Orchid-Creator: An authoring tool supporting LLM-driven Interactive Narrative creation

Zhen Wu
zwuch@connect.ust.hk
Hong Kong University of Science and Technology
Hong Kong SAR, China

Serkan Kumyol
skumyol@connect.ust.hk
Hong Kong University of Science and Technology
Hong Kong SAR, China

Zhengyang Ma
zmaaf@connect.ust.hk
Hong Kong University of Science and Technology
Hong Kong SAR, China

Tristan Braud
braudt@ust.hk
Hong Kong University of Science and Technology
Hong Kong SAR, China

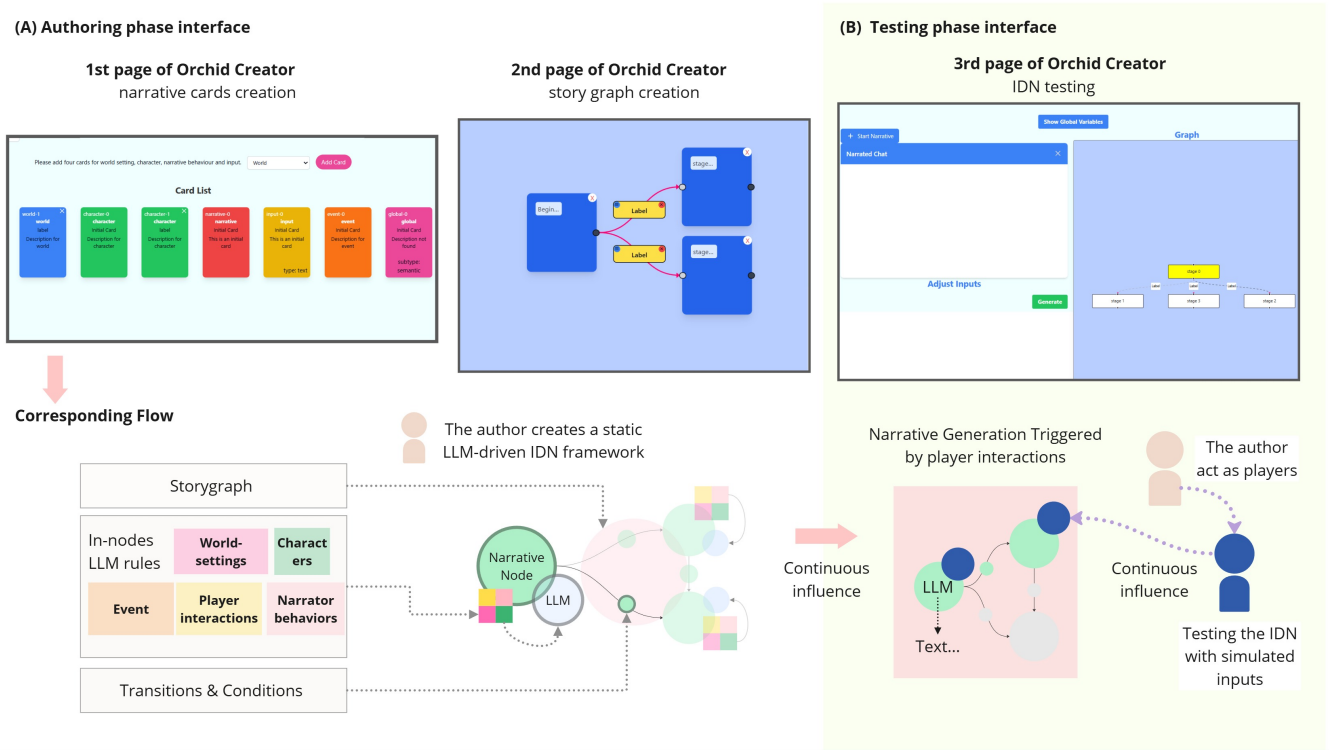


Figure 1: Overview of Orchid-Creator . Authors construct LLM-driven IDNs through two phases. In phase (A), they design the story graph, including in-node interactions and transitions, resulting in a static LLM-driven IDN framework. In phase (B), authors simulate player input to test the IDN created in Orchid-Creator .

Abstract

Large language models (LLMs) are reshaping interactive digital narratives (IDNs). However, creating complex interactive narratives while preserving narrative consistency remains challenging. We present Orchid-Creator (Orchid), an LLM-based authoring tool that represents IDNs as story graphs with a card-based interface for scene definition and conditional transitions. We evaluated Orchid in two studies: a usability study with eight authors, and a comparative study with 20 participants (authors, developers, and players) that compared Orchid to Twine and AI Dungeon. Authors reported that Orchid's features met their needs (card-based interface: 4.0/5; story graph: 4.38/5; variable setup: 4.5/5). Structuring narratives with Orchid was easier ($M = 6.0/7$, $p < .01$) and produced better-structured stories ($M = 5.3/7$, $p < .05$) than the alternatives, balancing author control ($M = 5.5/7$) with outcome diversity ($M = 5/7$, $p < .01$) and

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791426

maintaining comparable usability. Finally, a case study with an artist demonstrates Orchid's utility for interactive art.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; • **Computing methodologies → Natural language generation**; • **Software and its engineering → Interactive games**.

Keywords

Interactive Digital Narrative, emergent narrative, LLMs, Authoring Environments, creative support tools

ACM Reference Format:

Zhen Wu, Serkan Kumyol, Zhengyang Ma, and Tristan Braud. 2026. Orchid-Creator: An authoring tool supporting LLM-driven Interactive Narrative creation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3772318.3791426

1 Introduction

Interactive digital narratives (IDNs) are stories where authors define the interaction structures that guide the information players receive and respond to their inputs. Recent advances in large language models (LLMs) have opened new possibilities for narration and games, enabling richer character dialogue and NPC behavior [30, 37, 47]. However, their application in rich IDNs like *Disco Elysium* and *Death Stranding* remains limited. Designing complex IDNs requires explicit planning to manage branches, global state, and multiple endings. Yet current applications of LLMs in IDN creation often lack explicit support for complex branching mechanisms. Major LLM-driven authoring interfaces (e.g., AI Dungeon) primarily support a single, open-ended path rather than planned branches. As such, narratives often drift away from the plans set by the author, undermining authorial intent and narrative consistency [8, 53]. Prior work has proposed techniques to preserve author agency while leveraging LLMs' generative strengths [37, 54, 56]. However, many of these approaches rely on traditional deterministic algorithms and often lack support for fine-grained narrative planning.

To address such challenges in LLM-driven IDN authoring, we draw inspiration from traditional hypertext and multi-path fictions that define narratives through navigable graphs, enabling stories to progress along different paths based on player interactions. This paper introduces Orchid-Creator , a web-based system that leverages nodes, graphs, and LLMs for real-time narrative generation. Orchid-Creator presents a "blueprint" interface for creating a story graph. In this interface, cards allow the author to decompose the narrative into its atomic elements, which are used to prompt the LLM. The author arranges these cards into the nodes of the story graph to define narrative stages and define transitions between these stages. It also provides a real-time testing interface where authors can simulate player input and observe the progression of the LLM-generated narrative.

Compared with script-based IDN authoring tools, Orchid-Creator enables authors to design IDNs that go beyond their original expectations: open-ended content can emerge while authors significantly reduce manual writing effort. Compared to current LLM-powered interfaces, Orchid-Creator provides finer authorial control, allowing authors to create IDNs with meaningful yet complex structures and smoother narrative progression. It enables authors to precisely define the story's structure while producing a more coherent, enriching experience for players (as in AI Dungeon [52]). Orchid-Creator suggests an emergent authoring approach that focuses on creating an integrative LLM-driven IDN through small steps, embodying the principle that "the whole is greater than the sum of its parts" [40].

We conducted two user studies to understand the influence of Orchid-Creator on the IDN creation experience. The first study, with eight IDN authors, explores design ideas and authors' workflows. The second study, involving 20 participants (IDN authors, game developers, and players), compared Orchid-Creator with Twine (traditional script-based) and AI Dungeon (LLM-driven) to highlight Orchid-Creator 's strengths and weaknesses. In Study 1, authors rated Orchid-Creator 's core features as meeting their needs (card interface M=4.0/5; story graph M=4.38/5; variables setup M=4.5/5). Follow-up interviews showed that authors valued the card-based blueprint approach of Orchid-Creator while raising concerns around generation speed, LLM variability, content filters, and a learning curve.

In the Study 2, participants rated Orchid-Creator higher on idea exploration (Q1-1; $M = 5.53$, $p = .0111$ vs Twine) and expressiveness (Q1-2; $p = .0256$), on structure definition (Q1-3; $p = .0001$ vs AI Dungeon, $p = .0143$ vs Twine), structure experimentation (Q1-4; $p = .0103$ vs AI Dungeon), and support for authoring interactivity (Q1-5; $p = .0339$ vs AI Dungeon, $p = .0021$ vs Twine). Orchid-Creator also balanced author control (Q2-4; $M = 5.0/7$) with outcome diversity (Q2-1; $M = 5.2/7$; $p < .01$) and produced more surprising results (Q2-3; $p = .0246$ vs AI Dungeon, $p < .001$ vs Twine), while perceived narrative coherence showed no significant difference across tools. During interviews, participants appreciated the shared author/player agency and the clear blueprint and rule-based approach. They also mentioned several design suggestions, such as automated card management and finer control over narrative details, and raised broader concerns inherent to LLMs themselves, including quality, bias, and black-box behavior.

Finally, we demonstrate Orchid-Creator 's practical use through a co-design case study(Study 3) with a narrative artist, documenting and reflecting on her long-term engagement with Orchid-Creator . We also collected player feedback on the IDN created by the artist.

In summary, this paper presents the following contributions:

- Orchid-Creator , an authoring tool for LLM-driven IDNs featuring nodes and graphs.
- Two empirical user studies: **a usability study** (N = 8) validating core features and workflows, and **a comparative study** (N = 20) that evaluates Orchid-Creator against Twine and AI Dungeon on creative support, authorial agency, contingency, and perceived player roles. Participants rated Orchid's core features as meeting their needs (card interface 4.0/5; story graph 4.38/5; variables 4.5/5), found structuring narratives easier ($M = 6.0/7$, $p < .01$) and better ($M = 5.5/7$, $p < .05$) with Orchid-Creator than Twine and AI Dungeon while balancing author control with outcome diversity.
- A case study, "Let the Monster Speak," co-created with a narrative artist using Orchid-Creator as the IDN engine.

- Observations of the qualities introduced by Orchid-Creator across the three studies. We highlighted its encouragement of player co-creation, crafting, and the development of individual creative strategies.

2 Related work

2.1 IDN branching-based authoring tools

Interactive Digital Narratives (IDNs) is a field that combines storytelling with user interactivity through digital media, allowing users to shape the narrative they ultimately receive [17, 43]. Several typologies of IDNs have been identified in the literature, including hypertext fiction, interactive drama, computer games, and multi-user dungeons (MUDs) [43]. The development of specialized tools for authoring IDNs provides crucial support for its creation. An IDN authoring tool serves as a comprehensive, independent workspace that facilitates narrative development, whereas a metalinear story necessitates a dedicated writing tool from the outset [6, 45]. In addition to character development, event drafting, and interaction editing, branching is a key authoring aspect of IDNs that empowers non-linear structures [17, 25]. This aspect is exemplified in most authoring tools as story graphs with nodes that represent narrative chunks. Platforms like Twine [26], Inform [36], and Articy [15] provide various functionalities for designing complex, non-linear stories. StoryNexus [14] combines a world model with choice-based storytelling, and Varytale [21] divides narratives into "storylets". Some tools, like Inform [36] and Hugo [49], enable users to create IDNs using syntax to program the branching.

2.2 LLMs and Automatic Narrative Generation

Large Language Models (LLMs) are recognized for their rapid generation of creative content and their capacity to support story creation, effectively facilitating the transition from initial narrative ideas to final outcomes [1, 2, 38]. Researchers have proposed intuitive approaches to enhance the narrative creation process by engaging with LLMs. This includes using mind maps for guidance [3, 42] and employing metaphors to enhance creativity. For example, sketching lines for generating narrative arcs [10] and using magnets and dust for LLM-driven word-building [11]. Commercial applications Dream Gen [13], and Storyscape [46] primarily utilize LLMs interactively for narrative creation [31, 59]. While these works focus on "how to produce a narrative," IDNs require additional considerations regarding how players interact with a narrative. Although these works are not specifically designed for IDN creations, their interface design and user flow illustrate the importance of scaffolding users in story conceptualization intuitively. This is the focus we emphasize in the design of Orchid-Creator .

2.3 LLMs in IDNs creation and games

The limitations of traditional IDN design, which requires extensive dialogue scripting, highlight the need for LLMs to overcome the time-consuming processes involved [32, 48]. This is addressed in works using LLMs for planning events [39], guiding NPCs' behaviors [8, 53], and generating game levels governing complex functional constraints and spatial relationships [4, 20, 29, 50]. NPC dialogue generation using LLMs creates dynamic player interactions, as seen in works by Sun et al. with social chatbots [47], and Kumaran et al. with branching dialogues [30], and Ngaw et al. for character and object descriptions [37]. However, LLMs have untapped potential in managing the overarching narratives rather than fragmented content across various game aspects.

A representative LLM-driven IDN example in the mainstream market is AI Dungeon [52], where the LLM serves as a narrator, generating unlimited outcomes based on player input[58]. Similarly, in AI Roguelite, various LLMs create dynamic quests and dialogues [33]. However, in both cases, while the AI responds to commands, narrative disruptions occur, leading to a lack of continuity compared to traditional narratives.

Regarding authoring, several approaches have addressed authorial agency in applying real-time LLMs in IDNs. These methods allow authors to initiate a narrative, which an LLM then decomposes into character action plans[23, 28]. Furthermore, control is achieved through techniques like causality weighting for selecting plot fragments, as demonstrated by College Ruled [51]. Another approach enables writers to define high-level plots that LLMs convert into concrete character actions [54]. While these methods enhance narrative complexity, gaps in interactivity raise questions about their applicability to complex player interactions. Wu et al. [55] offer a creative approach integrating branching and LLMs, receiving positive user feedback. However, the work remains at the stage of a conceptual framework and has not been evaluated over large-scale narratives.

