism model has been constructed. Within this model: 'Scenario Resonance Intensity (C)' serves as a psychological mediator for emotional transference and semantic immersion, forming a transitional mechanism between information input and cognitive output; ultimately, 'Depth of Cognitive Construction (U)' functions as the outcome variable reflecting users' deep understanding, meaning reconstruction, and value generation, thereby indicating the actual transformative effect of narratives on users' cognitive structures.

Experimental Environment and Case Design

To validate the constructed cognitive mechanism model for digital narratives, this study designed a controlled experimental environment, established a digital narrative interaction platform, and configured three types of typical immersive narrative cases within this platform. These cases simulate immersive communication scenarios involving different modalities of input and varying intensities of interaction, enabling systematic verification of the research hypotheses. The digital narrative interaction platform was developed within a controlled laboratory environment. It integrates functions including content control, behavioural recording, and automated data extraction, enabling comprehensive tracking of multimodal input and interaction operations throughout immersive narrative scenarios. Case design adheres to the dual-gradient construction principle of modal input intensity and interaction complexity within immersive communication research, ensuring diverse experimental scenarios effectively replicate real-world user interactions. Specific case designs are detailed in Table 1:

Table 1. Case Design

Case Attribute	Characteristics of Modal Density (M)	Characteristics of Interaction Frequency (I)	Description of Content Design
Case A: Visual Short Film	Unimodal (visual + audio) with single-dimensional information presentation	No interaction; linear viewing only	3-minute documentary-style video clip; no user participation required—viewing-only mode

Case B: Multimodal Interactive Drama	Multimodal integration (text, images, sound effects, video, etc.)	Moderate interaction; plot path selection available	Web-based interactive drama; users make plot choices at key nodes to influence subsequent content
Case C: AI-Driven Narrative Generation	Highly multimodal combination (visual, auditory, textual, dynamically generated content)	High-intensity interaction; user input dominates content generation	Generative AI-based story-building platform; users input or select keywords to generate personalized story texts and images in real time

During the experimental implementation, all three case types followed the standard procedure of 'narrative information input-interactive behaviour execution-cognitive feedback generation', ensuring users completed the full process of information reception, behavioural engagement, and cognitive processing throughout the experience. The experimental process and data for the paper were obtained with the consent of the participants, who signed informed consent forms, and received ethical approval from Xi'an Technological University. The design of these three case types was controlled and standardised in terms of content difficulty, narrative structure, and interactive tasks. It should be clarified that Case A was not intended to represent a fully developed multimodal interactive narrative. Instead, it served as a baseline condition within the experimental design. By adopting a low-modality, non-interactive narrative format, this study established a reference point for systematic comparison. This comparison evaluated how increased modality density and interaction frequency in Cases B and C affected cognition. This gradient-based design allowed the analysis to focus on how incremental changes in input structure were associated with variations in cognitive construction outcomes. This ensured comparability and measurability across different experimental conditions, based on the gradient changes in modal input and interaction frequency. Upon completing the experience, users were required to fulfil corresponding meaning feedback and text generation tasks as prompted by the system, thereby providing data support for measuring the Depth of Cognitive Construction (U) and analysing its mechanisms. The overall experimental process comprised four stages, detailed as follows:

Guidance and Familiarisation (5 minutes): Assist participants in understanding platform functionality and conduct foundational operational training to ensure consistent interaction behaviour; Immersive Narrative Experience (3×10 minutes): Participants sequentially complete full experiences of Case Types A/B/C, each lasting 10 minutes, with the platform automatically recording behavioural logs; Feedback Generation and Semantic Tasks (10 minutes): Following the experience, participants complete text-based tasks (content summarisation, narrative reconstruction, or viewpoint articulation) prompted by the system to extract cognitive construction depth metrics (U); Subjective Evaluation and Emotion Measurement (5 minutes): Participants complete scales for emotional resonance (C) and immersion experience (5-point Likert scale), while collecting emotional lexical density and co-occurring hot words to support composite measurement of the C metric. The measurement pathways and data sources for the four core variables are outlined in Table 2.

