A Method for Generating Immersive Narratives in Digital Shadow Puppetry Based on ReAct Planning and Multi-Agent Coordination

Enze Li[1], Xiaowan Xie[2,*]
[1]Animation Department, Huanghuai University, Zhumadian Henan 463000, China
[2]Art and Design Department, Huanghuai University, Zhumadian Henan 463000, China
Corresponding author: Xiaowan Xie (e-mail: 20192068@huanghuai.edu.cn).

ABSTRACT Technological developments in the field of artificial intelligence and intelligent algorithms are the foundation of digital shadow puppetry to move towards a dynamic, interactive and multimodal storytelling. This paper aims to solve such problems as inadequate plot continuity, lack of constraints on cultural rules and conflicts between multi-agent outputs, and proposes a method for creating an immersive narrative in digital shadow puppetry based on ReAct planning and multi-agent collaboration. This approach builds an association model between narrative elements and cultural characteristics of digital shadow puppetry, and then it uses the 'reasoning–action–observation' cycle to do the narrative goal decomposition, cultural constraint path search and local dynamic re-planning, and finally it achieves the collaborative generation of multimodal content by modelling the director, screenwriter, characters, scenes, actions and music as agents. Plot coherence, cultural conformity and multimodal alignment were achieved at 92.4 per cent, 94.1 per cent and 92.8 per cent respectively while task completion was at 94.5 per cent and overall user satisfaction score of 4.63 in experimental results. The effective combination of cultural expression norms, narrative consistency and real-time interaction requirements has ensured this method can be a viable approach for the diffusion and interactive performance of digital shadow puppetry.

INDEX TERMS ReAct planning; Multi-agent collaboration; Digital shadow puppetry; Immersive storytelling; Multimodal generation

I. INTRODUCTION

The development of digital storytelling, from the playback of a coherent script to a perceptible, rational and collaborative production, has opened new possibilities for the living transmission and immersive dissemination of digital shadow puppetry, thanks to the progress of artificial intelligence and intelligent algorithms. Gao and Juluri proposed a planner–executor–evaluator system to improve the quality of complex content generation with iterative feedback [1] and Marah and Challenger used adaptive hybrid reasoning in agent digital twins to improve the ability to update the state in dynamic environments [2]. Previously, Younes et al. [3] used multi-agent interaction to create coordinated actions for physical characters a priori, Zhang et al. [4] integrated multi-agent system with the connection of narrative structure and playable scenarios, Wu et al. [5] analysed information acquisition, character interaction and logical decision-making in reasoning games, and Tütüncü et al. [6] manipulated characters and props in extended reality to create structured narrative segments. Ahmed et al. presented an overview of the collaboration and scaling mechanisms of distributed multi-agent systems [7] while Uddin et al. pointed out that communication failure, memory corruption and behavioral inconsistencies can lead to loss of confidence in real time systems [8]; Wu et al quantified collaborative structures from causal relationship and sub team perspectives [9]. In general, state-of-the-art approaches are still unable to meet stylised cultural constraints, the need to re-plan a plot based on the user's feedback, and the need to keep text, actions, scenes and music consistent.

Therefore, an immersive approach in storytelling has been designed which combines ReAct planning, cultural constraint search, director agent scheduling, shared memory and conflict resolution. The research encompasses task modelling, dynamic programming, multimodal generation and interactive evaluation to improve the coherence of the narrative, cultural appropriateness, operational stability and user immersion.

II. TASK MODELLING FOR IMMERSIVE NARRATIVE GENERATION IN DIGITAL SHADOW PUPPETRY

A. REPRESENTATION OF NARRATIVE ELEMENTS AND CULTURAL CHARACTERISTICS IN DIGITAL SHADOW PUPPETRY

Digital shadow puppetry narrative elements include characters, plot, scenes, actions, music and props as well as repertoire conventions and stylistic norms, vocal styles and ethical rules [10]. Various elements are mapped to different kinds of heterogeneous nodes and a semantic association network is built based on participation and trigger relationship, dependency and constraint relationship. The combination of text, visual and audio features creates a single representation with an associative structure as illustrated in the Figure 1.

B. ANALYSIS OF IMMERSIVE NARRATIVE SCENARIOS AND INTERACTION REQUIREMENTS

Immersive digital shadow puppet storytelling is designed for virtual theatres, digital exhibition spaces and mobile interactive scenarios, and must comprehensively respond to users' voice, gestures, gaze and plot choices [11]. Interaction requirements can be expressed as follows:

$$D_t = \{I_t, E_t, P_t, R_t\} \tag{1}$$

where I_t represents user intent, E_t represents the environmental state, P_t represents the current plot, and R_t represents real-time feedback [12]. The system dynamically adjusts character behaviour, scene transitions and narrative pacing in response to changing requirements to ensure interactive coherence, content controllability and an immersive experience.

C. STRUCTURED ORGANISATION OF MULTIMODAL NARRATIVE RESOURCES

Multimodal narrative elements for digital shadow puppetry are: script dialogue, characters designs, scene imagery, sequence of actions, vocal melodies and music, and ambient sounds. These resources are uniformly encoded in character, phase of plot, emotion, timestamp and category of culture and relationships of cross-modal indexing are set up [13]. Semantic tags and vector features are used to achieve the resource retrieval, association matching and temporal alignment, which provides a standardised data foundation for the following plot planning, agent invocation and multimodal collaborative generation.

D. DEFINITION OF STATES, ACTIONS AND CONSTRAINTS FOR NARRATIVE GENERATION TASKS

Digital shadow puppet narrative generation can be modelled as a constrained sequential decision-making process, with its states and actions defined as follows:

$$S_t = (u_t, p_t, e_t, m_t) \tag{2}$$

$$a_t = \arg\max_{a \in A} Q(s_t, a), \quad C(s_t, a_t) \le \varepsilon \tag{3}$$

where u_t represents user intent, p_t represents plot state, e_t represents scene environment, m_t represents shared memory; A represents the set of candidate actions, Q represents the action utility function, C represents the cost of cultural and logical constraints, and ε represents the permitted threshold [14]. This definition provides a unified basis for plot planning, content generation and dynamic replanning.