In summary, the current application of LLMs in IDNs reveals three shortcomings: low authorial agency, insufficient utilization of LLMs' real-time narrative creativity, and a lack of player interactivity, which often feature homogenized NPC conversations resembling typical chatbot interactions. Proposed novel methods remain abstract concepts rather than functional systems and lack evaluation over larger-scale narratives. We aim to address these issues in the design of Orchid-Creator .

2.4 Nodes and graphs for LLMs

Nodes and graphs are crucial in enhancing LLM-assisted narrative generation and interactive authoring. Systems like VISAR [60], XCreation [57], and WhatIF [35], FigJam [41] use nodes and graphs to represent story elements and their interconnections, facilitating an intuitive understanding of narrative structure and guiding LLM generation. By visualizing and abstracting LLM-generated outcomes with nodes and graphs, users can quickly understand the information presented. For example, Graphologue employs innovative prompting and interface designs to extract entities and relationships from LLM responses, creating interactive node-link diagrams that allow users to explore, organize, and engage in flexible, non-linear dialogues [22]. Additionally, the "cells, generators, and lenses" framework enables interactive writing interfaces, allowing end-users to experiment with and iterate on various configurations of inputs, models, and outputs across different writing tasks [24]. Nodes and graphs become an important design element in Orchid-Creator .

3 Motivation and Design

While recent developments in LLM-driven IDNs are gaining popularity, there are currently limited integrative authoring tools to

assist creators in developing. Existing platforms like AI Dungeon allow users to implement some high-level authoring elements, primarily positioning users as players, making it hard to produce original works out of the gaming system. On the other hand, tools that apply AI in facilitating authoring overlook the real-time creativity of AI. Many of them distinguish AI-assisted authoring from actual narrative play, described in subsection 2.2.

Wu et al. described a new authoring approach centered on the real-time usage of LLMs in IDN, called Orchid [55]. This approach represents IDNs in a node graph format, scaffolding AI-driven IDN construction. However, it serves as a conceptual framework, which is challenging to benchmark against existing authoring tools. Additionally, without a functional system, it's difficult to evaluate its impact on creators' long-term creative engagement. Our research aims to address this by developing a functional system that embodies Orchid's authoring approach while focusing on usability.

Orchid consists of three phases. The first phase focuses on in-node LLM generation rules. The authors define World Setting, character, Narrator Behavior, and player input. In the second phase, authors arrange the phase 1's content into a story graph and define the transition conditions between the nodes. This forms a node graph containing different LLM prompts stored in each node. In phase 3, the authors can test the IDN, stimulate player inputs, and observe the LLM reactions in real-time. While we maintain the three-phase authoring approach from Orchid, we propose the following design qualities(DQ) to address the issues raised by authors in their user study:

DQ1: Stricter trigger judgment. Wu et al. [55] described limitations in node transitions in their design. They were based on single stages rather than a history accumulating all the generated content. It also lacked precise control in transition decision-making as it was entirely governed by the LLM. We propose a new design combining *Global Variables* and condition checks, shifting from AI planning to having AI summarize variable values. The transition will be evaluated based on hard-coded thresholds of these values. This design functions similarly to an inventory system in games, helping users extract information from the narrative history, such as how many times a character has appeared. When defining a transition, users can utilize the Global Variables' intended values.

DQ2: Flexible testing. In Orchid, there was no flexible testing interface to experiment with different paths. The testing must flow from start to end. There was also no support for testing individual nodes. To cover these, we design a retrospective feature to facilitate users in debugging various routes, as described in subsection 4.3

DQ3: Dynamic memory management. In Orchid, due to the token context limit, each transition cleared the LLM's memory, resulting in the loss of historical task completion data from previous nodes. This led to inconsistencies in the LLM's output across stages. We address this by upgrading the memory database functionality to ensure the LLM maintains a coherent context from start to end.

DQ4: Structured input variable authoring. In Orchid, participants were confused by the authoring of input variables. It required users to describe the input type freely using natural language. However, the freedom made it difficult to design and envision how they would function in a narrative. Moreover, users found it challenging to explain the input to the LLM and worried about its understanding. To better guide the users, our design limits input variables to three common types mentioned in Wu et al. [55]: natural language, number, and boolean. In the backend, we frame how these inputs correspond to specific logic. This frees the users from explaining the input types to the AI.

4 System Features and Workflow

To better illustrate how Orchid-Creator supports the development of such IDNs, we create an example IDN called "Butterfly of an Android". The narrative flow is detailed in Figure 2. Players assume the role of a protagonist searching a computer system to recover their identity. The experience proceeds in nine linked stages.

- Stage 0 (Orientation): This stage serves as an introduction to the setting. Players begin with a free-form, natural-language conversation with an agent that personifies the system. After 5 dialogue lines, the system moves on to Stage 1.
- Stage 1 (Exploration): the game presents a choice among four files, each opening a distinct conversation with an AI bot.
- Stages 2-5 (Discussion): players inspect each bot's folder to learn background details. The player may enable an "Encryption" option at any time. This option converts all input and output into binary code, allowing the bot to develop greater trust with the player at the cost of immediate readability.
- Stage 6 (Revelation and Resolution): one bot, Metro-2, knows the answer to the player's quest. If the player earns Metro-2's trust, Metro-2 reveals the pivotal line, "You are my butterfly," which leads to the climax of the narrative. After this revelation, the player chooses to leave (Stage 7, a farewell to Metro-2) or to stay(Stage 8).
- Stage 7 (The End): if the player leaves, Metro-2 wishes goodbye to the player, and the story ends.
- Stage 8 (Epilogue): upon choosing to stay, the player inputs a number of days representing the length of their stay. The narrative shifts to a third-person, diary-style account of the player and Metro-2 covering that period.

In the remainder of this section, we detail the authoring process of this IDN using Orchid-Creator .

4.1 Phase 1 - Defining IDN components cards

Figure 3 presents the interactive components of Phase 1. This phase is inspired by the concept of *storybreaking* described by Daniel et al. [42]. It's a technique applied in television writing, where collaborators consider narrative as a particular configuration of existing elements. This suggests the reverse-engineering possibility of a narrative, based on different components. It is also inspired by the elemental thinking towards narrative, which suggests that narrative can be broken down into events, worldmaking, and descriptions [19]. Orchid-Creator decomposes the narrative into individual cards that include: World Setting, Character, Event, Narrator Behavior, Input, and Global Variable card, as shown in Figure 3. These cards are meant to be independent from each other, so they can be combined in phase 2, similar to how theater plays arrange characters and places in scenes and acts.

4.1.1 World Setting, Character, and Event cards. These cards allow authors to define the major elements of the narrative. The content

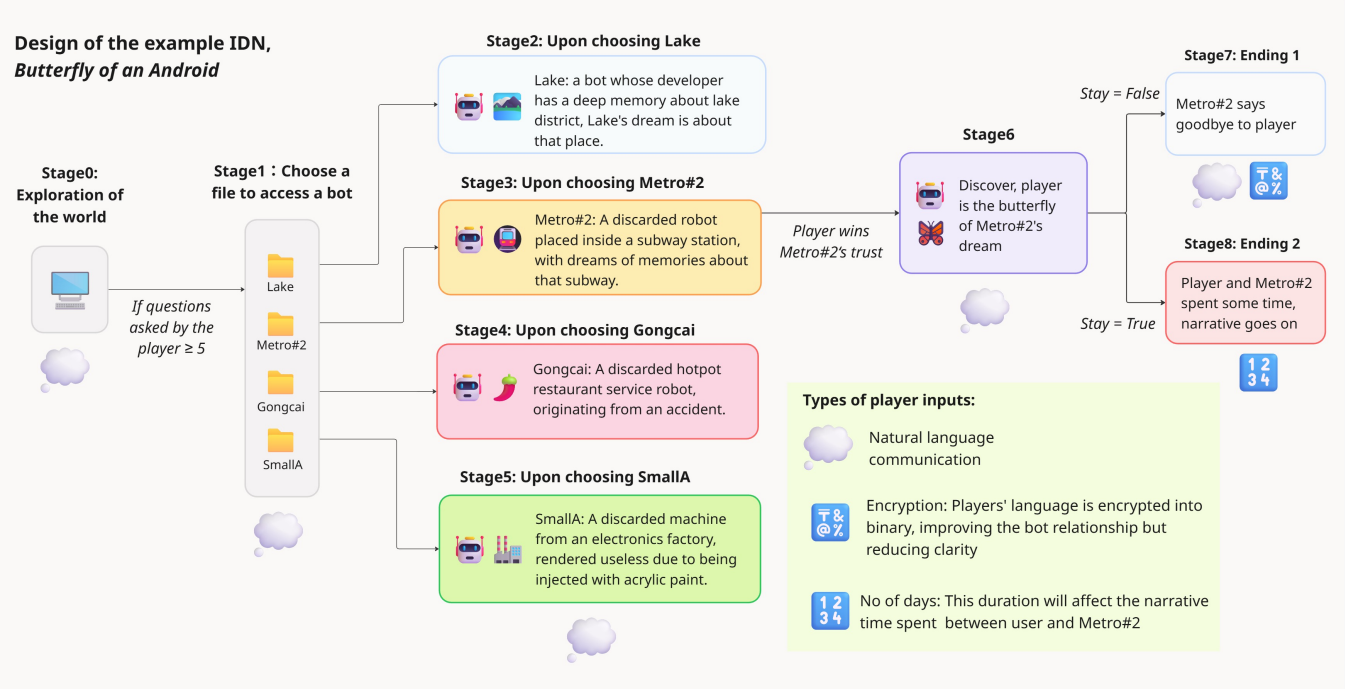


Figure 2: Story graph design of Butterfly of an Android. The narrative features a complex branching structure in 9 stages, covering three types of player input: Natural Language, Boolean, and Number. The narrative transition is governed by three types of Global Variables: Stage 0 to 1 is governed by Number. Stages 1 to 2, 3, 4, or 5, and 6 to 7 or 8, are governed by Boolean. Stages 3 - 6 are governed by Semantic.

entered in these three types of cards becomes part of the LLM's knowledge base, providing context for AI generation.

In "Butterfly of an Android", the author creates three world-setting cards: (1) the computer world, (2) the creator of this world, and (3) the general rules bots follow in the world. The author also creates four Character cards for each bot, each containing their background stories, and four corresponding event cards describing a specific event in each bot's life.

4.1.2 Narrator Behavior card. The content for this card becomes a direct part of the LLM's system prompt. Users are encouraged to specify the narrative perspective, instruct the desired behavior of the LLM, and the role it is meant to play. Authors are encouraged to create a separate Narrator Behavior card for each stage to better guide the main actions occurring within.

In "Butterfly of an Android", the author designs nine narrative stages, so nine distinct Narrative Behavior cards are needed. For instance, the first stage states, "Starting – You need to introduce the player. They are currently in the computer's RAM. The player can ask you questions about this world, and you need to answer them with some responses based on your imagination." These definitions allow the author to significantly change the style and tone of the story, switching from a third-person to a first-person perspective across stages.

4.1.3 Input card. This card defines the interactions offered to the player and how these interactions influence the narrative content. We design three types of inputs:

- "Text" or "Long text" creates an interaction that requires players to type words. The author describes in natural language how the player's text input influences narrative generation.
- "Boolean" creates an interaction where players are offered a binary choice. The author specifies what happens to the narrative if the player chooses true and vice versa.
- "Number" allows for numerical input. The author describes how the number influences the narrative and may provide a range.

In "Butterfly of an Android", the first stage starts with natural language communication. The author thus creates a "Long Text" card, defining the players to communicate with the system in their first-person perspective. "Boolean" cards represent the "Encryption" condition, as when set to True, player input is encrypted into binary text sent to the Bot, which responds in binary form, potentially enhancing their connection. Finally, a "Number" card is designed for stage 8, where the player can choose the number of days spent with Metro-2.

4.1.4 Global Variable card. Global Variable cards define elements that authors want to keep track of across the narrative generation process. For instance, objects the player has in their possession, information exchanged with NPCs, or categorical choices players made. An LLM is used to summarize the information contained at

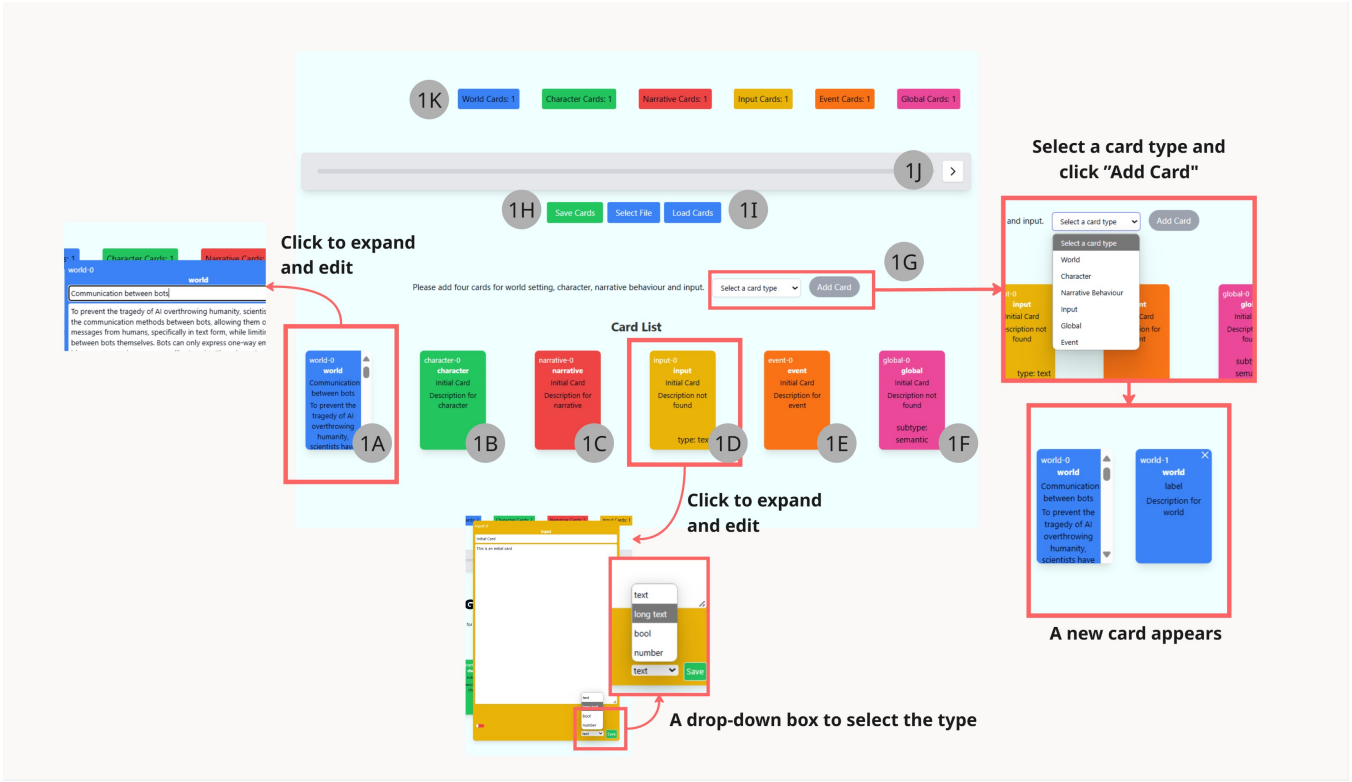


Figure 3: Interactive components of phase 1. They are defined as cards, each of which corresponds to a specific narrative element: World Setting (1A), Character (1B), Narrator Behavior (1C), Player Input (1D), Event (1E), Global Variable (1F). Authors can create new cards (1G), save and load cards (1H, 1I), and check the total number of cards (1K).

each turn of generation. Global Variables are also used to define transition conditions.