Table 2. Indicator Measurement and Data Collection Methods

Indicator	Operationalization Method	Data Source
Modal Density (M)	The system backend records the number of multi-modal calls per unit time × weight coefficient	Automatically generated by platform logs
Interaction Frequency (I)	Average number of user behaviors (clicks, inputs, jumps, etc.) per minute	Behavioral log tracking
Contextual Resonance (C)	Composite measurement: Likert scale scoring + emotional word density + semantic similarity calculated by embedding model	User questionnaires + platform text analysis
Depth of Cognitive Construction (U)	Three-level classification of user texts via semantic coding system (U1/U2/U3)	User text feedback + natural language processing

The three cases were not intended to provide a comprehensive representation of immersive media formats. Instead, they were deliberately selected to construct an experimental gradient along two core dimensions: modality density and interaction frequency. This design did not aim for platform-level generalizability. Rather, it sought to examine, under immersive narrative conditions, how controlled variations in multimodal and interactive input configurations were associated with cognitive construction outcomes.

Variable Definition

To construct and empirically validate a model of cognitive mechanisms in digital storytelling, this paper defines four core research variables based on existing theoretical frameworks and experimental conditions: modal density (M) [10], interaction frequency (I) [11], contextual resonance (C) [12], and Depth of Cognitive Construction (U) [13]. Operationalisation of these indicators and measurement pathway design were subsequently completed. The independent variables comprise modal density and interaction frequency, with contextual resonance (C) as a mediating variable and Depth of Cognitive Construction as the dependent variable. The operationalised measurement methods for each variable are as follows:

For modal density, this study employs systematic automated recording to quantify the frequency of multimodal invocations per unit time during user experiences. Weighting coefficients are assigned based on the differing information-carrying capacities of each modality (e.g., visual > auditory > textual), thereby constructing a composite score. The specific measurement method is expressed by the formula:

$$M=\sum_{i=1}^{n} \mu_i \cdot w_i \tag{1}$$

Where μ_i is the usage frequency of the i-th modality per unit time, and w_i is its weight coefficient. In this study, the weighting coefficient (w_i) was operationally defined as reflecting the relative information-carrying capacity and perceptual salience of different modalities within an immersive narrative environment. Based on prior research on multimodal communication and considerations of experimental controllability, the following standardized weights were adopted: visual modality (w = 0.5), auditory modality (w = 0.3), and textual modality (w = 0.2). These values were held constant across all experimental conditions to ensure internal consistency and comparability. This weighting scheme was not intended to represent an absolute hierarchy of human perception. Rather, it functioned as an operational approximation that balanced theoretical plausibility with measurement stability. Visual information was assigned the highest weight because it plays a dominant role in immersive perception and spatial cognition. Auditory and textual information were assigned lower weights to reflect their supportive yet distinct roles in narrative comprehension. Similar proportional weighting approaches have been adopted in prior multimodal modeling studies to enable the quantitative integration of heterogeneous input channels. Additional robustness checks using alternative weighting schemes yielded directionally consistent results, indicating that the observed relationships were not sensitive to minor variations in modality weighting.

Interaction frequency is quantified through platform logs, where the system automatically records user actions such as clicks, inputs, and selections during each case experience. This data is used to calculate the average number of interactions per unit time (minute), forming a quantitative metric. Contextual resonance employs a composite metric, integrating subjective ratings with objective textual features. Subjective scoring: Employing a five-point Likert scale, users are invited to rate their emotional resonance. Objective metrics: Calculating the density of emotional words within user feedback text, combined with sentiment vectors extracted via semantic embedding models, to compute similarity against standard sentiment templates. The final score is generated through weighted integration.

The Depth of Cognitive Construction is assessed through semantic encoding analysis, categorising user feedback texts into three levels of cognitive construction: information restatement (U1), meaning rewriting (U2), and value generation (U3). Large language models generate user text embedding vectors, which are then compared with standard hierarchical centre vectors via cosine similarity calculations. The formula is as follows:

$$U=\frac{N_{U3}+0.6\cdot N_{U2}}{N_{total}} \tag{2}$$

Where N_{U3} and N_{U2} are the numbers of text units at levels U3 and U2, respectively, and N_{total} is the total number of text units.