E. OVERALL SYSTEM ARCHITECTURE DESIGN

The whole system architecture includes an interaction perception layer, a resource management layer, a ReAct planning layer, a multi-agent generation layer, a shared memory layer and an immersive presentation layer [15]. After the intent is recognized and states are updated, the ReAct planner takes the user input and breaks it down into objectives and chooses the paths of the narrative, which is then collaboratively created by the scriptwriter, character, scene, action and music agents, with consistency being checked by the director agent using shared memory. The closed-loop dynamic replanning system is continuously fed back to the planning module with the environmental feedback. The overall structure is shown in Figure 2.

III. REACT-BASED IMMERSIVE NARRATIVE PLANNING METHOD

A. CONSTRUCTION OF THE REACT NARRATIVE PLANNING MODEL

The ReAct narrative planning model takes user intent, the current plot, shared memory and cultural knowledge as joint inputs, and progressively generates narrative paths through a 'reasoning–action–observation' cycle [16]. The planning reasoning state is represented as:

$$z_t = \phi_\theta([s_t, g_t, m_t, k_t]) \tag{4}$$

The action selection probability is defined as:

$$P(a_t \mid z_t) = \frac{\exp(W_{a_t}^T z_t)}{\sum_{a_i \in A_t} \exp(W_{a_i}^T z_t)} \tag{5}$$

where s_t is the current narrative state, g_t is the stage goal, m_t is the shared memory, k_t is the cultural knowledge, ϕ_θ is the inference encoding function, A_t is the set of candidate actions, and W_a is the action weight [17]. The results of action execution are fed back via environmental observations to update the next round of planning; the structure is shown in Figure 3.

B. USER INTENT RECOGNITION AND NARRATIVE GOAL DECOMPOSITION

User intent recognition integrates interactive information such as text commands, speech semantics, gesture categories and plot selections, and employs an attention mechanism to calculate the contribution weights of different modalities:

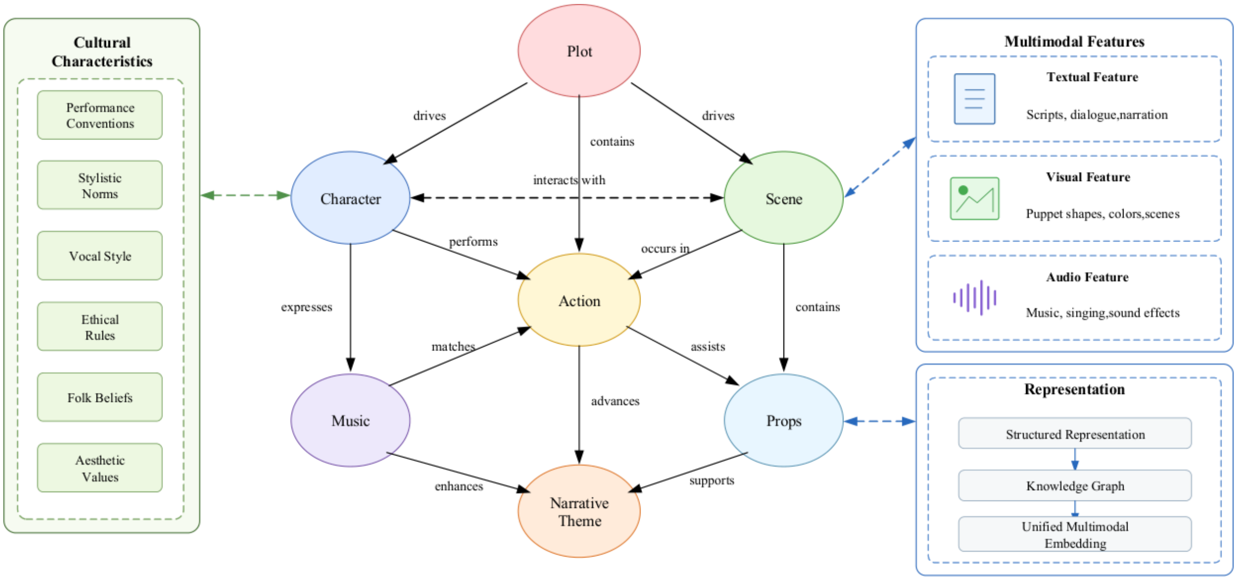

Fig. 1. Association Model of Narrative Elements in Digital Shadow Puppetry.