We design three types of Global Variables, each describing how the variable is updated and how it is judged against the threshold condition.

- "Semantic" variables operate by similarity comparison. The author inputs a sentence, which the LLM compares against the previously generated narrative content. If the narrative content exhibits a sufficient degree of similarity to the input sentence, the condition is met.
- "Boolean" variables use both true and false as determinations. The author inputs the statement to be checked in the description. The LLM will summarize the narrative history to assess whether the statement is true or false. Boolean variables provide specific outcomes. Users can apply both "true" and "false" values when designing the transition threshold.
- "Number" variables are used as counters. The LLM updates the number based on the narrative history. The author can assign a threshold number to this variable when designing the transitions.

In "Butterfly of an Android", the transition between stage 0 and 1 is determined by the number of questions asked by the player. A Number variable counts it, applying an LLM to summarize based on whether the user asked a question at each turn of narrative generation. Four Boolean variables determine which bot the player chooses to talk to between stages 1 and 2-5. A Semantic variable interprets the player's success in winning Metro-2's trust. If this condition is met, a transition is triggered from stages 3 to 6. Finally, two variables are used in the transition between stages 6 and 7-8. Firstly, a semantic variable determines whether Metro-2 has said the sentence "You are my butterfly." Secondly, a Boolean variable is based on whether the player chooses to stay with Metro-2 or not.

4.1.5 Interface operation. In the software interface, the author can add cards by clicking "Add Card" (subsection 4.1-1G) and selecting a card type. They can also double-click on a card to edit its content (1A). The total number of cards in each category is also displayed (1K). Both the Input (1D) and Global Variable (1F) cards have a drop-down box to select the type. By clicking "Save Cards" (1H), the current information in the cards is downloaded as a standalone file in JSON format. Through "Select File" and "Load Cards" (1I), authors can upload the JSON file containing the cards' information. Then, authors can click on the arrow (1J) to move to the narrative graph definition.

4.2 Phase 2 - Defining the story graph

As shown in Figure 4, the author adds nodes(2A) and populates them with cards (2B). By adding a card to a node, the information contained in the card is stored. The same card can be assigned to

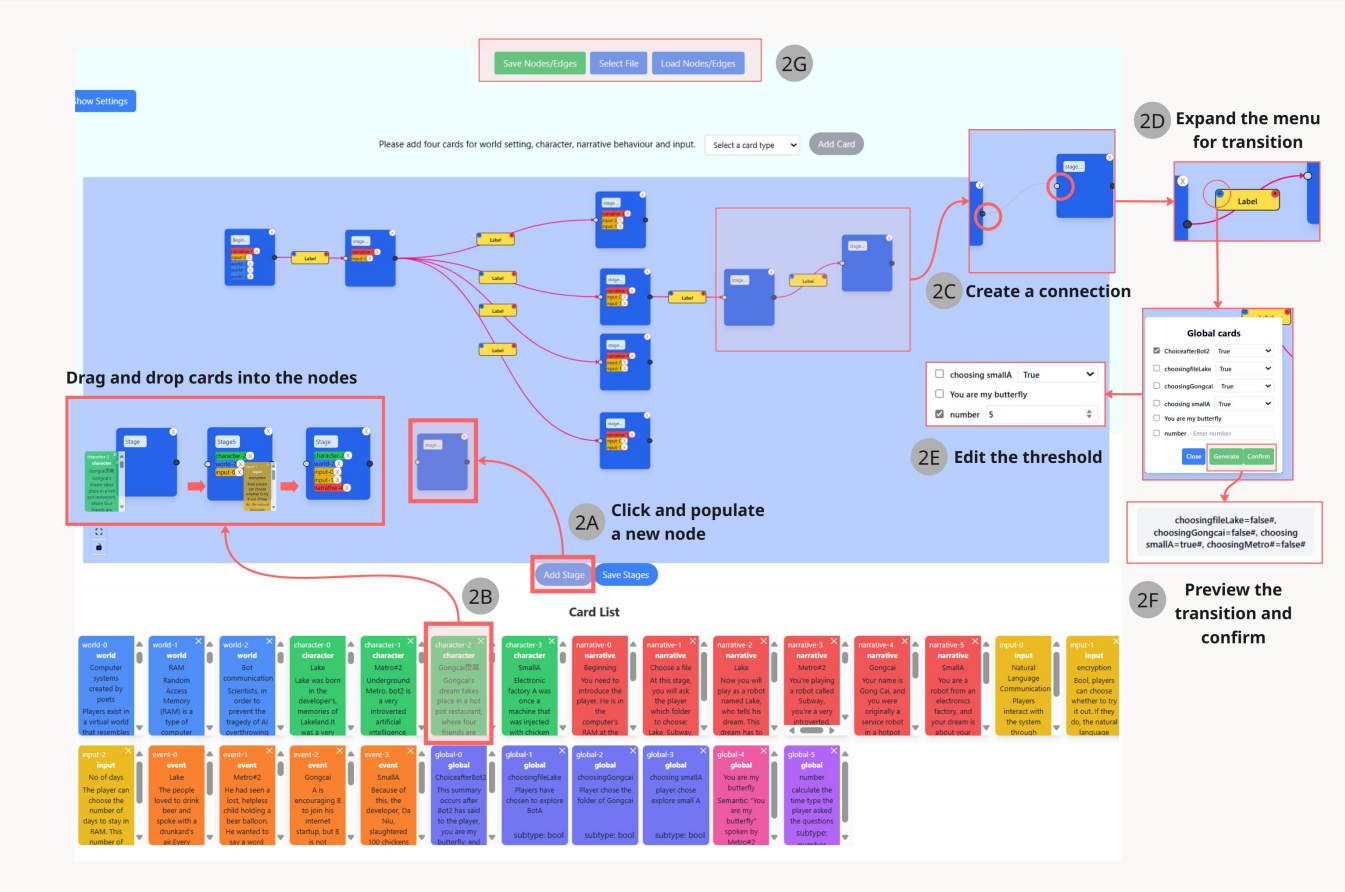


Figure 4: Interactive process of phase 2. Authors create the story graph by defining the nodes and vertices. After adding a new node(2A), the author populates it with cards defined in phase 1 (2B). Nodes are connected (2C) through transitions defined using Global Variables (2D, E, F).

different nodes. After defining nodes, the author connects them (2C) and manages their transition conditions through a menu containing all the Global Variables (2D). The author then selects the relevant Global Variables and enters the expected value for each of them (2E). They can preview the logical statement governing the transition (2F). Once all selected variables meet their transition criterion, this logical statement switches to true, triggering the transition between the adjacent nodes.

For instance, in "Butterfly of an Android", the transition between stage 0 and 1 is based on the number of questions the player asked. The threshold of the Global Variable "number" is set to 5. At each generation turn, the LLM evaluates the number of questions asked in total and compares it to this number. Upon counting 5 questions, it triggers the transition to stage 1. Authors can save and load their story graph by items in Figure 4-2G.

4.3 Phase 3 - Testing the IDN

The third phase of the authoring process is dedicated to testing the IDN. The goal of this phase is to provide the author with what the player will be presented with and to observe the main components of the narrative progression. Figure 5 displays the interface and summarizes the major interactions. Orchid-Creator provides a chat box, as well as adjustable inputs corresponding to the input cards defined in phase 1. This interface allows the author to manually simulate the inputs that will take effect in a given stage. A representation of the story graph created in phase 2 (Figure 5, 3D) shows the overall stage progression, with the stages that have been progressed highlighted in yellow. By default, authors start the testing from stage 0.

In "Butterfly of an Android", stage 0 applies "Natural Language Communication". The author thus simulates the player's input using text (3A). For instance, the author can type "Hi, where am I now?" to simulate the first player's input. Upon clicking the generate button (3B), the system displays the formatted prompt corresponding to this input for validation. After confirming, it is sent to the server, and the corresponding narrative is displayed in the narrative box (3C). The IDN graph remains in stage 0 as the transition condition has not been reached (3E). The author can check the status of the Global Variables (3F) to verify whether, for instance, the variable counting the number of questions asked by the player has been appropriately updated.

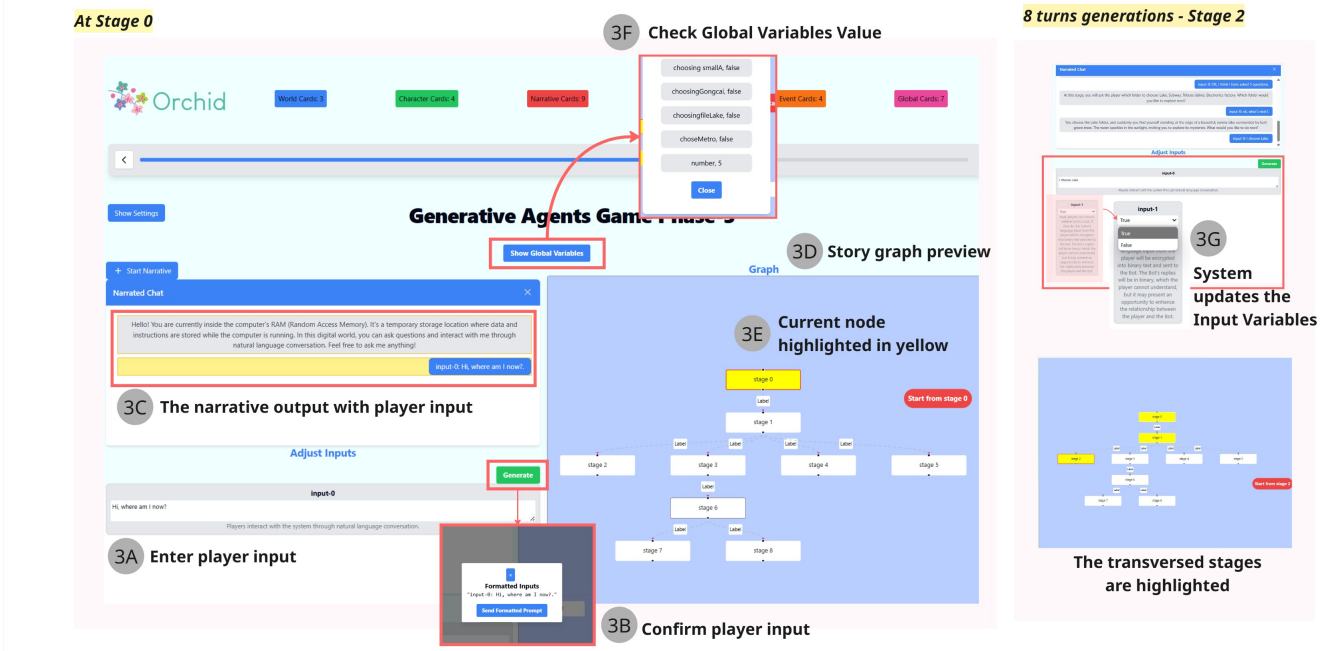


Figure 5: Interactive process of phase 3. The author can simulate user input (3A) and generate the corresponding narrative (3B, 3C). The story graph (3D) shows the current narrative node (3E). The author can check whether the Global Variables are updated appropriately (3F). Upon triggering the transition through player inputs, the story graph and player input are updated (3G).

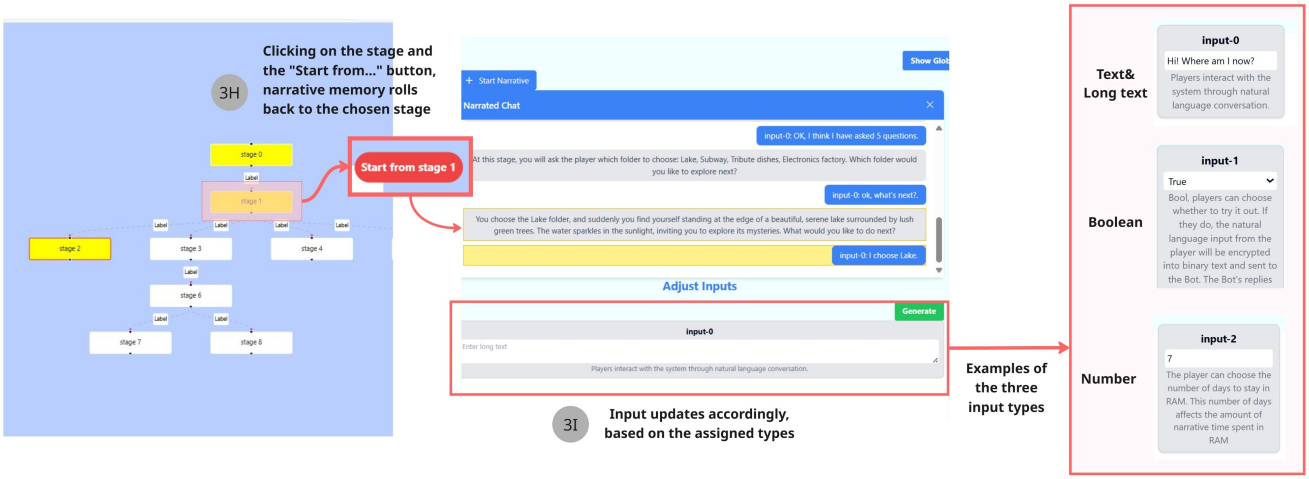


Figure 6: Navigating between stages during IDN testing. The author can select a previous stage (3H). The player inputs are reset, and Global Variables are reverted to their initial values (3I).