EXPERIMENT AND RESULT ANALYSIS

Model Path Configuration and Analysis Strategy Design

Model Deconstruction and Causal Pathway Establishment

Following the establishment of the variable system and the collection of experimental data, it is necessary to systematically deconstruct the causal pathway structure within the model to clarify the subsequent direction of solution and the logic for pathway validation. The variable relationship model established in this study exhibits a 'parallel input + serial mediation' structural characteristic. Modal density (M) and interaction frequency (I), as two parallel independent input variables, exert direct influence on the Depth of Cognitive Construction (U). Simultaneously, both form an indirect pathway through the mediating variable 'contextual resonance(C)', constituting a layered mediation mechanism with both direct and indirect effects. Consequently, the direct effect path 'M→U, I→U' characterises the immediate driving effect of information input intensity on user cognitive construction depth; the mediating effect path 'M→C→U, I→C→U' represents the transitional regulatory role of the contextual response mechanism between information input and cognitive construction depth.

Derivation of First-Order Regression Models and Parameter Calculation

To identify the direct driving effects of input variables on the Depth of Cognitive Construction in users, a first-order multiple linear regression model is constructed, as shown in Equation (3). This model analyses whether M and I exert a significant positive influence on U, while also examining the explanatory power of variable C. This provides a benchmark effect size for subsequent mediation effect modelling.

$$U=\alpha M+\beta I+\gamma C+\varepsilon \tag{3}$$

Where U denotes Depth of Cognitive Construction; α, β and γ are coefficients to be estimated; and ε is the residual term.

The model employs the least squares method for parameter estimation. Variables are first standardised to eliminate dimensional interference, with M, I, and C undergoing Z-score standardisation. Subsequently, multiple linear regression calculations are performed using SPSS to generate the coefficient matrix. This is followed by significance testing, where t-tests yield p-values for each variable, with a confidence level α = 0.05 set. Next, the model's coefficient of determination was calculated to assess its overall explanatory power. Finally, variance inflation factors (VIF) were analysed to diagnose multicollinearity among variables, ensuring VIF < 5 for validity.

Building upon the first-order regression analysis, this study employs structural equation modelling to model and test the mediating pathways, thereby further identifying the mechanism role of contextual resonance(C) within the 'input variable–cognitive construction depth' pathway. Two mediating pathways were constructed: modal density mediating via contextual resonance to cognitive construction depth 'M→C→U', and interaction frequency mediating via contextual resonance to cognitive construction depth 'I→C→U'. Both pathways share C as a common mediating node, forming a parallel nested structure of 'dual input-single mediator-single output'. This structure captures the mechanisms of affective transference and meaning processing during cognitive construction, thereby revealing the pathways of cognitive formation in immersive communication.

Modelling employed maximum likelihood estimation. First, a path diagram of variables was constructed. Fitting quality was assessed using root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI): RMSEA < 0.08; CFI > 0.90; TLI > 0.90. Finally, standardised regression weights and Bootstrap sampling (n = 2000, repeated self-sampling) were employed to conduct dual tests for the significance and confidence intervals of mediating effects, alongside testing the significance of path coefficients. Model estimation yielded the following path effect values: direct effects 'M→U, I→U'; Indirect effects: 'M→C→U, I→C→U'; Total effect: 'Direct effect + Indirect effect'.

Should the indirect path coefficient prove significant, and the direct path coefficient decrease upon introducing the mediating variable, the 'partial mediation' model is confirmed. This mechanism suggests that the perceived depth of contextual resonance and the degree of emotional triggering form a significant cognitive buffer between narrative input and user cognitive output.

Cognitive-Level Modelling of User Semantic Feedback

The Depth of Cognitive Construction in users (U), serving as the core dependent variable in this study, relies not only on numerical metrics but also on structural deconstruction and semantic hierarchical analysis of user-generated content. To this end, a text-semantic-based cognitive grading model must be established to achieve precise assessment of 'depth of cognitive construction'. Integrating the variable definitions outlined previously, this study categorises cognitive outputs into three processing levels, as presented in Table 3:

Table 3. Graded Framework for Cognitive Outputs

Level	Cognitive Type	Description of Processing Characteristics
U1	Information Retelling	Direct extraction or minor adjustment of text content; lack of structural reconstruction and meaning elevation
U2	Meaning Rewriting	Structural adjustment of event logic, character relationships, or plot sequence
U3	Value Generation	Development of new perspectives, interpretive frameworks, or extensible imagination based on the original narrative