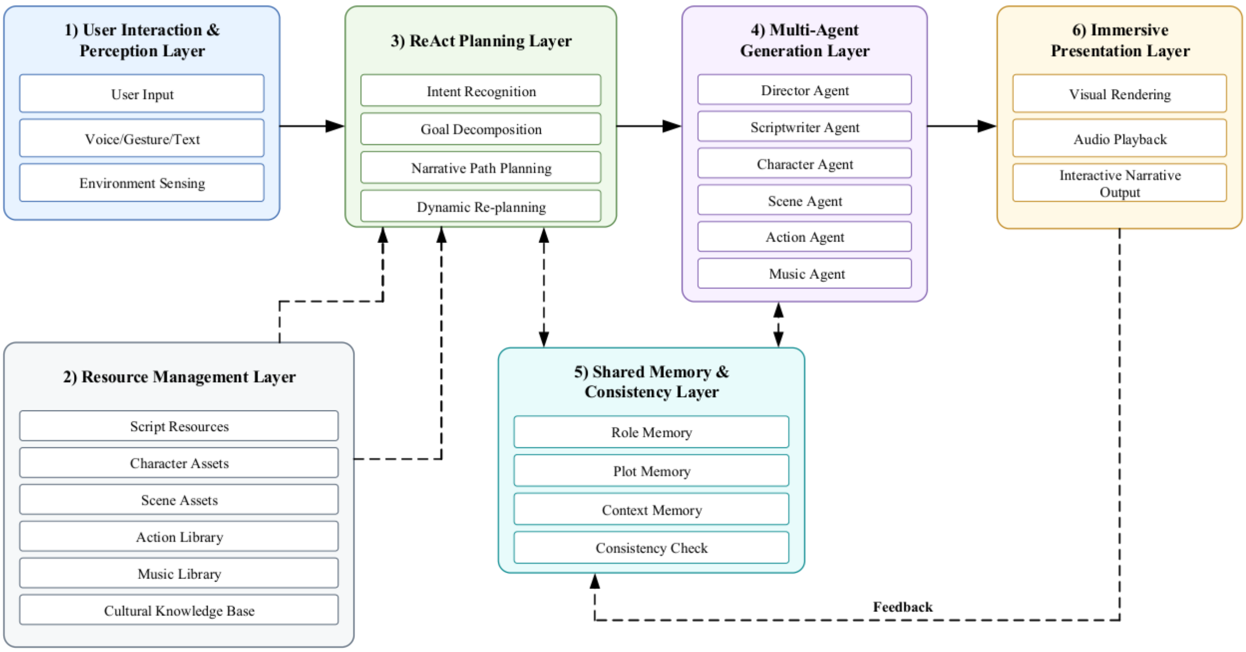

Fig. 2. Overall System Architecture.

$$h_u = \sum_{j=1}^{M} \alpha_j h_j \tag{6}$$

$$\alpha_j = \frac{\exp(q^T h_j)}{\sum_{k=1}^{M} \exp(q^T h_k)} \tag{7}$$

Here, h_j represents the interaction modality features of class j, M denotes the number of modalities, q is the intent query vector, α_j represents the modality weights, and h_u is the fused user intent representation [18].

The system recognizes eight interaction events as recognition windows, and intents are classified into six classes, namely, Plot Advancement, Character Interaction, Scene Exploration, Information Retrieval, Branch Modification and Narrative Termination with a recognition confidence threshold of 0.70. Later, according to character, event, scene and temporal relations, the detailed narrative goal is broken down into at most six sub-goals and finally a task dependency graph with a depth of at most 4 is built to give structure to the subsequent planning and agent scheduling [19].

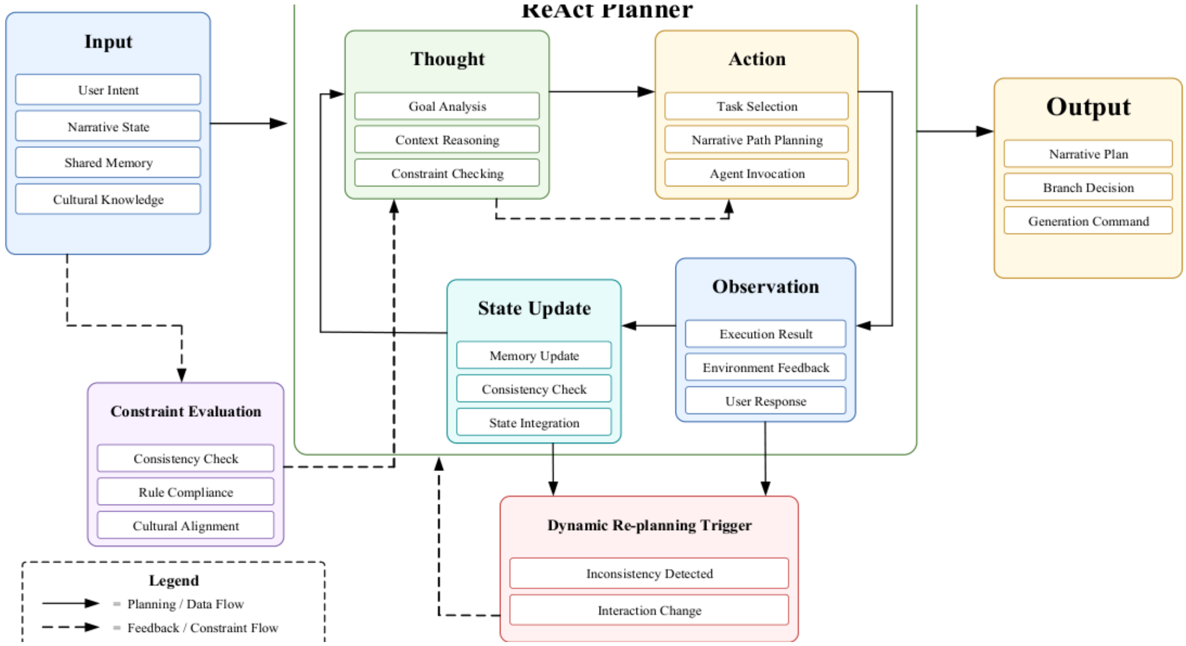

Fig. 3. The ReAct narrative planning model.

C. PLOT PLANNING MECHANISM BASED ON THE ALTERNATION OF THOUGHT AND ACTION

Plot planning adopts a local, incremental generation approach, organising goal analysis, cultural retrieval, candidate construction and action execution in an alternating sequence to avoid plot drift caused by generating a complete story in a single step [20]. The comprehensive score for candidate actions in the tth round is defined as:

$$V(a_i) = \lambda_1 G_i + \lambda_2 L_i + \lambda_3 R_i - \lambda_4 U_i \tag{8}$$

where G_i represents the match between the candidate action and the stage objective, L_i represents the logical coherence of the plot, R_i represents the coverage of available text, action and music resources, U_i represents the uncertainty of execution, and λ_i represents the weighting coefficient [21], and their sum is 1.

For each round, the top five actions are kept; the max planning depth is 8; and each action corresponds to 2–4 narrative sentences and changes in character states [22]. Once an action is completed, the results of the action, user responses and resource invocation statuses are logged in the observation log, and then the next round of inference is carried out, allowing the narrative path to grow progressively throughout the interaction process.