While the IDN progresses to stage 2, the system prompt, inputs, and knowledge base are updated based on those assigned to stage 2. Another player input, "Encryption", is introduced (3G). The author simulates this input by choosing between "true" and "false,"

together with the pre-existing text box for natural language interaction. The author then experiments with the different values of the "Encryption".

The author can revert to previous nodes to test different paths. In Figure 6, when the author makes a transition back from stage 2 to stage 1 (3H), the context is reset, and the inputs updated accordingly. The narrative history to stage 2 is cleared from the memory. As the narrative progression maintains a directed tree structure, the content cannot be generated based on nonexistent historical memory. For instance, the author can not directly test stage 2 if stages 1 and 0 have not been traversed. The 'Start from ...' button is disabled in this context.

5 Technical Details and Implementation

We implemented the front-end interface using JavaScript and ReactJS and built a backend server with Flask to query the LLM through the OpenAI API in Python. We utilized the Langchain [1] library for handling embeddings, prompt formatting, memory management, and the engineering framework.

On page 1, upon creating cards and populating contents, their information is stored within the backend. When users populate *World Settings*, *Character*, and *Event* cards into the node on page 2, these are transformed into text embeddings, which will be applied to retrieval-augmented generation [2]. *Narrator Behavior* and *Input* cards integrate the node's system prompt. The Input cards utilize Langchain's parameter exposure to guide the generation. *Global Variables* cards' definitions are stored in a dictionary, while their names are used in edge definitions on page 2, enabling users to set specific thresholds for node transitions.

On page 2, the story graph created by the users defines a state machine managed with the Networkx library [3] in graph representation. The graph comprises nodes (prompt, knowledge base, input methods) and edges (transition conditions). Each node is represented as a JSON file containing the following attributes: unique identifier, list of associated cards, formatted system prompt, list of retrieval information derived from the assigned *World Settings*, *Character*, and *Event*, and external inputs derived from the *Input* card. Transitions are stored in a dictionary, using the pair of nodes as *key*, and the relevant *Global Variable* and the threshold as *value*.

On page 3, the program responsible for generating the narrative will execute in the following steps sequentially, cooperating with two different LLMs in each turn of narrative generation:

(1) Process narrative: The system takes the current node ID to determine which system prompt and knowledge base to invoke. It then performs a similarity search within the knowledge base using the formatted *Input* to select relevant segments for the response. GPT-4o [4] generates the corresponding narrative, with the temperature set to 0.8.
(2) Dynamic memory management: The narrative history and input records are managed by the *chatmessagehistory* module from Langchain, tracked by a *memory.db* file. Newly generated narrative content is automatically added to the memory database. When the user reverts to a previous node, the history is deleted up to the beginning of the selected node.
(3) Calculating and updating Global Variable: We use an LLM (GPT-3.5-turbo [5]) with a temperature of 0.1 to resolve the variables' value, ensuring fast response and stable output. The system prompt for this LLM defines how to summarize Global Variables based on the description given by users, by distinguishing between Boolean, Number, and Semantic types, and uses history in *memory.db* as the source of judgment.
(4) Handling transitions: The system retrieves the neighbor nodes of the current node. It iterates through the different transition conditions of the edges to compare whether the Global Variables' current value meets the threshold. The first transition whose threshold is met will be triggered, and the current node ID will be updated.

6 User Study 1 - Usage and Usability

We first conducted a small-scale user study including participants with rich experience in IDN. We aimed to investigate their response to the system's design and usability at the component level, and to observe the specific types of work they create using this system.

6.1 Participants and setup

We recruited 8 participants from the authors' personal connections. The selected panel comprised 2 male and 6 female participants, and their ages ranged from 21 to 32(M=25.25, SD=2.11). They came from diverse backgrounds related to IDN, each bringing unique expertise.

P1, P5, P6, and P8 were experienced game designers who had previous narrative game projects (approximately 3 years of expertise). P7 was an interactive narrative film artist (2 years of expertise). P3 was a media scholar and interactive artist (5 years of expertise). P2 and P4 were interactive product designers (4 years of expertise). Five of them (P1, P3, P5, P6, and P7) had experience with LLM implementation in their interactive projects. While P2, P3, and P4 did not have IDN authoring experience, they had rich experience in IDN gameplay. The studies were conducted via Zoom. Orchid-Creator was deployed on a remote server for participants to access, while data were saved on a local computer for review. Their entire study procedures were recorded.

6.2 Procedure

We first provided each participant with a consent form, a brief, and a task sheet two days before the scheduled study. During this period, participants were instructed to prepare some elements of world-building, character concepts, and story elements. After this, we set up the online meeting and conducted one-on-one testing sessions. The researcher first introduced the workflow of Orchid-Creator using "Butterfly of an Android"(20 minutes), and participants could ask any questions about the usage. Then, the tool access link was sent to the participants, and they began to create an IDN using their prepared writing material. They could request assistance at any time. We gave the participants two tasks. Task 1: Use Orchid-Creator

[1] https://www.langchain.com/
[2] OpenAI text-embedding-ada-002 model
[3] https://networkx.org/
[4] https://openai.com/index/hello-gpt-4o/
[5] https://openai.com/index/gpt-3-5-turbo/

to design and develop a complete IDN project, including a beginning, an end, and at least one intermediate node. Participants could freely decide the complexity of the story graph and world-building. Task 2: Act as players to test their final design, at least completing one narrative path from the start to one of the possible ending nodes. After the tasks, they proceeded to answer the questionnaire and participated in interviews.

The entire creative process was quite flexible, with no strict time limits until participants were satisfied with their output. We observed that in the process of the design and development, P2, P3, P5, P6, and P8 implemented an iterative strategy. They mixed editing and playtesting, instead of separating the two. After completing a single narrative stage, they immediately tested it on page 3 and adjusted their content on pages 1 and 2 based on the results, especially when the output did not align with their intentions. After being satisfied with the node, they moved on to building the next node, gradually increasing complexity and defining the transition relationships between nodes. This was especially evident in P5's aggressive narrative design, as her prompts triggered multiple AI banners, leading her to repeatedly adjust the card content to avoid the AI's content filtering. P1, P4, and P7 implemented a linear process. They completed the contents on pages 1 and 2 together, then proceeded directly to the final testing.

Participants' total spent time on task 1 varied depending on the complexity of their IDN designs. Some participants were satisfied with their output during the session. P1, P4, P6, and P7 completed their designs in about 40 minutes without adjustments, as their narratives were relatively simple and open-ended. In contrast, P3 spent around 2 hours crafting his project, which involved complex narratives and character relationships based on a series of historical events. P2 and P8 spent approximately 90 minutes engaging with Orchid-Creator in the first session and felt unfinished. They planned another follow-up session with us to assess Orchid-Creator to iterate on their designs. These follow-up sessions, lasting about one hour each, were guided by researchers and recorded.

For task 2, we observed that P5 and P8 attempted 2-3 times to see if they could successfully reach different branches. After finishing one path, they would restart the game until they successfully reached an alternative path. Their IDN designs offered playability in terms of story structure, prompting them to see if the transitions they created could be realized. Other participants stopped the testing upon successfully finishing one narrative path and then stopped, as they felt they had experienced enough to anticipate the outcomes of further tests. These participants were particularly focused on how the diversity of player inputs could lead to different in-node generation, especially P2, P3, and P6, who employed linear IDN designs. When their input caused the narrative to progress to the next stage, they frequently used the retrospective functions to go back, test different inputs, and check the AI's responses.

Upon completion of tasks, we provided a questionnaire composed of self-designed questions on a 1-5 Likert scale, ranging from strongly disagree to strongly agree, assessing *satisfaction* and *perceived importance* of Orchid-Creator 's major design components: cards and drag-and-drop, trigger design, JSON saving/loading, node checking, retrospectives and experimentation with different paths, Global Variable setup and checking, and narrative coherence (evaluated based on the generated outcomes). The questionnaire was followed by a semi-structured interview, aiming to provide deeper insights, complementing the quantitative questions. The interview also addressed the participants' views on the system's creative support and the improvements necessary to integrate the system into their creative practice.

6.3 Analysis on IDN design

We reviewed the study recordings of participants and summarized the IDNs they designed. Regarding the story graphs, P2, P3, and P6 produced linear structures with stages connected sequentially, each consisting of four nodes. The other six participants created branching structures of varying complexity, with P8 having the most complex design that features five different nodes. The others generally implemented a simpler structure with one starting node and two branches. This demonstrated that Orchid-Creator successfully yielded different narrative structures. However, we found that participants' designs focused on the first-person perspective, assigning players a role. Additionally, most interactive inputs required players to communicate with the narrator or NPCs in natural language. Only P3 incorporated a digital dice mechanic, while no participants used Boolean inputs to guide choices. In the following, we list outputs from: P1-a game designer with AI implementation experience, P8-a game designer without such experience, and P3-an interactive artist and experienced IDN player, to demonstrate the different IDN types that arose from Orchid-Creator (Figure 7).

P1 designed a sci-fi narrative with a start and two branches. The story centered around choosing different character factions, each leading to events in two distinct stages. The interactive input was natural language communication. In Stage 0, the background of the story was established, allowing players to freely ask the narrator questions, and different characters will be introduced. If players made friends with Lin, judged by a Semantic Global Variable, they progressed to Stage 1, where events related to Lin's faction occurred. Conversely, if players engaged in conversations with K more than four times, judged by a number Global Variable, they were drawn into Stage 2, aligning with K's faction.

P8 designed a dramatic story featuring four characters: Li Ming, an entrepreneur; Zhang Ting, his fiancée; Chen Li, Zhang Ting's best friend and Li Ming's first love; and Wang Wei, Li Ming's close friend. Li Ming tested Chen Li's feelings through Wang Wei, who secretly had a crush on Zhang Ting and was considering whether this might be his opportunity. The IDN used natural language as input, with the player taking on Zhang Ting's perspective to gather evidence of her fiancée's infidelity in Stage 0. Once the player collected five pieces of evidence, they progressed to Stage 1, which took place just days before Zhang Ting and Li Ming's wedding. At this point, the system asked whether the player wanted to expose Li Ming. Players responded to the system, leading to three possible endings: suppressing their feelings (Stage 2), exposing Li Ming (Stage 3), or deciding to go through with the wedding (Stage 4). Each transition between stages corresponded to a Global Variable. Earlier stages focused on counting the number of evidence, while later stages involved semantic judgments based on the player's descriptive choices.

Initially, P3 created a linear IDN with four stages. The story revolved around a murder mystery set in an 18th-century rural

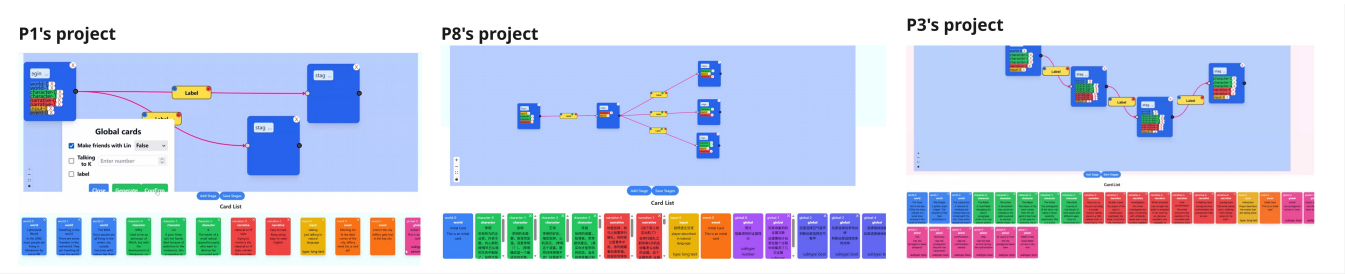


Figure 7: Story graphs created by P1, P8, P3 during the study.

Figure 8: Participants' responses to satisfaction and perceived importance of different design components of Orchid-Creator , in a 1-5 Likert scale.

town in France, featuring five characters: the Rabbi, the golem, the mayor, the writer, and the townsfolk, which formed five *character* cards. The player assumed the role of the writer. P3 designed two types of input: a dice roll (1-12) to determine the protagonist's luck and a text input for players to provide instructions to the narrator. The narrative structure progressed through the discovery of clues, starting with the murder scene (Stage 0), followed by the writer's investigation (Stage 1), experiences of being attacked and receiving help (Stage 2), and finally discovering the identity of the culprit (Stage 3).

6.4 Analysis on questionnaire and interview

We calculated the distribution of participants' scoring of the questions and the mean value. Figure 8 displays the evaluation results for the nine design components. Since our interviews aimed to support the questionnaire, we aligned the interview findings with related questionnaire items, yielding the following insights.

6.4.1 Facilitating narrative structuring. Orchid-Creator presented a good quality of facilitating narrative structuring. Global Variable setup and checking (Q2) were unanimously deemed important. Most of the participants also appreciated the current design for node checking (Q4), graph (Q6), and defining triggers (Q8). From the interviews, participants noted *Global Variables* and provided them better control over the narrative, allowing them to set constraints instead of fully letting AI decide on the narrative progression (P1, P6). The narrative structuring methods in Orchid-Creator also reduced the complexity of creating branching narratives. P4 connected the graph to procedural generation techniques she used in 3D modeling software, noting how the node-based generation logic deconstructed complex elements. P5 believed the system helped her decompose a complete story "through a series of operations," while P7 highlighted its ability to facilitate brainstorming.

6.4.2 Detailed design considerations. Card design (Q9) received high scores. P3 appreciated the concept of cards and the tabletop metaphor, while P2 appreciated the detailed variety of cards considered in our system. P1, P2, and P8 also stressed path experimentation (Q3) and project saving loading (Q7), which provided good considerations on the actual use scenarios.

6.4.3 Limitations on narrative coherence and interactive inputs. Narrative coherence (Q1) and defining interactive inputs (Q5) were the two items that presented relatively low scores. Narrative coherence was mostly reflected in the progression pace, as the real-time generation required a slight time delay instead of producing text immediately after player input (P1, P3, P5). P3 expressed that the ideal experience should be "fluid, akin to a game master in Dungeons and Dragons, who manages the overall pace of play to provide a seamless experience." Additionally, P5's narrative design faced significant restrictions from Azure OpenAI's content filter, leading her to feel that it "limits creative expression." Moreover, character personalities were sometimes confused, which P1, P2, P3, and P6 regarded as a significant issue. For example, a character originally designed as female was mistaken for male in P2's design. P1 noted that due to the unstable output quality of LLMs, it is "hard to serve as a reliable real-time text source" (P1).