To convert user text feedback into structured data, the following semantic modelling process is established: (1) Text cleansing and standardisation processing, removing stop words, punctuation, correcting misspellings, and unifying morphological transformations; (2) Keyword and syntactic unit extraction, employing dependency parsing algorithms to extract keyword-subject-predicate structure-sentiment word triplets; (3) Semantic vector embedding and hierarchical matching: Generate text embedding vectors using BERT, calculate cosine similarity between these vectors and three levels of typical semantic centres, and perform hierarchical classification; (4) Manual calibration and consistency verification: To ensure accuracy, employ double-coder cross-validation and calculate Cohen’s Kappa coefficient (setting Kappa ≥ 0.80 as valid).

Data Processing Workflow and Error Control Mechanism

(1) Variable Normalisation and Dimensional Standardisation

Multivariate modelling requires comparable scales for all input variables. In this study, the modal density (M), interaction frequency (I), and contextual resonance(C) exhibited significant differences in their original magnitudes. Without transformation, this would introduce parameter bias. Therefore, Z-score standardisation was applied, as shown in Equation (4):

$$X' = \frac{X-\bar{X}}{\sigma_X} \tag{4}$$

Here, X denotes the original variable, $\bar{X}$ represents the mean, and σ_X signifies the standard deviation. The standardised variables enter regression and SEM models with a mean of zero and a variance of one, ensuring the interpretability of path coefficients and the convergence stability of the models.

(2) Construction of the structural input matrix

All variables required for modelling are converted into a two-dimensional matrix form of user variables. Discrete variables are transformed into continuous variables through coding, constituting the modelling input matrix as per formula (5):

$$\mathbf{X} = \begin{bmatrix} \mathrm{M}_1 & \mathrm{I}_1 & \mathrm{C}_1 & \mathrm{U}_1 \\ \mathrm{M}_2 & \mathrm{I}_2 & \mathrm{C}_2 & \mathrm{U}_2 \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{M}_n & \mathrm{I}_n & \mathrm{C}_n & \mathrm{U}_n \end{bmatrix} \tag{5}$$

This structure ensures compatibility between the input format and the tensor structure required by regression functions and path models, thereby enhancing computational efficiency and error control.

(3) Model Hierarchy Error Control Mechanism

To mitigate potential statistical bias and structural distortions during modelling, a multi-layered error control mechanism is established by incorporating collinearity diagnostics, path residual tests, parameter robustness verification, and textual consistency analysis.

(4) Multiple regression model analysis

A standardised multiple linear regression model was employed to model the three independent variables: Modal density (M), Interaction frequency (I), and Contextual resonance (C). The results are presented in Figure 1.