D. DYNAMIC RE-PLANNING BASED ON ENVIRONMENTAL FEEDBACK

Environmental feedback comprises changes in user behaviour, character execution status, scene resource availability and consistency check results. To determine whether the current plan requires adjustment, a comprehensive deviation metric is constructed:

$$\rho_t = \beta_1 \|e_t - \hat{e}_t\|_2 + \beta_2(1 - c_t) + \beta_2 r_t \tag{9}$$

where e_t represents the actual environmental state, $\hat{e}_t$ represents the predicted state during the planning phase, c_t represents the current plot consistency score, r_t represents the conflict rate of resources or cultural rules, and β_i represents the normalised weight [23].

Dynamic re-planning is triggered when ρ_t exceeds 0.45. The system's focus is on back tracking the last 2–3 plot nodes affected, with user choices confirmed and content completed, while generating only the tasks of the characters involved, scene transitions and the resulting branches. No more than three times of repair is carried out for every task, and once local repair is still not able to meet the constraints, the director agent will try to invoke a backup narrative path, which will control the range of re-planning and the extra computational work [24].

E. NARRATIVE PATH SEARCH AWARE OF CULTURAL CONSTRAINTS

Narrative path search incorporates the identities of shadow puppet characters, stylistic conventions, ethical relationships, vocal styles and scene taboos into the candidate evaluation process, and classifies cultural rules into hard and soft constraints [25].

The search cost for a candidate path P is defined as:

$$J(P) = \sum_{t=1}^{T} \left[\eta_1 d_t^g + \eta_2 d_t^l + \eta_3 d_t^c\right] + \kappa N_h(P) \tag{10}$$

where d_t^g represents the deviation between the tth node and the narrative goal, d_t^l denotes the logical divergence between adjacent plot segments, d_t^c represents the penalty for soft cultural constraints, $N_h(P)$ indicates the number of hard constraint conflicts, η_i denotes the evaluation weight, and κ is the large penalty coefficient [26].

The search is based on the constraint bundle search technology, and the results are immediately eliminated when the character identity is inconsistent; when it is inconsistent with traditional ethics, or a forbidden combination of styles is found; when the vocal style is inconsistent, the penalty coefficient is 0.1-0.4, and when the colour style is inconsistent, the penalty coefficient is 0.2-0.8, which is a balance between extensibility of the story and cultural expression norms and the width of the bundle is set to 6, and the maximum depth of the branch is set to 10.

F. REACT NARRATIVE PLANNING ALGORITHM WORKFLOW

The ReAct narrative planning algorithm inputs user's input, the current state of the plot and the state of the culture knowledge, and then iteratively performs intention recognition, goal decomposition, candidate path generation, constraint checking and action selection. After action execution, the system gathers the user's response, the changes in the environment and the generation result of the multi agents, and refreshes the shared memory & the state of the narrative; If the stability value is below the threshold or the change rate of interaction exceeds the set range, it will trigger the re-planning of the local area; If not, continue to narrate the plot until the narrative goal is reached [27]. The entire algorithm operation process is displayed in Figure 4.

IV. MULTI-AGENT COLLABORATIVE DIGITAL SHADOW PUPPET CONTENT GENERATION

A. MULTI-AGENT ROLE ALLOCATION AND COMMUNICATION MECHANISMS

In the multi-agent system we have agents representing the director, the screenwriter, the characters, scenes, actions, the music and the cultural review, the last one being the director agent, which has the responsibility of decomposing the tasks and coordinating the execution of them. The agents use a regular message format that includes task ID, context, constraints, execution status/execution output results, and share important elements of the plot and character status, thus creating a traceable collaborative generation chain. Each agent in the system has its role defined and functional configuration as given in Table 1.

TABLE 1. Multi-Agent Roles and Functional Configuration

Agent	Main Function	Output
Director Agent	Decomposes tasks and coordinates agents	Scheduling commands
Scriptwriter Agent	Generates plots, narration and dialogue	Narrative scripts
Character Agent	Controls character speech, emotion and behavior	Character responses
Scene Agent	Configures backgrounds, lighting and props	Scene descriptions
Action Agent	Generates puppet movements and action sequences	Action instructions
Music Agent	Matches singing, music and sound effects	Audio sequences
Cultural Review Agent	Checks cultural rules and narrative conventions	Review results

B. TASK SCHEDULING DRIVEN BY THE DIRECTOR AGENT

The task dependency graph output of the ReAct planner is given to the Director Agent who then generates a scheduling sequence according to the priorities of the sub-tasks, precedence/succession relationships, cultural constraint levels, and the agent's workload. The plot and character tasks are performed sequentially based on the dependency of the tasks and the scene, action and music tasks are generated sequentially when conditions are satisfied. The Director Agent constantly tracks the task statuses and retries, replaces or reallocates tasks in case of time-outs, conflict or low-confidence results, and writes the reviewed content to shared memory to ensure that the multimodal generation process is carried out in an orderly fashion and that the overall content is consistent.

C. COLLABORATIVE GENERATION BY THE SCREENWRITER AND CHARACTER AGENTS

The screenwriter agent generates event frameworks, draft dialogue and plot turning points based on narrative objectives, plot context and cultural constraints, whilst the character agent generates personalised responses by combining identity attributes, emotional states and historical behaviour. Character state updates are represented as:

$$r_i^{t+1} = F_r(r_i^t, e_t, g_i, m_t) \tag{11}$$

where r_i^t is the state of character i at time t, e_t is the current event, g_i is the character's goal, and m_t is the shared memory. The consistency score of the generated content is defined as:

$$C_i = \alpha C_i^p + \beta C_i^c + \gamma C_i^h \tag{12}$$

where C_i^p, C_i^c and C_i^h represent plot match, character consistency and historical consistency, respectively, and $\alpha + \beta + \gamma = 1$. Responses falling below the set threshold are returned to the screenwriting agent for revision, thereby forming a collaborative closed-loop system comprising plot generation, character feedback and consistency review.