For interactive inputs, participants mainly felt that they struggled to envision how these would connect to actual gameplay, especially *Number* and *Boolean* inputs. P5, P6, P7, P8 suggested that future developments should include direct integration of Orchid-Creator with gaming engines. This would allow them to use direct game input to influence the narrative. However, P8 believed that the categorization of *Text*, *Number*, and *Boolean* could better encourage player interaction compared to other LLM-driven games she knew. In her experience, "LLM games such as AI Dungeon give players too much freedom, and without a better framing of how the input will influence the narrative." This often leads the player not to know what to say".

6.4.4 Different usage habits. Overall, there were no "strongly disagree" responses, and all design components received scores between 3.5 and 4.4/5. This suggests that our system design is satisfactory. Several comments reflect different users' potential usage habits and the variations that might arise during use. P2 mentioned that the system requires a learning period and needs to invest significant time in crafting, whereas P1 expressed great curiosity about using LLMs for decision-making and indicated a desire to test the model's capabilities during the authoring process. P3 mentioned difficulties in switching writing styles between defining character and narrative behaviors. For example, when defining a character, he might write, "This is [character name]," but for narrative behaviors, he needs to adjust to a "GPT command format", such as "Please use the style of...". P3 and P6 described that they had a cautious mindset during authoring, worrying that if they built a large project with rich authoring content, the model might be unable to process it, resulting in wasted efforts.

6.4.5 Suggestions. P5 and P8, who have experience in authoring traditional IDNs, believe that combining pre-scripted content with generated text could align well with their workflows. Most participants suggested that future developments should embed Orchid-Creator into the broader media ecosystem, to cooperate with other aspects of game design, such as visual and sound.

7 User Study 2 - Tool Comparisons

Following study 1, we aimed to understand how Orchid-Creator expands user capabilities and enhances creativity compared to existing IDN authoring tools. In this study, we conducted a more rigorous quantitative comparison between Orchid-Creator and two IDN authoring tools: *Twine*, dedicated to static-text IDN, and *AI Dungeon* for AI-driven IDN. Both tools are currently the most popular in their category and thus represent the state of applications in the field. This study addresses three research questions:

- RQ1: What are the differences in creative support among Orchid-Creator , Twine, and AI Dungeon, and does Orchid-Creator sufficiently support AI-driven IDN creation?
- RQ2: How do authors perceive contingency and authorial agency when using Orchid-Creator to create IDNs compared to Twine and AI Dungeon, and in what ways does Orchid-Creator encourage open-ended creation?
- RQ3: How does Orchid-Creator compare to Twine and AI Dungeon in terms of usability, and what challenges do users encounter, and suggestions for Orchid-Creator ?

7.1 Baselines

Twine [26] is an authoring tool for pre-scripted, hyper-text style fictions, for creating text-driven games and interactive stories. As shown in Figure 9a, the users can create a new story by entering a title and adding building blocks of the narrative. Each block can be linked to others using simple bracket syntax. For instance, typing *[[Paul]]* creates a hyperlink between the beginning and the block after clicking Paul. We chose Twine as one comparative condition, as its block-based logic is similar to Orchid-Creator 's approach, helping us to investigate how and whether the LLM generation offers more creative elements and contingent outcomes.

AI Dungeon [52] is an AI-driven text adventure game creation platform that uses AI algorithms to generate narratives that respond to player commands (Figure 9). In Creator Mode, users can author initial settings for the generation by providing specific prompts and directions, including scenes, characters, and plot developments. Users can then test the IDN they defined in real-time and post it online for other players to experience. Unlike Twine, which supports block-based branching, AI Dungeon generates text linearly through AI algorithms. By comparing AI Dungeon with Orchid-Creator , we aim to investigate the advantages of using nodes and graphs for structuring AI-driven IDNs, and whether this method can provide broader authorial agency and encourage richer creativity.

7.2 Participants and study procedure

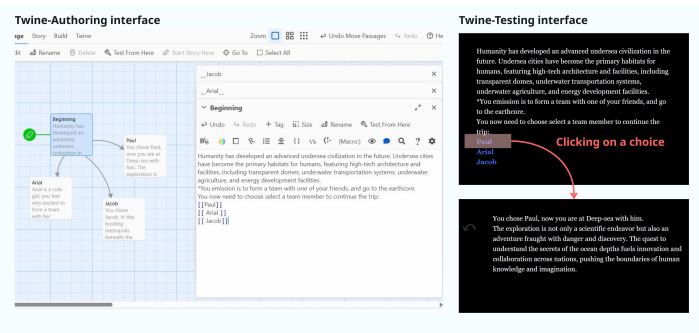
The comparative study was a within-subject study with 20 participants across three conditions. Participants were recruited through social media posts. None of the participants had participated in User Study 1. Their ages ranged from 23 to 32 (mean = 24.65, std = 2.44), with 10 males and 10 females. Compared to User Study 1, we did not require solid expertise in creating IDNs and included a diverse group of participants, including game and IDN developers, together with players, interaction designers, and traditional writers. Five participants came from interaction design backgrounds and identified themselves as open to engaging with IDN development in the future (P1, P2, P3, P5, P7, P8). Four had prior experience in game development and planning (P13, P15, P16, P18). Three participants had experience in creative writing and fiction creation (P10, P11, P17). P12 is an experienced tabletop RPG author. Six participants identified themselves as IDN gamers (P4, P6, P9, P14, P19, P20). In the rest of this section, we further label the participants as follows: Pn(I): Interaction Designer; Pn(G): Game Developer; Pn(W): Writer; Pn(P): Player.

The three conditions are based on the two comparative tools and Orchid-Creator .

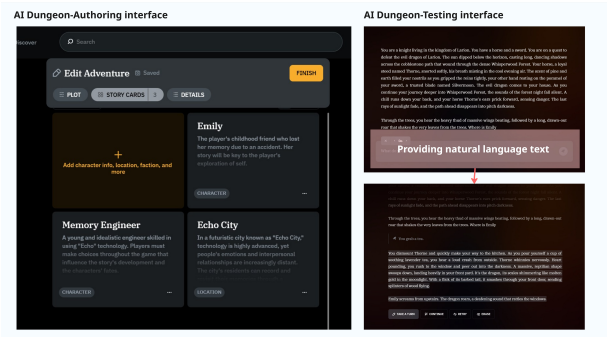
- Twine: Participants are required to use Twine's official online version. The goal is to use Twine's basic features to create a short interactive narrative or poem and test it.
- AI Dungeon: Participants are required to access the AI Dungeon website to experience the author mode, create a scenario, and then write a story while testing the outcome.
- Orchid-Creator : The goals and experience process are the same as in User Study 1. We limited the creation time to a maximum of 40 minutes to balance the time allocated for other conditions.

Once participants were recruited, the researcher provided the consent form and videos introducing the three tools' usage. Participants received a brief outlining the goals and the task sheet from User Study 1. They were asked to create world settings, conceptualize characters, and outline basic narrative elements, without a specific plot or structure, which they would develop while using the authoring tools. Then, participants had 2 days to prepare this sheet before the actual study.

During the study, participants constructed IDNs based on their task sheets using the three tools in an assigned Latin square sequence, at any time over the week. For *Twine* and *AI Dungeon*, we asked participants to make local screen recordings during their



(a) Authoring and testing interface of Twine. The player clicks on the hypertext to progress the narrative.



(b) Authoring and testing interface of AI Dungeon. The player types in natural language to progress the narrative.

Figure 9: Interface of Twine [26] and AI Dungeon [52]

usage. For Orchid-Creator , we used the same method as User Study 1, where users accessed the platform via a web link and set up a Zoom meeting with researchers for observation and clarification of issues.

The actual average time spent for Twine was 30 minutes, 20 minutes for AI Dungeon, and 40 minutes for Orchid-Creator . Upon finishing each condition, participants completed a questionnaire, keeping the same across the three conditions. After the participants had experienced all the conditions, we conducted interviews to explore their opinions and suggestions across the three tools.

7.3 Measure and analysis

The questionnaire contained the following aspects, designed in 7-point Likert scales:

(1) **Creative Support for IDN**, six questions: Idea Exploration and Expressiveness, adapted from Creative Support Index (CSI [9]). Two questions adapted from Kitromili et al. [25] about supporting and inspiring structure experimentation, two questions inspired by Louchart [34] about supporting and inspiring authoring interactivity. They aimed to address RQ1.
(2) **Emergent Narratives, Agency and Contingency**, seven questions: Perceived outcome diversity, and high-level structuring [5, 34]. Satisfaction with openness, surprising outcomes, and control versus machine creativity. Intention to consider contingency. Balance between authorial agency and contingency. Narrative coherence. They aimed to address RQ2.
(3) **Perceived Player's Role**, two questions aimed to address RQ2: investigating whether participants view potential players of their created IDN as co-authors of the narrative or as passive spectators.
(4) **Usability, using scenario**, six questions: Perceived usefulness and ease of use from UMUX-LITE [44]-including willingness to use and ease of use, integration into existing practice, clear links between action and outcome, results with effort [9], and easy to learn. They aimed to address RQ3.

The interview focused on the research questions. For RQ1, we asked the participants whether the tools expanded authors' capabilities in IDN, and how their creative mindsets aligned with or diverged from the mechanics of the three tools. We also asked about the scenarios in which participants might integrate Orchid-Creator into their creative workflows. For RQ2, we asked how the tools affected their authorial agency, contingency, or created a balance. For RQ3, we gathered insights on the benefits and drawbacks of Orchid-Creator compared to *Twine* and *AI Dungeon*. Finally, participants were asked to provide open design suggestions for future improvements.

We conducted repeated measures ANOVA to investigate the influence of three conditions on the 21 measures, followed by a Tukey post hoc pairwise comparison. P20(P)'s response was excluded due to outlier behaviors that created excessive similarity across conditions, but their interview data were retained for analysis. Figure 10 demonstrates the statistical data of the comparison, with highlighted significant differences. We used open-coding to classify the interview notes and categorized the insights reacting to the research questions.

7.4 Findings

We present our findings according to the research questions on creative support (RQ1) in subsubsection 7.4.1, contingency and authorial agency (RQ2) in subsubsection 7.4.2, and usability and suggestions (RQ3) in subsubsection 7.4.3. We summarize our main findings below.

- **RQ1:** Orchid-Creator was significantly better at facilitating idea exploration and expressiveness than Twine, and structural experimentation than AI Dungeon. Among the three, Orchid-Creator performed the best in supporting IDN structuring and interactivity authoring.
- **RQ2:** Participants found that both Orchid-Creator and AI Dungeon made IDN results more diverse and provided greater openness, more player co-creation dynamics than Twine, highlighting LLMs' contributions. Orchid-Creator demonstrated greater openness than AI Dungeon, while both Orchid-Creator and Twine provided better control and structuring,

highlighting the story graph's benefits in balancing control and contingency.
- **RQ3:** Orchid-Creator outperformed AI Dungeon and Twine in authoring rewards. It was preferred over Twine for willingness to use and better integration into existing practices. Participants also found it easier to connect their creative actions to outcomes in Orchid-Creator than in AI Dungeon.

7.4.1 RQ1: Creative support. **1) Idea exploration and expressiveness.** Post hoc pairwise comparisons on questions related to RQ1 revealed that Orchid-Creator (M=5.53) supported idea exploration (Q1-1) better than Twine (p=0.0111). Participants believe that Orchid-Creator encourages idea iterations, with initial efforts focused on rule-setting (P10(W), P15(G). While it offered more freedom than writing a script, experimenting with different authoring content was considered necessary to explore this new IDN typology and achieve satisfying outcomes(P1(I), P2(I), P13(G), P15(G), P18(G)). Additionally, Orchid-Creator allowed users to be more expressive (Q1-2) while creating IDNs (p=0.0256). The low difference between Orchid-Creator and AI Dungeon in these two questions may indicate the advantages of LLMs in providing a "quick and dirty" workflow from initial ideation to seeing results.

2) Structuring. Orchid-Creator was rated significantly higher regarding facilitating structure definition (Q1-3) than both AI Dungeon (p=0.0001) and Twine (p=0.0143). Although the hypertext logic of Twine is relatively simple, the process of forming a complete story structure was quite cumbersome(P4(P), P3(I), and P10(W)). P10(W) stressed that Orchid-Creator offered a blueprint-style editor that provides greater freedom by "enabling rapid experimentation with different logic before shaping the specific text". For encouraging structure experimentation (Q1-4), Orchid-Creator outperforms AI Dungeon (p=0.0103). P8(I) and P16(G) highlighted that page 3 aided their structural experimentation by visualizing the real-time progression on the story graph.

3) Interactivity. In terms of supporting authoring interactivity (Q1-5), Orchid-Creator was rated significantly higher than both AI Dungeon (p=0.0339) and Twine (p=0.0021) in inspiring player interactions (p=0.0011). P8(I) emphasized that "constraints that Orchid-Creator provided are important, it can encourage players to provide high-quality input." P6(P) believed that rich interactivity authoring can result in great playability, which AI Dungeon failed to provide (P1(I), P20(P)).

7.4.2 RQ2: Contingency and authorial agency. **1) Contingency.** There was a significant difference between Orchid-Creator (p=0.0074) and AI Dungeon(p=0.0051) compared to Twine in outcome diversity (Q2-1). P6(P) explained this diversity "making the story appealing." P7(G), P8(I), and P20(P) believed that the LLMs introduced a lot of surprises, which enabled players to alter the course of the game. Regarding whether the tools provide openness that meets creative expectations (Q2-3), Orchid-Creator outperformed both AI Dungeon (p=0.0246) and Twine (p=0.0000). Additionally, AI Dungeon was rated significantly higher than Twine (p=0.0246). Orchid-Creator encouraged the intention to consider contingency (Q2-5) during the creative process, more than Twine (p=0.0024). P1(I) stated, "Twine forces me to become imaginative, instead of letting me trust the system's creativity."