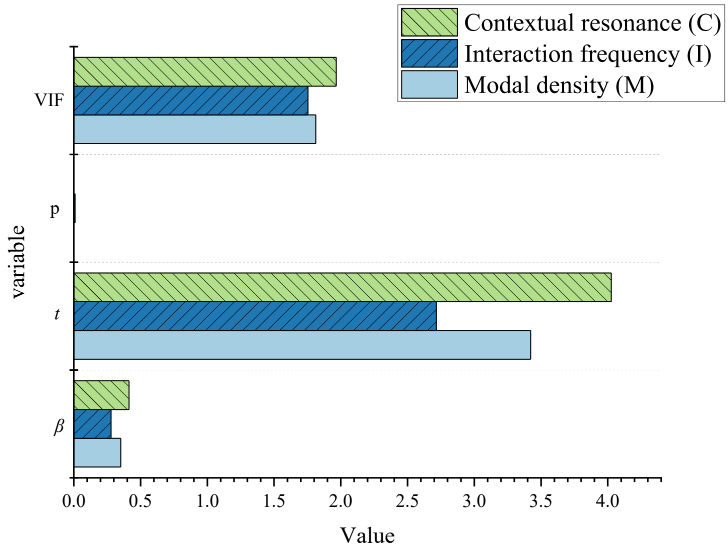

Figure 1. Regression Coefficient Estimates

To comprehensively evaluate the regression model's fit for modal density, interaction frequency, and contextual resonance, key statistical summaries-the coefficient of determination (R2), F-test value, and multicollinearity diagnostic parameters—were analysed. Model summary: R2 = 0.487, adjusted R2 = 0.462; F (3,36) = 11.391, p < 0.001. In Figure 1, both modal density (β = 0.351, p = 0.001) and interaction frequency (β = 0.279, p = 0.009) exerted significant positive effects on Depth of Cognitive Construction, while contextual resonance (β = 0.414, p < 0.001) exerted significant positive effects on cognitive construction depth magnitude, indicating that all three input variables possess independent cognitive facilitation effects within immersive narrative environments. Notably, the standardised coefficient for contextual resonance was the highest, signifying this variable's strongest predictive effect on cognitive construction depths. This suggests that users' emotional engagement and depth of resonance occupy a pivotal position in cognitive construction. This finding provides a causal hypothesis foundation for subsequent mediation effect modelling. The model demonstrated satisfactory overall explanatory power (adjusted R2 = 0.462), meeting the general expectation for multivariate modelling in social behaviour research (> 0.3). The F-test yielded significant results, rejecting the null hypothesis and confirming that at least one independent variable possesses statistically significant predictive power for the dependent variable. Variance inflation factors remained below 2, ruling out multicollinearity interference. This indicates sound structural differentiation among the three input variables and demonstrates the mathematical stability of the regression model estimates.

(5) Analysis of the Mediating Role of Situational Resonance

To examine the mediating effect of contextual resonance (C) on the pathway from modal density (M) and interaction frequency (I) to Depth of Cognitive Construction (U), a structural equation model featuring dual inputs and single mediation was constructed. Model paths are represented by standardised regression weights, with all manifest variables being Z-standardised measured indicators. The estimated structural equation path coefficients are presented in Figure 2:

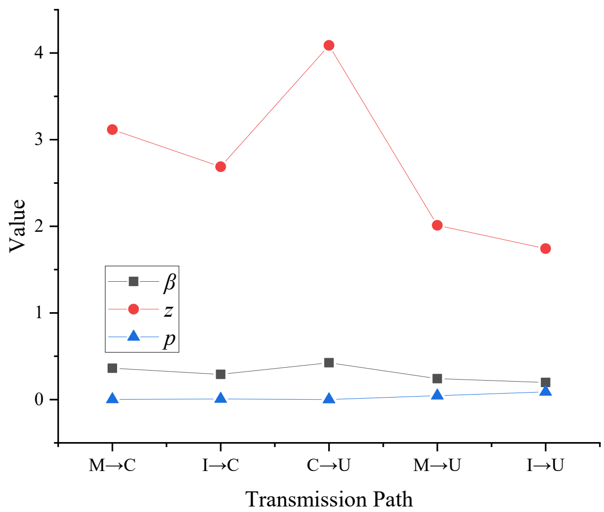

Figure 2. Path Coefficient Estimates for Structural Equation Model

The model fit indices are as follows: RMSEA = 0.051, CFI = 0.946, TLI = 0.932, $\chi^2/df = 1.812$, meeting the fit criteria. As shown in Figure 2, the structural equation model exhibits overall good fit (RMSEA<0.06, CFI>0.90), supporting the theoretical validity of the path settings between variables. Path coefficient analysis revealed: Modal Density (M) exerted a significant positive influence on Contextual resonance (C) ($\beta = 0.362$, $p = 0.002$), indicating that higher modal complexity facilitates greater emotional engagement and situational immersion among users; Interaction Frequency (I) also significantly positively influenced Contextual resonance ($\beta = 0.291$, $p = 0.007$); Scenario resonance significantly influenced Depth of Cognitive Construction (U) ($\beta = 0.426$, $p < 0.001$), validating the crucial mediating role of emotional perception in cognitive construction.

To further validate the statistical significance of the mediating effect, Bootstrap resampling (n = 2000) was employed to generate confidence intervals for the indirect path. Results are presented in Table 4:

Table 4. Results of Mediating Effect Identification

Mediating Path	Indirect Effect Value	95% CI	Significant Mediation?
M→C→U	0.154	[0.082, 0.244]	Yes
I→C→U	0.124	[0.049, 0.216]	Yes

Table 4 indicates that the confidence intervals for both mediating paths do not encompass zero, confirming that contextual resonance exerts a partial mediating effect within these pathways. Specifically, the M→U path remained significant after controlling for the mediating variable ($\beta = 0.241$, $p = 0.045$), whereas the I→U path lost significance upon introducing C ($p = 0.089$). This indicates that the effect of interaction frequency is almost entirely mediated through the context.