D. MULTIMODAL COORDINATION OF SCENES, ACTIONS AND MUSIC

It is a multimodal content generation by the scene, action and music agents working in parallel, with a common plot timeline and action sequence. The background, lighting and props are set by the scene agent, combining location, mood and cultural tags; the action agent generates the trajectories of the shadows puppet's joints and movement rhythms based on the intent of the characters; and the music agent adapts the music to the intensity of the plot, including: vocal styles, percussion and ambient sound effects. Semantic and temporal alignment is done via the event number in the various modalities, the director agent keeps track of the

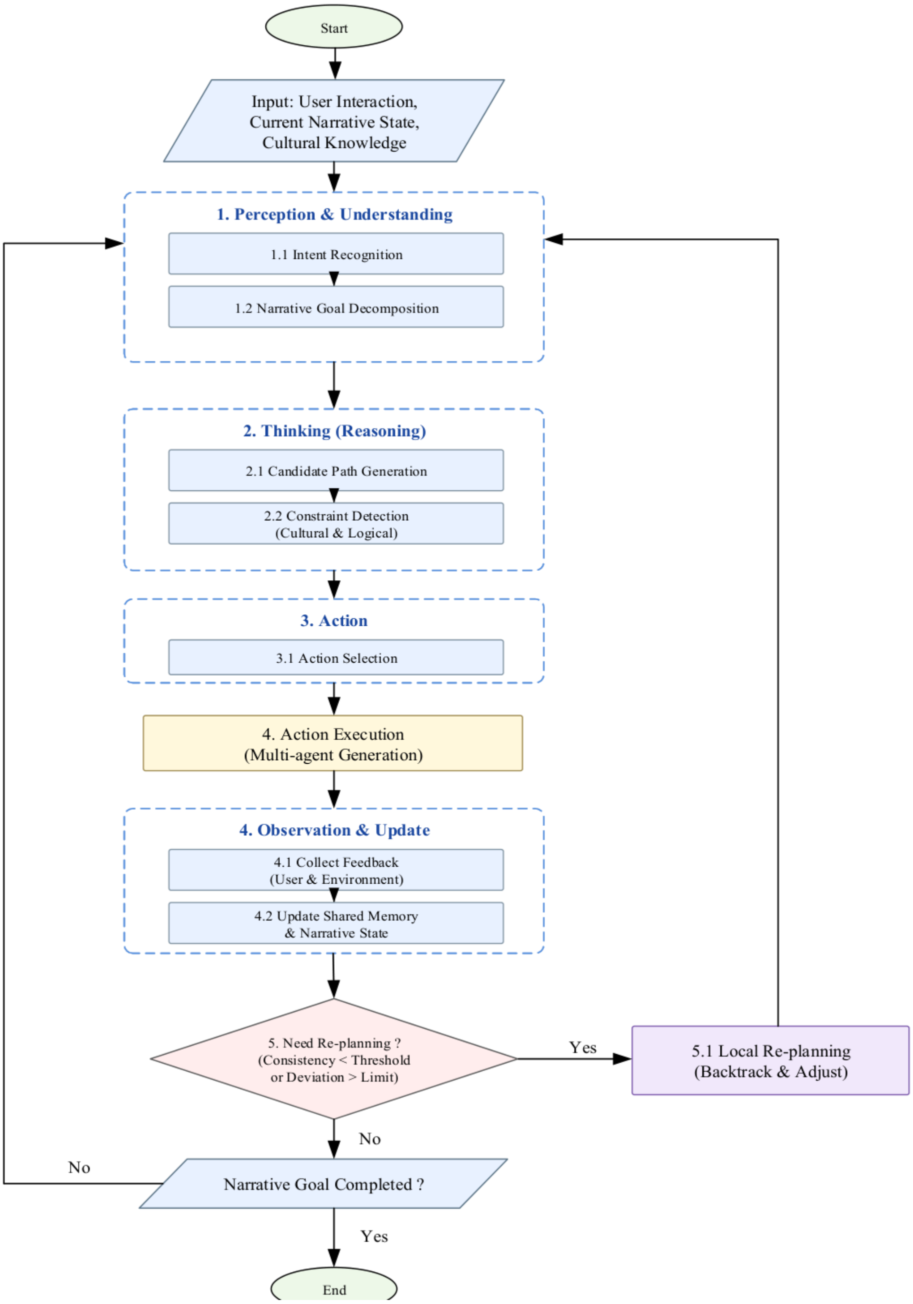

Fig. 4. Flowchart of the ReAct narrative planning algorithm.

times and beats, and detects deviations in action duration, scene transitions and music rhythm to do local corrections and synchronised output. The collaborative generation mechanism is shown in Figure 5.

E. NARRATIVE CONSISTENCY CONTROL BASED ON SHARED MEMORY

This shared memory is stored centrally, and can be broken down into short term working memory, plot memory, character memory and cultural memory, and is used to store past plot events, character states, scene changes, prop usage, and user choices. Each agent fetches all the relevant historical information before producing content and stores them back into the memory repository after the content is produced. The consistency control module checks identity of characters, causal relationship, chronological order as well as cultural norms. If the odds of conflict surpass a threshold, then the director agent is called to do local corrections/version rollbacks to reduce information lost, character drift and plot inconsistencies in long-term stories.