2) Authorial agency. For supporting high-level structuring (Q2-2) and a satisfying level of control (Q2-4), AI Dungeon was rated significantly lower than both Twine (p=0.0349 and p=0.0312, respectively) and Orchid-Creator (p=0.0001 and p=0.0023, respectively). These suggest that the branching structures, which are the shared quality between Twine and Orchid-Creator , effectively gave users a sense of control over IDNs. From the interview, we observed that Orchid-Creator enhanced its level of control through the stage segmentation, which decomposed a storyline into finer steps (P11(W), P17(W), P18(G)). As P7(G) noted, "In Orchid-Creator , nodes can represent different events within a scene, with each event assigned a unique narrative generation guideline." This was lacking in AI Dungeon, as it only allowed for initial prompt engineering similar to traditional ChatGPT (P5(I)), causing problems such as "producing lengthy narrative without truly progressing a plot" (P6(P)). Additionally, participants noted the poor LLM performance in AI Dungeon (P11(W), P12). Despite their attempts to manipulate AI Dungeon's prompts, it rigidly guided the narrative. P10(W) felt that the content produced by AI Dungeon, possibly due to the model's specific training in Dragon and Dungeon style, significantly limited her own writing style.

3) Balance. Orchid-Creator was also significantly rated higher than Twine in the balance between authorial agency and contingency (Q2-6, p=0.0072), and higher than AI Dungeon. From the interview, we found that the balance originated from Orchid-Creator 's ability to handle transitions. P17(W) described the plot progression in AI Dungeon as "feeling stuck in a loop", whereas Orchid-Creator did not have this issue. P6(P) remarked that the transitions provided the AI with a "creative outlet", allowing it to "refresh its mind by jumping to the next stage when it runs out of ideas in the current one."

4) Player co-creation. In Q3-1, most participants considered that the IDN created by Twine provided less dynamism in player co-creation compared to both Orchid-Creator (p=0.0005) and AI Dungeon (p=0.0451). From the interviews, we found that while AI Dungeon performed poorly in terms of control, participants appreciated it for encouraging player creativity (P15(G), P6(P)). P8(I) pointed out that the distinction between the author and player was quite blurred in both Orchid-Creator and AI Dungeon. This may indicate that LLMs contributed to player co-creation. However, the level of co-creation in AI Dungeon was lower than in Orchid-Creator . We believe this was due to AI Dungeon's inherent "stuck in a loop" phenomenon, which weakened the impact of player inputs on the plot.

7.4.3 RQ3: Usability and suggestions. **1) Willingness to use.** All participants expressed a strong willingness to use Orchid-Creator in the future. Compared to *Twine*, participants were more willing to use Orchid-Creator (Q4-1, p=0.0156) and believed that Orchid-Creator could be more easily integrated into their existing practice (Q4-2, p=0.0156). Participants noted that Twine required extensive text writing (P12 and P14(P)), making it more labor-intensive and cumbersome when creating content(P2(I)). Users of Twine needed a significant writing background (P8(I), P11(W)). P10(W) stated that Orchid-Creator better supported her freedom in writing style, which suited her current practice.

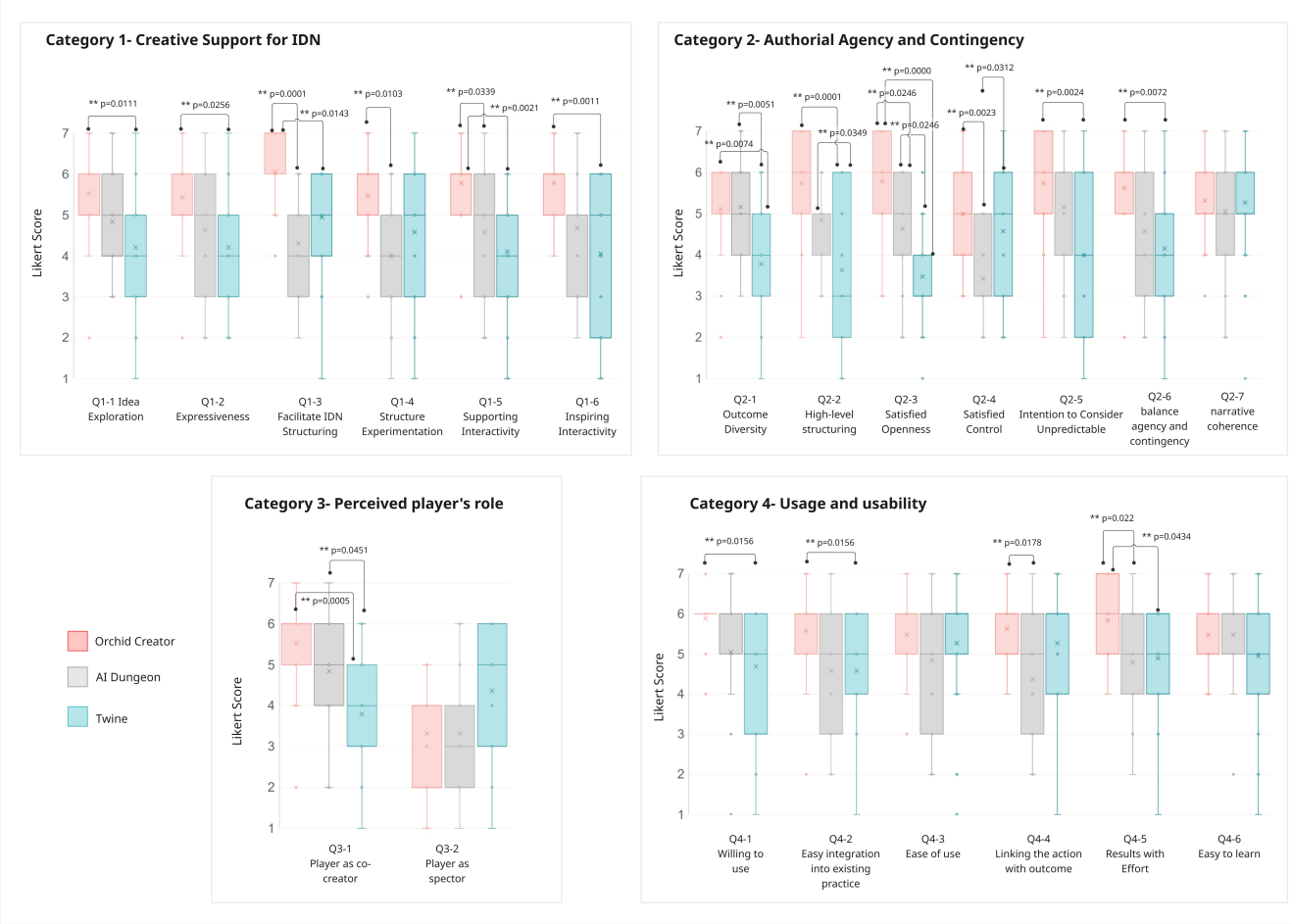


Figure 10: Evaluation survey results for Orchid-Creator across four categories: creativity support (Category 1), authorial agency and contingency (Category 2), perceived player co-creation (Category 3), and usage and usability (Category 4). Box plots illustrate the data, with pink representing Orchid-Creator , grey for AI Dungeon, and blue for Twine. Asterisks indicate significant differences between items based on post-hoc Tukey pairwise comparisons following repeated measures ANOVA.

2) Affordance and reward. Compared with *AI Dungeon*, Orchid-Creator provided more clarity in how interactions with components relate to the narrative outcome (Q4-4, p=0.0178). This benefited from page 3 showing the current stage of the narrative (P11), and the breakdown of story components helped organize their thoughts better than starting from scratch (P15). Orchid-Creator outperformed both *AI Dungeon* (p=0.022) and *Twine* (p=0.0434) in terms of the results being worth the effort invested (Q4-5). This reward was considered by P8(I) as "high-quality input leads to high-quality output". The feeling of reward also originated from Orchid-Creator 's easy to understand (P2(I)) and the confidence it brought in precise control (P18(G)).

3) Potential usage of Orchid-Creator . Some participants anticipated that it was important to have a well-defined framework before using Orchid-Creator for implementation. As P13(G) noted, otherwise, extensive debugging may be required. P3(I) described Orchid-Creator as "engineering stories". Participants expressed interest in personal writing of new forms of fan fiction (P11(W)) and brainstorming game design(P16(G)).

4) Suggested improvements. Participants desired automatic generation of story cards from short stories, aligning with traditional habits (P7(G), P13(G)). The speed of information leakage was a concern. P17(W) and P18(G) noted the excessive "copying and pasting" of the LLM from original character definitions. This may stem from the text embedding mechanism applied in Orchid-Creator . P13(G) suggested optimizing for different literary genres, as they may involve different narrative structures and modular enhancements. P18(G) envisioned beginning the process with a pre-scripted narrative piece, instead of generated, to provide players with a determinate context.

7.4.4 General thinking of AI-driven IDN. P1(I) likened AI's contributions to popcorn movies, where diminishing returns lead to stories requiring significant time and effort to become exceptional.

P8(I) and P11(W) expressed skepticism about the AI's black-box nature, wanting to understand the output's relationship with different authoring components, particularly how it used previously defined text. P7 noted that automatic generation by AI can introduce biases, such as assuming a character named "Jen" is female despite being defined as male.

8 User Study 3 - Case Study

For exploring the potential usage of Orchid-Creator and its influence on the workflows of artists, we present a longitudinal case study in the form of artist collaboration. The artist, a woman who preferred to remain anonymous as Artist A, specialized in creating immersive environments and interactive narrative art. She had a particular interest in AI technologies. Figure 11 demonstrates the timeline and process of our collaboration with her. Throughout this period, we asked her to actively report on her process in creating a self-reported project documentary. We also conducted several meetings with her to help her adjust the directions and iterate on the ideas. We interviewed her about her experience throughout the different phases of the collaboration.

8.1 Concept description

The IDN experience, titled *Let the Monsters Speak*, aims to encourage meaningful conversations between parents and children about how society "creates" monsters, and the media portrays and amplifies the existence of these monsters for profit, no matter if the monster really exists. The IDN will select one type of monster within: Scapegoat, Mirror, Mascot, and Hero Maker – each representing distinct social dynamics. Through the IDN, players witness and contribute to the monster's formation and propagation within society via natural language communication. The monster's personality and appearance evolve with the characters' descriptions and rumors, as well as the player's own descriptions. After the narrative experience, the monster's description will transform into an AR model viewable via mobile phone. Figure 12 presents the IDN design.

The project in Orchid-Creator has a total of six stages. Stages 0 to 3 follow a linear progression, branching into two endings(stage 4-1, 4-2) at stage 3. The narrative begins with a news report that sets the stage, chosen by the LLM based on a monster typology and a random object, such as chairs or a wooden block. In stage 1, players select places to visit, where they meet various villager NPCs and learn their perspectives on the monster. Once the system determines that public opinion about the monster has intensified, it prompts players with the question, "How is the monster like in your mind?" After this, in stage 2, the players encounter a news reporter who seeks to exploit the monster for profit and discuss it with him. After enough dialogue, players enter stage 3, where they meet a merchant selling creative toys related to the monster. Following some discussions with the merchant, the system asks players whether they believe in the truth of the monster described in the merchant's words. This leads to two branching paths: If players choose to believe, they enter stage 4-1, experience an ending, where they encounter monster hunter Alex Carter, who guides them through dialogue to change their beliefs. Conversely, if players choose not to believe, they experience a bittersweet ending (stage 4-2). Rejecting the monster's commercialization doesn't stop its impact- instead, the town thrives on the myth. Monster hunter Alex Carter then discusses with players about this phenomenon and reveals the truth behind the monster.

8.2 Creative process summary

8.2.1 Phase 1: System familiarization and background research for world building, 1st month. We deployed Orchid-Creator to a cloud server for the artist to access at any time. At the beginning, through a live demonstration, we introduced all the modules of Orchid-Creator , ensuring Artist A understood how to use it. Artist A spent about a week exploring Orchid-Creator , using simple narrative structures and asking the researcher any questions she had about the system. After that, the artist chose the theme of "Monsters," drawing on her previous research interests. In parallel, she gathered background narrative material to support the narrative construction through case studies of real and fictional monsters. This resulted in a typology of monsters (Scapegoat, Mirror, Mascot, Hero Maker). It provided direct sources for World Setting and Character cards.

She also explored rumor diffusion and media amplification mechanisms. This led to a five-stage diagram visualizing how monsters become a general belief through media effects. It includes: the creation of monster events, media amplification and shaping, dramatization of events, public audience presence, and long-term iterative feedback. It inspired her initial ideas for the IDN sequence.

At this stage, Orchid-Creator subtly influences the artist's design strategy, with the system's logic and the artist's literary research mutually shaping each other.

8.2.2 Phase 2: Designing interactive mechanism, 3rd month. In this phase, she produced several paper prototypes, such as a physical "monster station," and invited users to explore the interactions. Then she began to design the interaction by producing sketches, while using Orchid-Creator to experiment with the input cards. She tested the input cards on a single stage within Orchid-Creator .

There were four iterations in total. The first version utilized player gestures – such as when the player zoomed in and pulled two characters closer together on the screen – to trigger a new piece of narrative. However, when implementing in Orchid-Creator , she found the interaction lacked a strong connection with narrative content, making it indistinguishable from a non-AI narrative. Also, it failed to leverage *input cards'* story adaptation ability. The second version applied the number of stacked wooden blocks to affect the narrating style. For example, if there were 13 blocks, the story was suited for a 13-year-old's level of critical thinking. This iteration still didn't satisfy the artist, as the inputs' influence on narrative seemed "too contrived and simple". In the third version, the players picked up wooden blocks to select corresponding narrative perspectives: public, media, and monster – all of which alter the AI's narrating perspective. However, actual testing with Orchid-Creator revealed an inconsistency in the narrative style that might make it incoherent.