Outcome-Based Classification of Cognitive Construction Levels and Hierarchical Distribution

Following the numerical estimation of the cognitive construction depth variable U, to gain deeper insight into its performance mechanisms and construction types within the user group, the study conducted a clustering analysis of all participants' cognitive outputs based on the hierarchical labels (U1, U2, U3) derived from semantic feedback. This was further correlated with input variables to explore differential patterns in construction types. The hierarchical distribution of user cognitive constructions is presented in Table 5:

Table 5. Distribution of Cognitive Construction Levels in User Outputs (n = 40)

Hierarchical Category	Meaning Description	Average Proportion (%)	Standard Deviation (SD)
U1	Information Retelling Level	41.70	9.60
U2	Meaning Rewriting Level	34.20	8.10
U3	Value Generation Level	24.10	7.30

As shown in Table 5, the results indicate that U1 (replicative cognition) accounted for the highest proportion, reflecting that most users primarily accepted the original narrative content. However, the combined proportion of U2 and U3 approached 60%, suggesting that in an immersive context, over half of the users possessed a certain capacity for structural processing or value generation. The overall distribution indicated that cognitive construction in immersive narrative contexts was not merely a process of information reproduction. A substantial proportion of user outputs reflected processes of structural reinterpretation and value-oriented generation. The results of the three-category clustering of user cognitive patterns are presented in Table 6:

Table 6. Hierarchical Distribution of User Perception Constructs (K-means, k = 3)

Cluster Group	Dominant Cognitive Type	Number of Samples	Average U Value
Reproduction-oriented users	U1-dominant	15	0.22
Reconstruction-oriented users	U2-dominant	14	0.38
Generation-oriented users	U3-dominant	11	0.54

Based on Table 6, the K-means clustering method ($k = 3$) was employed to classify the cognitive type distribution within participants' feedback texts. This yielded three distinct patterns, each corresponding to the predominant tendencies within the aforementioned cognitive hierarchy. The clustering contour coefficient of 0.611 indicates that the group structure exhibits moderate discriminative power, meeting the criteria for valid clustering. Further analysis employed one-way analysis of variance to examine differences in input variables across the three cognitive construction groups, as presented in Table 7.

Table 7. Results of Group Difference Analysis

Input Variable	F Value	p Value	Difference Conclusion
Modal Density (M)	6.284	0.004	Significant difference ($p < 0.01$)
Interaction Frequency (I)	4.912	0.012	Significant difference ($p < 0.05$)

Table 7 shows statistically significant differences in modality density and interaction frequency across groups with different cognitive construction outcomes. It should be emphasized that these groups were defined post hoc based on the observed cognitive construction depth (U). They did not represent independent or inherent user attributes. Accordingly, the observed differences reflected a structural association between the input conditions of the experimental configuration, namely modality density and interaction frequency, and cognitive construction outcomes. They did not indicate that user types themselves determined variations in modality density or interaction frequency. Comparative results for mean values across input variables for each cognitive construction group are presented in Table 8:

Table 8. Comparison of Mean Values Across Cognitive Construct Groups for Input Variables

Group Type	Mean of Modal Density (M)	Mean of Interaction Frequency (I) (times/minute)
Reproduction-oriented users	3.12	8.21
Reconstruction-oriented users	3.89	10.03
Generation-oriented users	4.58	12.14

The Tukey HSD post-hoc test results are presented in Table 9:

Table 9. Tukey HSD Post-hoc Test Results

Comparison Group	p Value of Modal Density (M)	p Value of Interaction Frequency (I)	Description of Significant Difference
Reproduction-oriented users vs. Reconstruction-oriented users	0.047	0.038	Significant difference exists
Reproduction-oriented users vs. Generation-oriented users	0.003	0.001	Stronger significant difference
Reconstruction-oriented users vs. Generation-oriented users	0.084	0.076	

Synthesizing the results of Tables 8 and 9, higher levels of cognitive construction outcomes under the experimental conditions were systematically associated with configurations featuring greater modality density and higher interaction frequency ($p < 0.01$). The observed outcome-based hierarchical pattern suggested a progressive relationship between information input intensity and cognitive construction depth. This pattern was not attributable to differences in users' inherent characteristics. These findings supported the interpretation that variations in multimodal and interactive input structures were closely related to the depth of cognitive construction achieved in immersive narrative environments.