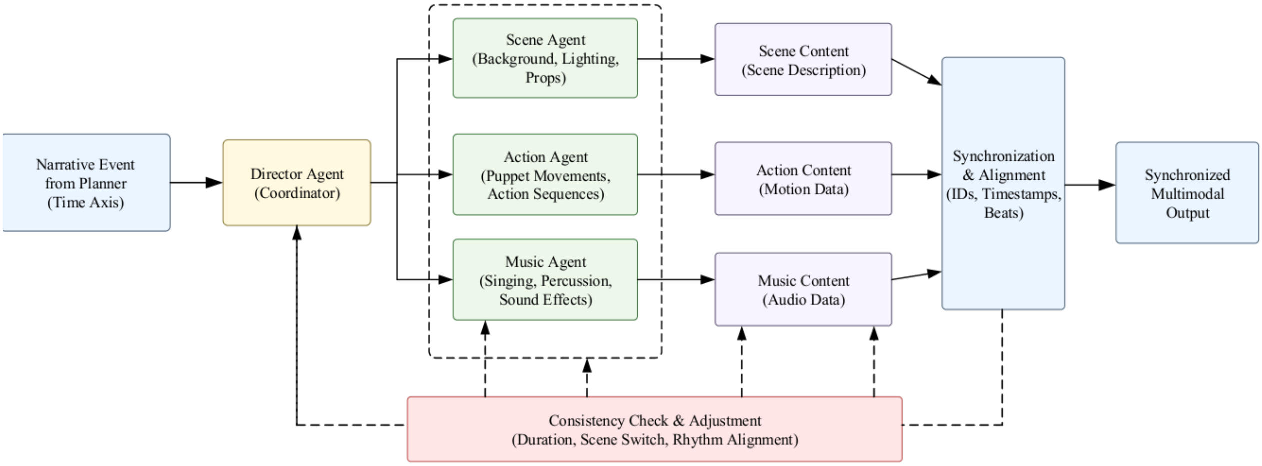

Fig. 5. Mechanism for the collaborative generation of multimodal content.

F. MULTI-AGENT NEGOTIATION AND CONFLICT RESOLUTION

Conflicts in multi-agent outputs primarily include inconsistencies between the plot and character behaviour, mismatches between the duration of actions and musical tempo, and contradictions between scene resources and cultural norms. The system employs a three-tiered processing mechanism comprising 'conflict detection—local negotiation—director's arbitration'. The comprehensive utility of a candidate solution x_j is defined as:

$$U(x_j) = \sum_{k=1}^{K} \omega_k q_{jk} - \lambda R_j \tag{13}$$

where q_{jk} is the score of solution x_j for the evaluation metric at item k, ω_k is the metric weight, R_j is the conflict risk, and λ is the penalty coefficient. The degree of consensus among agents is expressed as:

$$A = \frac{2}{N(N-1)} \sum_{i<j} \mathrm{sim}(o_i, o_j) \tag{14}$$

where N is the number of agents participating in the negotiation, o_i and o_j are the outputs of different agents, and sim is the semantic consistency. If A falls below the threshold, a local correction is performed; conflicts involving hard constraints result in immediate rejection, whilst conflicts involving soft constraints are resolved through weighted negotiation. If the maximum number of negotiation rounds is exceeded, the director agent selects the optimal solution or reverts to the previous stable state.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND DATASET CONSTRUCTION

The experimental platform has the Ubuntu 22.04 operating system and is equipped with an Intel Xeon processor, 128 GB of RAM and an NVIDIA RTX 4090 GPU. The implementation is done with Python3.10, PyTorch2.3 and a multi-agent communication framework. The data consists of traditional shadow puppet play script, character design, image of the scenes, action sequences, vocal music and cultural rules. Valid samples were created after performing deduplication, standardisation of formats, semantic annotation and expert review, and were split in a ratio of 7:2:1 for training, validation and test sets. Table 2 provides the number of each type of narrative resource used and the number of times the data were partitioned.

TABLE 2. Statistics of the Experimental Dataset

Resource Type	Total Samples	Training Set	Validation Set	Test Set
Narrative Scripts	1,800	1,260	360	180
Character Images	4,200	2,940	840	420
Scene Images	2,600	1,820	520	260
Action Sequences	3,600	2,520	720	360
Music and Audio Clips	2,400	1,680	480	240
Cultural Rules	800	560	160	80
Total	**15,400**	**10,780**	**3,080**	**1,540**

B. COMPARATIVE METHODS AND EVALUATION METRICS

The comparative experiment consisted of six model groups: template-based generation, Transformer sequence generation, general-purpose large-scale model generation, single-agent ReAct, multi-agent without ReAct and the proposed method. All groups used the same test set and generation length and each group tested five times. Other measures of evaluation were plot coherence, character consistency, cultural appropriateness, and narrative diversity and multimodal alignment, with response latency, replanning time and task completion rate being recorded. Twenty shadow puppet culture experts and 40 general users were invited to score the results of the subjective evaluation with a 5-point Likert scale. Reliability of the scale was done by using Cronbach's alpha and it was accepted as significant with the level of 0.05.

TABLE 3. Comparison of Narrative Generation Quality

Method	Plot Coherence (%)	Character Consistency (%)	Cultural Fidelity (%)	Narrative Diversity (%)	Multimodal Alignment (%)
Template-based	71.4	74.2	79.6	62.8	68.5
Transformer	77.6	76.8	78.9	72.5	74.1
General LLM	82.7	80.9	81.3	80.6	78.4
Single-agent ReAct	87.1	84.8	85.6	82.9	80.7
Multi-agent without ReAct	86.3	87.5	88.2	84.1	86.8
Proposed Method	92.4	91.7	94.1	88.6	92.8

TABLE 4. Results of Immersive Interaction Performance

Interaction Task	Response Latency (ms)	Replanning Time (ms)	Synchronization Error (ms)	Completion Rate (%)
Voice Dialogue	318	472	61	96.2
Gesture Branching	284	445	55	95.4
Scene Exploration	336	498	64	94.7
Multimodal Conflict	421	612	79	92.8
Continuous Interaction	389	574	72	93.6
Average	349.6	520.2	66.2	94.5

C. COMPARATIVE ANALYSIS OF NARRATIVE GENERATION QUALITY

The overall performance of different methods in digital shadow puppet narrative generation was compared across five dimensions: plot coherence, character consistency, cultural appropriateness, narrative diversity and multimodal alignment. Each method was run independently five times on the same test set; the evaluation results are shown in Table 3.

The proposed method resulted in 92.4 per cent plot coherence, 91.7 per cent character consistency and 94.1 per cent cultural conformity, each of which was 6.1, 4.2 and 5.9 percentage points higher respectively than the multi-agent approach that did not use ReAct; multimodal alignment also improved from 86.8 per cent to 92.8 per cent. Overall, the results show that dynamic programming can improve the coherence of the plots, and cultural constraints and cultural expression bias can significantly reduce the character drift.