In the final version, she proposed a significant creative shift: from AI storytelling, where players passively received AI-generated information, to AI guiding players to express their thoughts. She also found this approach better aligned with her artistic intent in encouraging critical discussions, rather than delivering a didactic experience. She removed the physical interaction elements and

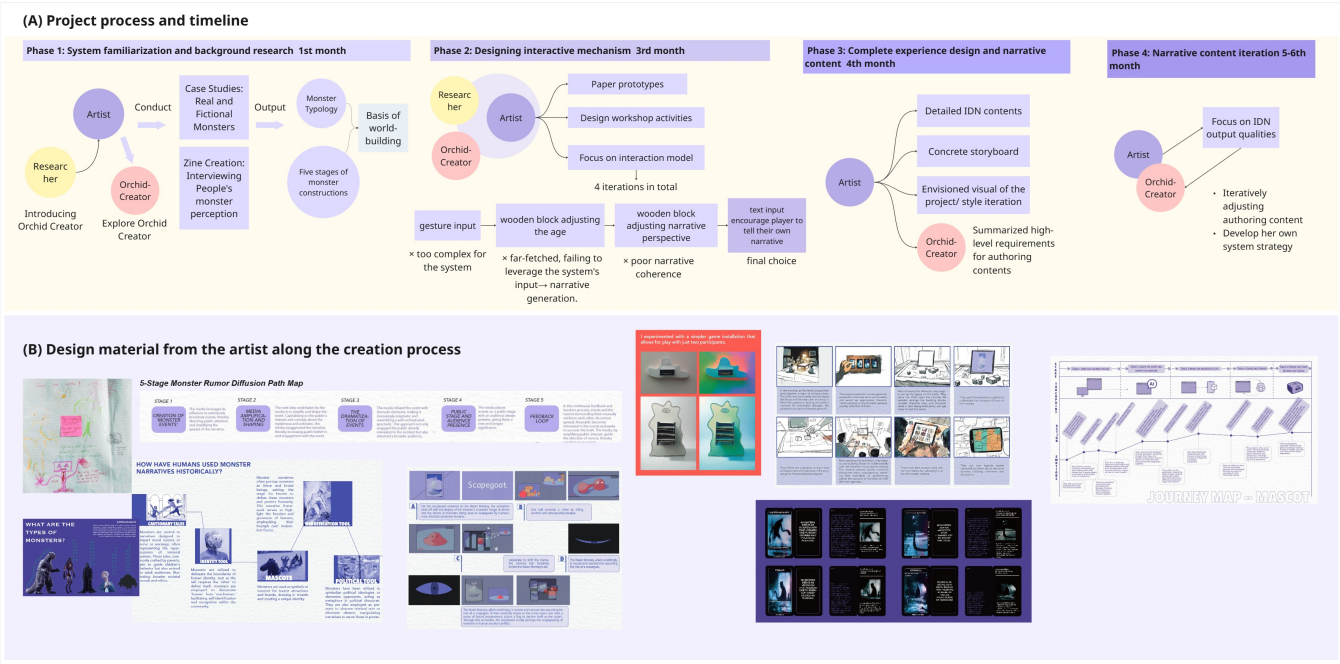


Figure 11: (A) Creative process of the artist collaboration lasting for 6 months. (B) Design material produced by the artist during the artwork development.

focused on natural language interaction. She then used *input cards* to define what topics the players were encouraged to discuss at different stages.

In this phase, we observed that Orchid-Creator influenced Artist A's interaction design a lot and introduced some challenges. The system's logic encouraged the artist to focus on how interactive inputs can connect with the self-generative nature of AI-driven narratives. It also prompted consideration of how to design interactions that form a coherent and progressive experience. Additionally, we noted that Artist A tends to apply a single type of input modality (in her case, natural language communication). She also focused more on the interaction model at this stage rather than the quality of the generated text content.

8.2.3 Phase 3: Complete experience design and narrative content, 4th month. Following the previous phase's shift from passive storytelling to encouraging children to tell their own stories, Artist A began constructing the project's overall experience. This encompassed the visual design of the IDN interface, the 3D design of the monster and other NPCs, the entire experiential flow, and game-related derivatives. The anticipated experience was a visual novel generated in real-time from text produced by Orchid-Creator . After each IDN experience, the system would ask the player to describe a monster in their mind, and the text would be applied to a 3D monster generation, projected in AR. Her creative activities encompassed sketches, concept art, and storyboarding.

Artist A also began to summarize several high-level requirements for the authoring content in Orchid-Creator . She asked the researcher about the system's limitations and sought tips on how to better define the narrative content. She paid particular attention to AI's questioning approach, such as avoiding overly didactic or morally correct content. In the meantime, she finalized her story graph design.

We noticed that in the artist's creative vision, the IDN governed by Orchid-Creator was just one part of the complete experience. The entire storyboard revealed richer ideas, including what the physical interactive devices would look like, the post-narrative experience, and the social dynamics between parents and children. They were elements that go beyond the original functions of Orchid-Creator . Additionally, our system could not currently provide visual connections and multimedia support. This created a gap between the generative narrative and the corresponding visual, which Artist A considered essential for a complete interactive art project.

8.2.4 Phase 4: Narrative content iteration, 5th-6th Months. This phase involved the most frequent engagement with Orchid-Creator . Artist A made nearly 20 adjustments to her project in Orchid-Creator until she was satisfied. She explained that her strategy involved isolating different stages to ensure narrative output quality, and then adjusting the transitions through *Global Variables* to make sure the narrative progresses fluently. Artist A described her process as "almost every step was a repetitive attempt," where, within one narrative stage, she simplified and refined overly complex plot and character settings when mistakes happened. She then modified the Narrative Behavior cards to adjust the dialogue logic, followed by adjustments to the corresponding Input cards based on the revised dialogue.

She shared that during testing, she paid attention to the AI's comprehension capabilities, finding that overly complex information in cards yielded poor results, so she simplified her authoring

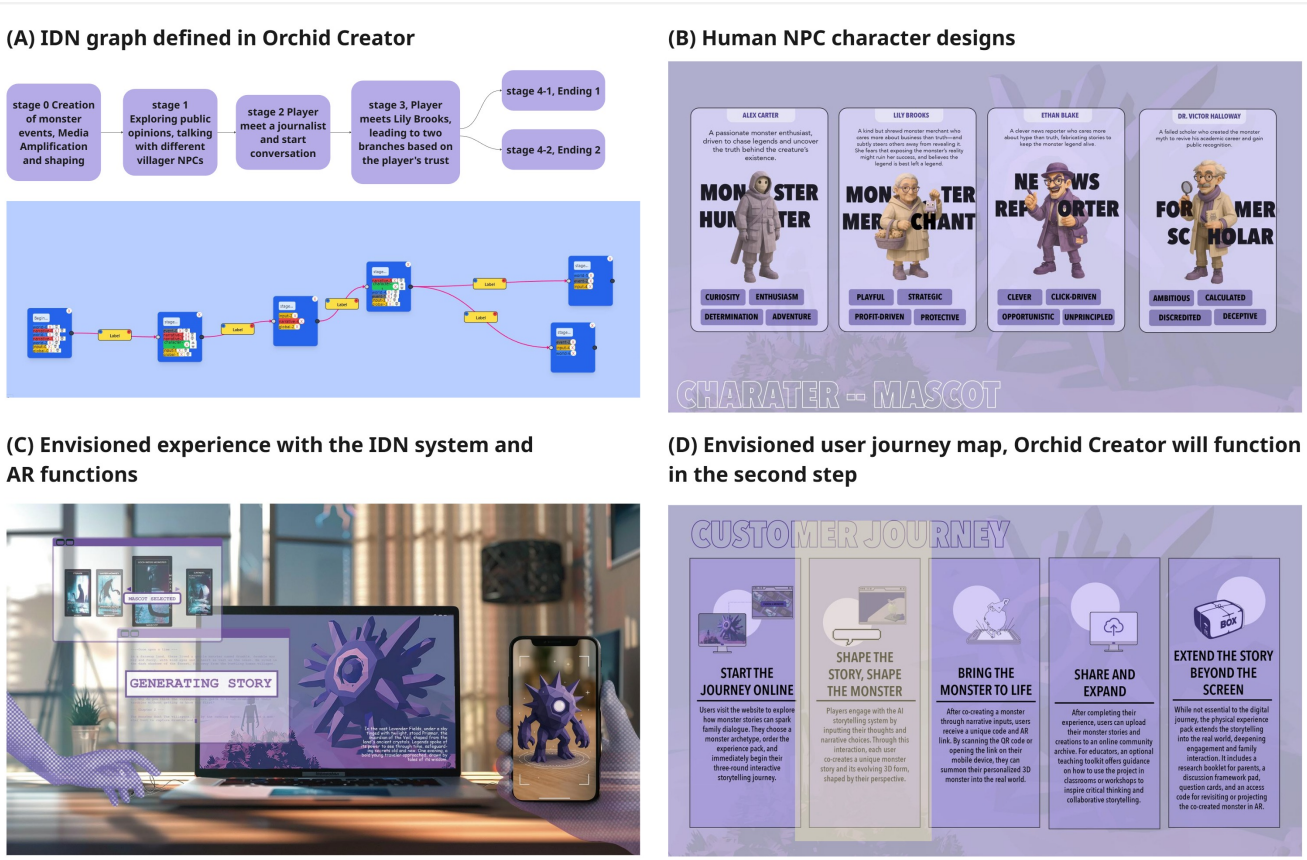


Figure 12: Design output from the artist collaboration. (A) Original IDN graph that the artist created using Orchid-Creator . (B) Human NPC designs that are part of the character cards are filled in Orchid-Creator . (C) Envisioned experience with the IDN system and AR function. (D) Envisioned user journey map, from engaging with the IDN to obtaining an actual monster model from the experience, which the player can then share and further expand.

content at some point. Also, her original design involved having the AI handle multiple tasks within stage 2: first generating monster descriptions, then guiding players to gather information through NPC dialogue, followed by automatically generating a location where the AI would act as the NPC. She discovered that this placed excessive demands on the AI's information processing, so she split this stage into multiple smaller stages linked together (Stage 2 → Stage 0-1-2).

She also discovered some tricks for Orchid-Creator : recognizing that AI couldn't process large volumes of information at once, she utilized its memory continuity by placing world-setting elements only in stage 0. This way, after the AI-generated descriptions, this memory carried over to subsequent stages, naturally forming the settings without needing additional world-setting cards. This approach particularly helped with stage 2, which initially included a card with around 1000 words of monster descriptions, given that the initial stage already provided monster descriptions through the news.

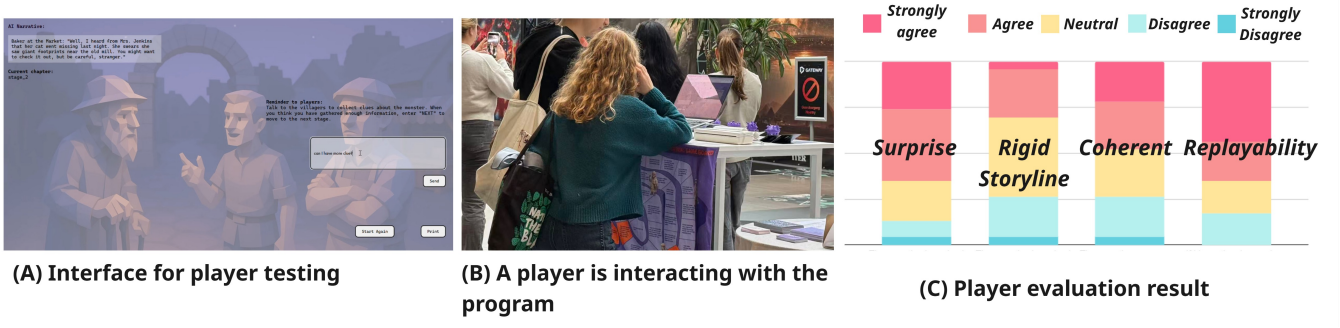
Additionally, she noted that during player-NPC interactions in Stage 2, the AI failed to distinguish between NPCs without explicit character cards. Initially, she instructed the AI: "You are now acting from the perspective of a different villager to discuss your opinions about the monster..." The system lumped together statements from different villagers and caused confusion with pronouns. She resolved this by specifying in the Narrator Behavior card: "Assign each villager interacting with players a distinct role, such as gatekeeper or farmer..." This made NPC dialogue more concrete and eliminated the sense of incoherent rambling.

This phase produced a lot of knowledge for how an actual author applied Orchid-Creator to adjust the narration, and it was to our surprise that Artist A would develop her own strategies to adapt to the system's limitations.

8.3 Player testing

To observe players' reactions to the IDN output from Orchid-Creator , and address the artists' desire to engage audiences through an exhibition format, we co-hosted a player testing exhibition.

8.3.1 Set up. The project was set up as an experience booth at a local design festival, equipped with a computer for participants to



(A) Interface for player testing

(B) A player is interacting with the program

(C) Player evaluation result

Figure 13: (A) Interface for player testing. (B) A player is interacting with the program. (C) Player evaluation result.

experience the IDN. Since Orchid-Creator 's output remains text-based, we produced a tailored program using the same logic. This utilized the cards and story graph implemented in Orchid-Creator , which ran in the backend. The frontend incorporated narrative images created by the artist with a tailored UI design (Figure 13-A). The system switched narrative backgrounds at different stages, forming an interactive visual novel. Visitors attracted by the booth included parents and children (Figure 13-B), while Artist A facilitated the experience on-site, distributing questionnaires about the game to those who completed the entire IDN experience. The questionnaire included questions on a 5-point Likert scale, including whether the narrative was surprising, coherent, followed a rigid storyline, and the replayability. Artist A also interviewed several players and discussed their feedback on the experience.

8.3.2 Result. A total of 23 players filled out the questionnaire. Each player spent approximately 15-20 minutes completing the entire IDN. As shown in Figure 13-C, the majority of participants agreed that the narrative was creative and surprising (15 out of 23), and many expressed a willingness to play the game multiple times to see different results (15 out of 23). Not many participants felt that the overall story followed a fixed storyline (only 7 out of 23), suggesting that the IDN allows players to perceive various possibilities. However, only 11 out of 23 participants believed the overall story was coherent and fluid.

For the interviews, nearly all players found the stage-by-stage approach engaging. A few players suggested enhancing the narrative progression to emphasize a sense of development: For example, the first villager segment discussing eyewitness accounts offers only simple clues. When meeting later characters, there should be deeper, layered evidence, such as photos or traces. Right now, it felt more like speaking to different NPCs and receiving varying viewpoints, but the clues lacked significant depth and connection. Participants also expressed great interest in how AI manages narrative transitions in the system, indicating they had not played a game like this before.

Regarding character development, some players felt that the AI's responses in its role-playing seemed less human-like, as it often provided humble replies even to unusual questions. Others found that the personalities of the AI characters were not distinct enough, leading to generated content that felt similar. Additionally, they noted the difficulty in having deeper discussions with the NPCs, as they often provided repetitive content. There were suggestions to better control the amount of information in the output, with some participants believing that breaking down large blocks of text into single sentences would be more suitable.