DISCUSSION

The structural equation modeling results revealed clear path heterogeneity. The effect of interaction frequency (I) was primarily mediated through contextual resonance (C), whereas modality density (M) retained a direct effect while also operating through C. This distinction indicated that the functional roles of the two input variables in immersive narratives were asymmetric. The underlying reason was that interaction behavior constituted a signal of agentive intervention. User actions such as clicking, selecting, and making path decisions first altered the sense of involvement and control. This process subsequently enhanced self-relevance appraisal and emotional arousal. Once emotion and immersion were stably activated, cognitive processing shifted toward deeper organization and integration. As a result, the influence of I on cognitive construction was mainly transmitted through contextual resonance. By contrast, modality density provided informational and structural resources. The juxtaposition of multiple modalities not only enhanced sensory immersion but also created semantic anchors and memory indices through the accumulation of multi-channel cues. This process encouraged users to establish correspondences across symbolic forms and to complete cross-modal integration. It directly supported meaning reorganization and frame generation. Consequently, M maintained an independent contribution to cognitive construction even after emotional resonance was controlled. In other words, I functioned more as a trigger for emotion and attention, whereas M operated as scaffolding for cognition and semantics. Together, they formed a coupled mechanism linking an emotional pathway with a structural pathway. This interpretation suggested that immersive narrative design should avoid equating interaction enhancement with cognitive enhancement. A more effective strategy was to use interaction to trigger resonance and to use modality to organize meaning. Such an approach could support a sustainable chain of deep cognitive construction. At the application level, the proposed mechanism model linking modality density, interaction frequency, contextual resonance, and cognitive construction offered implications for intelligent textiles, virtual fashion systems, and human–computer interaction design. In intelligent textile and wearable media contexts, multimodal information should not be treated as a simple accumulation of sensory stimuli. It should instead be configured in a structured manner across visual textures, tactile feedback, and dynamic information. This configuration can form stable semantic cues and experiential anchors that support users' deep understanding of product functionality and cultural meaning. In virtual apparel and digital fashion presentation, the value of interaction design lay primarily in contextual triggering. Moderately paced and well-calibrated interactive choices could strengthen emotional engagement and identity immersion. This process could enhance users' cognitive acceptance of virtual garment styles, materials, and narrative concepts. More broadly, human–computer interaction systems should not rely solely on interaction frequency or interface complexity as performance indicators. They should attend to the differentiated cognitive roles of distinct input pathways. Interaction should be used to activate contextual resonance, while multimodal organization should be used to support meaning construction. The proposed mechanism provided an interpretable, cognition-oriented framework for immersive product and interface design. It facilitated a shift in user experience optimization from sensory enhancement to support for cognitive construction. This shift may enhance communication effectiveness and user retention in practical applications of intelligent textile and virtual fashion systems.

CONCLUSION

This study constructs a verifiable mechanism pathway model based on four variables: modal density, interaction frequency, contextual resonance, and cognitive construction depth. Through empirical experiments and modelling analysis, it validates the model's intrinsic structure and direction of effects. The analysis identifies contextual resonance as mediating the relationship between M and U (indirect effect = 0.154, CI = [0.082, 0.244]) and between I and U (indirect effect = 0.124, CI = [0.049, 0.216]). CI = [0.082,0.244]) and between I and U (indirect effect = 0.124, CI = [0.049,0.216]). The findings of this study have important implications for the textile and apparel fields. For example, in the development of multimodal smart textiles (integrating visual and tactile feedback) or virtual try-on systems, designers should strive to enhance multimodal density (M) and interaction frequency (I) in order to effectively drive deeper levels of users' Depth of Cognitive Construction (U). Specifically, compared with experimental conditions characterized by lower multimodal density and interaction frequency, conditions with higher multimodal density and interaction frequency were associated with a higher mean cognitive construction depth (U = 0.54), whereas the former yielded a mean value of U = 0.22. This finding further indicated that optimizing multimodal presentation and interaction design in immersive experiences could significantly enhance users' cognitive engagement and value-oriented interpretations of textile-related digital narratives. The findings of this study should be interpreted within appropriate boundary conditions. Because the empirical analysis was based on three experimentally constructed narrative scenarios, the proposed MICU mechanism was intended to explain the structural relationship between input configurations and cognitive construction outcomes. It was not intended to be generalized to all forms of immersive media or user populations. Future research may extend this framework to other media environments and longitudinal research contexts in order to further evaluate its applicability and robustness.