D. ANALYSIS OF IMMERSIVE INTERACTION PERFORMANCE

To assess responsiveness in real-time interactive environments, five task categories were selected: voice dialogue, gesture branching, scene exploration, multimodal conflicts and continuous interaction. Response latency, dynamic re-planning time, synchronisation error and task completion rates were measured; the test results are shown in Table 4.

Average response latency for the five task categories was 349.6 ms, with the dynamic re-planning time of 520.2 ms, the cross-modal synchronisation error was less than 66.2 ms and the average completion rate of the tasks was 94.5 per cent. Of these, the lowest latency was the gesture branch task 284ms, and the multi-modal conflict task 421ms had a high completion rate of 92.8%, indicating that the ability to respond was not impaired even though the task was complex.

E. ANALYSIS OF ABLATION EXPERIMENT RESULTS

To see how much each of the following aspects contributes to the ReAct planning, the shared memory and conflict resolution module, multi-agent coordination module, and cultural constraints module were respectively stripped off the ReAct planning. Independent tests were performed on each of the variants, while keeping the model parameters and the experimental environment the same to maintain the consistency of the datasets, the results are presented in Table 5.

TABLE 5. Results of Ablation Experiments

Model Variant	Plot Coherence (%)	Cultural Fidelity (%)	Multimodal Alignment (%)	Completion Rate (%)	Latency (ms)
Without ReAct	86.9	92.8	90.6	87.3	302.4
Without Cultural Constraints	90.7	82.4	91.5	92.8	331.7
Without Multi-agent Collaboration	88.1	89.6	84.7	89.1	286.3
Without Shared Memory	84.9	90.5	88.6	90.2	320.5
Without Conflict Resolution	89.3	91.2	86.8	88.7	311.8
Full Model	92.4	94.1	92.8	94.5	349.6

TABLE 6. Comparison of System Performance

Method	Average Latency (ms)	GPU Memory (GB)	Throughput (Tasks/s)	Completion Rate (%)	Failure Rate (%)
Template-based	146.2	2.8	6.8	78.6	9.6
Transformer	221.5	4.7	4.9	83.4	7.8
General LLM	268.7	8.5	3.8	87.2	6.4
Single-agent ReAct	312.4	9.2	3.2	90.1	5.3
Multi-agent without ReAct	326.8	11.8	2.9	91.6	4.7
Proposed Method	349.6	12.4	2.7	94.5	3.2

TABLE 7. User Experience Evaluation Results

Method	Presence	Engagement	Agency	Cultural Perception	Satisfaction
Template-based	3.18	3.24	2.86	3.41	3.12
General LLM	3.72	3.81	3.54	3.68	3.69
Single-agent ReAct	4.08	4.14	4.21	3.93	4.07
Multi-agent without ReAct	4.16	4.22	4.03	4.18	4.20
Proposed Method	4.61	4.58	4.52	4.66	4.63

Excluding shared memory, cultural conformity dropped to 84.9% (7.5 percentage points lost) and excluding cultural constraints, to 82.4% (11.7 percentage points lost). When conflict resolution was turned off, multimodal alignment dropped to 86.8 per cent and the task completion rate dropped to 88.7 per cent. All modules had a significant impact on generation quality, with cultural constraints and shared memory making the most notable contributions.

F. PERFORMANCE COMPARISON OF DIFFERENT METHODS

The operational efficiency and stability of different narrative generation methods were assessed under uniform hardware platform and generation task conditions by statistically averaging generation latency, GPU video memory usage, system throughput, task completion rate and generation failure rate; the performance comparison results are shown in Table 6.

The proposed method has an average latency of 349.6 ms, utilises 12.4 GB of video memory, achieves a task completion rate of 94.5 per cent, and has a generation failure rate of just 3.2 per cent. Compared with multi-agent methods without ReAct, latency increases by 22.8 ms, but the task completion rate improves by 2.9 percentage points and the failure rate decreases by 1.5 percentage points. Although the planning loop and agent communication increase computational overhead, they significantly enhance the reliability of executing complex tasks.

G. EVALUATION RESULTS OF USER IMMERSION EXPERIENCE

The user experience experiment involved 20 individuals with expertise in shadow puppet culture and 40 general users. Participants were asked to rate their experience on a five-point scale across the dimensions of presence, engagement, sense of narrative control, cultural understanding and overall satisfaction, in order to assess the system's actual immersive performance. The results are shown in Table 7.

The proposed method scored 4.61, 4.58 and 4.52 respectively for immersion, engagement and narrative control, whilst cultural awareness and overall satisfaction reached 4.66 and 4.63, both of which were higher than those of the other comparison methods. The Cronbach's α coefficient for the scale was 0.912, indicating that the evaluation results possessed a high degree of internal consistency. Dynamic plot feedback and multimodal synergy effectively enhanced user engagement and cultural perception of digital shadow puppetry.

VI. CONCLUSIONS

ReAct planning and collaboration between multiple agents can be used to create the immersive narrative in digital shadow puppetry; with this method, a unified representation of cultural features, user intent, dynamic plot planning and multimodal content coordination can be achieved. The results of the experiment show that the advantages of this approach are focused on plot coherence, characters consistency, cultural conformity and multimodal alignment obtained by a 94.5 per cent of plot completion and a 4.63 user satisfaction score. The innovations include path search with awareness of cultural constraints, local re-planning based on provided environment feedback, and consistency control with the help of shared memory. There is still space for development in terms of the stability of complex and long-duration narratives, owing to the inadequate sample coverage of shadow puppet theatre styles, the high cost of expert rule annotation and multi-agent communication overhead. Cross-style knowledge transfer, memory compression, lightweight collaboration and real-time, multi-device interaction are all potential future research areas.

FUNDING

This work was supported by the Zhumadian City Youth Special Project on Scientific and Technological Innovation (No. QNZX202322), the Project Plan of Henan Science and Technology Research (No. 242102321143), the Henan Cultural Revitalisation Project—Special Programme for Cultural Research (No. 2024XWH025), the Young Backbone Teacher Support Program of Huanghuai University (No. 502300080013), and the China Postdoctoral Science Foundation Program (No. 2022M712484).