Author Contributions

Conceptualization – R.H. and D.X.; Methodology – R.H. and D.X.; writing-original draft preparation – R.H.; writing-review and editing – R.H. and D.X.; Validation – R.H. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research is supported financially by Project of The Social Science Foundation of Shaanxi (Program No. 2025J002): "Research on Digital Translation and Narrative Structure Innovation of Shaanxi Cultural Symbols in the Context of Digital Intelligence Communication".

Human Research Subjects

The experimental process and data for the paper were obtained with the consent of the participants, who signed informed consent forms, and received ethical approval from Xi'an Jiaotong University.

Acknowledgements

Not applicable

REFERENCES

[1] Bruni LE, Kadastik N, Pedersen TA, Kerr C, Phaal R. Digital narratives in extended realities. Roadmapping extended reality Roadmapping Extended Reality: Fundamentals and Applications. Hoboken, NJ: Wiley; 2022.p. 35-62. doi: 10.1002/9781119865810.ch2

[2] Cao S, Cao N. How Does Emotion Affect Information Communication. ArXiv; preprint arXiv:2502.16038. 2025. doi: 10.48550/arXiv.2502.16038

[3] Campbell M, Vidal R. The multimodal translation workshop as a method of creative inquiry: Acousmatic sound, affective perception, and experiential literacy. Target. 2024; 36(2):184-214. doi: 10.1075/target.22051.cam

[4] Ding J. Adopting the first-person perspective theory to safeguarding heritage craftsmanship in immersive virtual narrative design [dissertation]. Hong Kong, China: The Hong Kong Polytechnic University; 2024. 319 p. Available from: https://theses.lib.polyu.edu.hk/handle/200/13079

[5] Kourtesis P. A Comprehensive Review of Multimodal XR Applications, Risks, and Ethical Challenges in the Metaverse. Multimodal Technologies and Interaction. 2024; 8(11):98. doi: 10.3390/mti8110098

[6] Chen X, Ibrahim Z. A comprehensive study of emotional responses in ai-enhanced interactive installation art. Sustainability. 2023; 15(22):15830. doi: 10.3390/su152215830

[7] Norrie CS, Deckers SRJM, Radstaake M. A narrative review of the sociotechnical landscape and potential of computer-assisted dynamic assessment for children with communication support needs. Multimodal Technologies and Interaction. 2024; 8(5):38. doi: 10.3390/mti8050038

[8] Díaz EP, Ricaurte P. Doctors for the truth: Latin American antivaccine oppositional cultures on Telegram. Big Data & Society. 2024; 11(4):20539517241306359. doi: 10.1177/20539517241306359

[9] Moinnereau MA, de Oliveira Jr AA, Falk TH. Immersive media experience: a survey of existing methods and tools for human influential factors assessment. Quality and User Experience. 2022; 7(1):5. doi: 10.1007/s41233-022-00052-1

[10] Darbandi M, Ghafourian E. Statistical Evaluation of Multimodal Interfaces: Exploring User Preferences for Combined Input Methods. International Journal of Advanced Human Computer Interaction. 2022; 1(1):17-30. Available from: https://www.ijahci.com/index.php/ijahci/article/view/14

[11] Talukdar N, Yu S. Breaking the psychological distance: the effect of immersive virtual reality on perceived novelty and user satisfaction. Journal of Strategic Marketing. 2024; 32(8):1147-1171. doi: 10.1080/0965254X.2021.1967428

[12] Han DID, Bergs Y, Moorhouse N. Virtual reality consumer experience escapes: preparing for the metaverse. Virtual Reality. 2022; 26(4):1443-1458. doi: 10.1007/s10055-022-00641-7

[13] Schneider S, Beege M, Nebel S, Ziaian JT, Gillham D. The cognitive-affective-social theory of learning in digital environments (CASTLE). Educational Psychology Review. 2022; 34(1):1-38. doi: 10.1007/s10648-021-09626-5