REFERENCES

[1] J. Gao and S. Juluri, "From Idea to Co-Creation: A Planner-Actor-Critic Framework for Agent Augmented 3D Modeling," in Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems, 2026, pp. 1–5.
[2] H. Marah and M. Challenger, "Adaptive hybrid reasoning for agent-based digital twins of distributed multi-robot systems," Simulation, vol. 100, no. 9, pp. 931–957, 2024.
[3] M. Younes, E. Kijak, R. Kulpa, et al., "MAAIP: Multi-Agent Adversarial Interaction Priors for imitation from fighting demonstrations for physics-based characters," Proceedings of the ACM on Computer Graphics and Interactive Techniques, vol. 6, no. 3, pp. 1–20, 2023.
[4] S. Zhang, Y. Xiao, R. Ma, et al., "RPGAgent: Driving Coherent Story-to-Play Generation with an LLM-Based Multi-Agent System," in Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, 2026, pp. 1–22.
[5] D. Wu, H. Shi, Z. Sun, et al., "Deciphering digital detectives: Understanding LLM behaviors and capabilities in multi-agent mystery games," in Findings of the Association for Computational Linguistics: ACL 2024, 2024, pp. 8225–8291.
[6] E. K. Tütüncü, Q. Zhou, F. Brudy, et al., "PlayWrite: A Multimodal System for AI Supported Narrative Co-Authoring Through Play in XR," in Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, 2026, pp. 1–26.
[7] I. Ahmed, M. A. Syed, M. Maaruf, et al., "Distributed computing in multi-agent systems: a survey of decentralized machine learning approaches," Computing, vol. 107, no. 1, p. 2, 2025.
[8] S. Uddin, B. Hussain, S. Fareed, et al., "A review of fault tolerance techniques in generative multi-agent systems for real-time applications," International Journal of Ethical AI Application, vol. 1, no. 1, pp. 43–53, 2025.
[9] H. Wu, A. Ghadami, A. E. Bayrak, et al., "Evaluating emergent coordination in multi-agent task allocation through causal inference and sub-team identification," IEEE Robotics and Automation Letters, vol. 8, no. 2, pp. 728–735, 2022.
[10] Y. Ge, Y. Ren, W. Hua, et al., "Llm as os, agents as apps: Envisioning aios, agents and the aios-agent ecosystem," arXiv preprint arXiv:2312.03815, 2023.
[11] I. Adabara, B. Olaniyi Sadiq, A. Nuhu Shuaibu, et al., "Trustworthy agentic AI systems: a cross-layer review of architectures, threat models, and governance strategies for real-world deployment," F1000Research, vol. 14, p. 905, 2025.
[12] D. Parmar, S. Olafsson, D. Utami, et al., "Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents," Autonomous agents and multi-agent systems, vol. 36, no. 1, p. 17, 2022.
[13] J. Wei, X. Wang, D. Schuurmans, et al., "Chain-of-thought prompting elicits reasoning in large language models," Advances in neural information processing systems, vol. 35, pp. 24824–24837, 2022.
[14] W. Huang, P. Abbeel, D. Pathak, et al., "Language models as zero-shot planners: Extracting actionable knowledge for embodied agents," in Proceedings of the 39th International Conference on Machine Learning. PMLR, vol. 162, 2022, pp. 9118–9147.
[15] S. Yao, J. Zhao, D. Yu, et al., "ReAct: Synergizing reasoning and acting in language models," in The Eleventh International Conference on Learning Representations, 2023, pp. 1–34.
[16] N. Shinn, F. Cassano, E. Berman, et al., "Reflexion: Language agents with verbal reinforcement learning," in Advances in Neural Information Processing Systems, vol. 36, 2023, pp. 8634–8652.
[17] S. Venkatraman, N. I. Tripto, and D. Lee, "CollabStory: Multi-LLM collaborative story generation and authorship analysis," in Findings of the Association for Computational Linguistics: NAACL 2025, 2025, pp. 3665–3679.
[18] Y. Ran, X. Wang, T. Qiu, et al., "BOOKWORLD: From novels to interactive agent societies for story creation," in Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics, 2025, pp. 15898–15912.
[19] K. Park, M. Kim, and K. Jung, "A character-centric creative story generation via imagination," in Findings of the Association for Computational Linguistics: ACL 2025, 2025, pp. 1598–1645.
[20] Y. He, "ShadowPlayVR: Understanding traditional shadow puppetry performance techniques through non-intuitive embodied interactions," in Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology, 2023, Art. no. 73, pp. 1–2.
[21] Z. Yao, S. Lyu, Y. Lu, et al., "ShadowMaker: Sketch-based creation tool for digital shadow puppetry," in Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 2024, Art. no. 417, pp. 1–5.
[22] Y. Liu, R. M. Williams, G. Xie, et al., "Promoting the culture of Qinhuai River Lantern shadow puppetry with a digital archive and immersive experience," arXiv preprint arXiv:2410.03532, pp. 1–23, 2024.
[23] W. Chen and Y. Guo, "Embodied transmission: Preserving intangible cultural heritage through procedural VR experience—A case study on Chinese shadow puppetry," in Proceedings of the 2025 5th International Conference on Culture, Design and Social Development. Atlantis Press, 2026, pp. 272–278.
[24] Y. Zhou and S. J. Jong, "Reconfiguring Chinese shadow puppetry in digital media," SN Social Sciences, vol. 6, p. 216, 2026.
[25] F. Huot, R. K. Amplayo, J. Palomaki, et al., "Agents' Room: Narrative generation through multi-step collaboration," in The Thirteenth International Conference on Learning Representations, 2025, pp. 1–34.

[26] T. Yu, K. Shi, Z. Zhao, et al., “Multi-agent based character simulation for story writing,” in Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants. Association for Computational Linguistics, 2025, pp. 87–108.

[27] T. Zhou, Z. Duan, C. Chen, et al., “AgentStory: A multi-agent system for story visualization with multi-subject consistent text-to-image generation,” in Proceedings of the 2025 International Conference on Multimedia Retrieval, 2025, pp. 1894–1902.