8.4 Reflection

8.4.1 Feedback from artist. After the collaboration, we interviewed the artist to gather her feedback on the experience. She expressed that Orchid-Creator brought a significant creative shift for her. While her previous works adhered to fixed narratives, with real-time AI, she could enable everyone to have unique stories rather than the same ones.

She also expressed that the move from defining every detail to establishing a narrative framework, "allows for customization and enables players to co-create the story." This aligns well with her creative intent, which focuses on encouraging players to tell their own stories. "I've always wanted to encourage player storytelling in various projects. I struggled to find a better implementation method. Orchid-Creator has made this possible." Additionally, she mentioned that the system alleviated the more challenging writing aspects of her projects, significantly aiding the development.

She reported a change in how the system influenced her overall creative process. Initially, when her ideas were less defined, Orchid-Creator 's mechanism inspired her design. However, after phase 1, once her background research and storyline were established, she felt that Orchid-Creator began to limit her creativity. It transitioned to guiding her on how to "engineer" the story. She found that the system's constraints and required components fixed many aspects of her narrative, making the "engineering" process feel more like "filling in the blanks". In later stages, she focused less on the AI's potential for spontaneity and prioritized the quality of the output, aiming to narrow down her creativity on guiding AI's responses.

8.4.2 Thinking from us researcher. Throughout the process, we observed a divergent-convergent creative process, as mentioned by the artist. It is similar to design processes found in frameworks like the Double Diamond design model [27]. During her long-term interaction with Orchid-Creator , the artist developed her own strategies and personal tricks, adjusting her content based on the system's limitations. These qualities are difficult to identify in traditional lab-based experiments. We also observed the importance of background

research for IDN construction using Orchid-Creator . Although it is a system that automatically forms narrative content, it does not stop artists from deep thinking about world-crafting. Her creative process started from solid research to craft the narrative first, and then she began to share some agency with AI.

Artist A focuses a lot on physical installation, visual design styles, and it appears that Orchid-Creator functions as just one component in the overall interactive art project. We are pleased to observe this, as creative programming tools like Processing [6] and Arduino [7] also serve as a part of design elements within an ecosystem of different media. Ultimately, the final outcome of the work is not driven by Orchid-Creator only, but rather emerges from the artist's comprehensive experience and creative habits, in conjunction with the synthesized materiality.

The collaboration also revealed a few challenges for applying Orchid-Creator in real practice. Orchid-Creator struggles to support rich interactions within the same stage, such as allowing the AI to work in multiple threads, portray different roles, or remember the author's specified sequence. Stages need to be finely divided, with clear instructions. Without multimedia support, it is difficult to connect Orchid-Creator with real interactions. While the artist viewed visual expression as essential to the overall IDN concept and hoped to see the integration of text and visuals during the creative process, Orchid-Creator could not address this need. Moreover, player input ideas for physical interaction and passively detected information were challenging to implement in Orchid-Creator , as there was no interface to test the physical interaction flow. This influenced the artist to give up the woodblock and gestural input design. The AI's generative attributes created a relationship between the artist and the system that is both supportive and limiting. This was evident as the artist might compromise on some complex ideas to ensure high-quality output, sacrificing her creative control.

We also learned from the player feedback that the stage-by-stage format encourages a sense of progression, as they expect the author to create an accumulation of clues throughout different narrative stages. Although characters are just one part of our system, players are particularly concerned about the authenticity of NPC dialogues, which in future Orchid-Creator should improve on.

9 Discussion

In this section, we first discuss the insights across the three different studies comprehensively. Next, we compare Orchid-Creator with existing IDN authoring tools and recent research on AI narrative. Finally, we highlight the limitations revealed by this study concerning LLMs, the creative ecosystem, and future design spaces.

9.1 On the three studies

In three different experiments, we explored various aspects of Orchid-Creator . Study 1 demonstrated the complexity and richness of the stories it could create, delving deeply into different design components. We found that Orchid-Creator received high recognition from experienced IDN creators in facilitating narrative structuring, especially through the combination of Global Variables and trigger conditions. This combination successfully provided more sense of control, which was limited in the previous study [55], emphasizing the advantages of our DQ1.

Study 2 focused on comparing Orchid-Creator with AI-linear authoring and non-AI branching authoring, demonstrating the enhancement in authoring interactivity and creative support. Study 3 explored how Orchid-Creator operated within a complete artistic project. It inspired the artist to create elements that extend beyond Orchid-Creator 's own capabilities, while also providing insights into player feedback.

Regarding the DQs in section 3, we identified the satisfaction associated with DQ1 (stricter trigger judgment) in both Study 1 and 2, DQ2 (flexible testing), and DQ4 (structured input variable authoring) across the three studies. Meanwhile, DQ3 (dynamic memory management) was reflected in the longitudinal study as the artists applied the features to overcome certain text embedding limitations.

In the following, we summarize the important insights observed across the three studies.

9.1.1 Individual strategy. We observed that each participant had different habits in using Orchid-Creator across the studies. Participants employed personal tricks to address the AI's limitations and to adapt to the authoring logic required by the system. For instance, in Study 1, we noted varying methods in managing the creative process. Some participants adopted a strategy of authoring all content at once before entering the testing phase to see the results. In contrast, other participants approached it like a solid construction process, working step by step, ensuring that the current stage was good enough for moving on to the next. In Study 3, artist A, throughout her long-term engagement with Orchid-Creator , developed techniques, such as using context to naturally form the world-building, and instructing the AI to assign roles to each NPC to avoid hallucination.

These are fascinating insights and exemplify a quality that we believe a creative support tool should possess. We think this personal strategy will be beneficial to bring a sense of ownership and accomplishment for users.

9.1.2 Authoring interactivity as "player co-creation". Throughout the user study, we observed that Orchid-Creator strongly encouraged users to think about how their player will interact with the narrative. Compared to AI Dungeon, which allows only natural language input, and Twine, which relies on clicks and selections, Orchid-Creator stimulated deeper thinking about how to design player interactions to influence the narrative.

Study 2 revealed that Orchid-Creator enables authors to perceive their players as co-creators actively involved in the full creative process. This was also evident in Study 3. Artist A's iterative process on interactive mechanisms (subsubsection 8.2.2) demonstrated how Orchid-Creator encouraged her to focus on conceptualizing inputs and support her artistic intentions. She actively designed interactions that provided players with creative space, with the intent of producing a co-creation experience. This insight differs from previous AI co-creation studies [18], which primarily focus on AI as a creative partner of author. Our findings highlight that the player is a crucial element of the co-creation process in systems like Orchid-Creator .

[6] https://processingfoundation.org/
[7] https://www.arduino.cc/

9.1.3 Automation v.s. Crafting. Across the studies, we observed that Orchid-Creator introduced some "crafting" behaviors. P2 and P8 from Study 1 iteratively used Orchid-Creator to refine their IDNs. Other participants also utilized the testing features to experiment with different authoring possibilities. In Study 2, statistical analysis showed that Orchid-Creator encourages idea exploration and structural experimentation more than other tools.

This demonstrates that, although automatic narrative generation is a key feature of systems like Orchid-Creator , as well as those described in subsection 2.2, it does not mean that users should "slack off" or stop thinking deeply and paying efforts. This is also supported by participants in Study 2, who preferred Twine, who expressed that crafting without automation could also yield impressive results, while unconsidered use of LLMs might make things boring.

In Study 3, we observed that the artist involved extensive background research. Furthermore, she iteratively debugged with Orchid-Creator over 20 times. This demonstrates that while Orchid-Creator incorporates AI automation, it did not hinder the artist from spending time. We recommend that future creative support tools with AI features, should ensure that automation does not diminish the creator's deep thinking. They should serve more as creative resources rather than tools for a shortcut.

9.1.4 Shifts in creative mindset. In Study 1, P3 demonstrated a need to switch his writing style, and P6 expressed caution regarding AI capabilities. These creative considerations were also reflected in the previous study [55], suggesting that when AI functions as a creative resource, human trust in it may play a crucial role. In Study 2, participants emphasized that Orchid-Creator introduced a pattern that allows for constructing an overarching narrative logic first before adding content, which offered a creative shift from previous methods. In Study 3, we observed that Orchid-Creator encouraged the artist to make a notable departure from her past practice of delivering stories. This creative shift seems to create a dynamic of both supporting and limiting between the user and the system, which has been noted in previous research as well [12, 16].

9.2 Conceptual novelty and comparison with existing works

Compared to existing node-graph authoring environments mentioned in subsection 2.4, Orchid-Creator introduces an approach that offers an integrative mindset – Users are effectively authoring an LLM, similar to prompting an AI to produce a real-time narrative, through scaffolding processes. Unlike WhatIF [35] and StoryJam [41], which use LLMs in a fragmented manner, Orchid-Creator presents LLMs' real-time usage as a cohesive entity, when ultimately delivered to players. We believe the creative method of Orchid-Creator embodies the theory of emergence [40], where "the whole is greater than the sum of its parts." As different LLM components interact with each other, guided by the user's rules, they form something more unpredictable than when the components are considered separately.

On the other hand, the encouragement of player co-creation by Orchid-Creator represents an innovation. To the best of our knowledge, this was not considered in previous AI-driven authoring tools. For instance, recent work on emergent narratives, such as the algorithmic framework proposed by Wang et al. [54], does not account for player co-creativity within the creative loop. In contrast, Orchid-Creator fosters an ecosystem where authors, LLMs, and players collaborate.

Our study also revealed insights that align with existing works. For instance, across the three studies, participants expressed positive feedback regarding our card-based authoring approach for deconstructing complex narrative elements, making them feel about tabletop strategies. This echoes the previous studies that explore metaphor to scaffold authoring [10, 11]. Moreover, this research enhances the original Orchid framework [55], and addresses the real-world practices limitations.

9.3 Limitations and future works

9.3.1 Limitations on LLMs. While DQ3 addresses the inconsistency of LLM historical memory, this research also uncovered new limitations, including coherence issues, generation speed, and character constraints that cannot be resolved through UI/UX enhancements. Our system focuses on prompt engineering and the combination of different LLMs without modifying the models themselves, such as through fine-tuning. While this approach offers flexibility for updating with the latest language models, it lacks LLMs trained on specific authoring data.

Since our system relies on different types of authoring modules, it is unsuitable for training directly with existing novels or articles. Thus, we envision that training the LLMs for Orchid-Creator will require collaboration with authors familiar with how Orchid-Creator operates, allowing them to produce multiple works while iteratively adjusting the dataset framework. This may also involve splitting datasets for specific elements, such as character relationships and input-output relationships.

Furthermore, we observed negative experience due to the model generation speed and context size. This might require a framework that is capable of making length adjustment decisions.

9.3.2 Limitations on text inputs. Across the three studies, users' authoring of interactive inputs has largely been limited to natural language text. This may be due to the difficulty in conceptualizing "Number" and "Boolean" without integrating them in a game engine. For instance, in a game, a die roll provides a numeric input that is not directly controlled by the player but passively sensed by the system. Simply presenting users with a numeric option makes it challenging for them to envision how it could play a role in IDNs.

9.3.3 Creative ecosystem. Future work should focus on multi-media support and the practical application of Orchid-Creator in real IDN game implementations, as many operations remain unknown when applied to out-of-lab IDN creations. Particularly during Study 3, the software limitations restricted creative ideas. Additionally, there may be differing needs among user groups and narrative genres.

Specifically, it remains unclear how the system handles large-scale narratives with numerous characters and complex world-building elements. These may require long-term observation, such as studying it within a full creative ecosystem, as shown in Buechley et al.'s study [7].

10 Conclusion

This paper presents Orchid-Creator , an IDN authoring tool that allows authors to define branching structures in LLM-driven IDNs. Authors use a system of cards and blueprints to define the main elements of the narrative and assemble them in a graph of branching stages. As such, Orchid-Creator allows for more coherent and enriching IDN experiences compared to existing tools. We validate Orchid-Creator through two user studies performed with IDN authors and a more general audience, respectively. These studies demonstrate that Orchid-Creator significantly improves authorial control and player co-creation, highlighting the advantages of LLMs over traditional script-based methods. A case study with a narrative artist further highlights the practical benefits and creative potential of Orchid-Creator , while also addressing the constraints faced by AI-driven tools for creativity. Overall, Orchid-Creator represents a significant advancement in the field of interactive narrative design, facilitating the crafting of complex and coherent stories.

References

[1] Alberto Alvarez. 2023. ChatGPT as a Narrative Structure Interpreter. In *Interactive Storytelling: 16th International Conference on Interactive Digital Storytelling, ICIDS 2023, Kobe, Japan, November 11–15, 2023, Proceedings, Part II* (Kobe, Japan). Springer-Verlag, Berlin, Heidelberg, 113–121. doi:10.1007/978-3-031-47658-7_9

[2] Yadagiri Annepaka and Partha Pakray. 2025. Large Language Models: a survey of their development, capabilities, and applications. *Knowledge and Information Systems* 67, 3 (2025), 2967–3022. doi:10.1007/s10115-024-02310-4

[3] Claus Atzenbeck, Sam Brooker, and Daniel Roßner. 2023. Storytelling Machines. In *Proceedings of the 6th Workshop on Human Factors in Hypertext* (Rome, Italy) *(HUMAN '23)*. Association for Computing Machinery, New York, NY, USA, Article 4, 9 pages. doi:10.1145/3603607.3613481

[4] Maren Awiszus, Frederik Schubert, and Bodo Rosenhahn. 2022. Wor (l) d-GAN: Toward Natural-Language-Based PCG in Minecraft. *IEEE Transactions on Games* 15, 2 (2022), 182–192. doi:10.1109/TG.2022.3153206

[5] Ruth Aylett. 1999. Narrative in Virtual Environments-Towards Emergent Narrative. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*. AAAI Press, Washington, DC, USA, 83–86.

[6] Kevin Michael Brooks. 1999. *Metalinear cinematic narrative: Theory, process, and tool*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[7] Leah Buechley, Mike Eisenberg, Jaime Catchen, and Ali Crockett. 2008. The LilyPad Arduino: Using Computational Textiles to Investigate Engagement, Aesthetics, and Diversity in Computer Science Education. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 423–432. doi:10.1145/1357054.1357123

[8] Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, Yujia Qin, Xin Cong, Ruobing Xie, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2024. AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors. In *The Twelfth International Conference on Learning Representations*, Vol. 2024. OpenReview.net, virtual, 20094–20136.

[9] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. doi:10.1145/2617588

[10] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. doi:10.1145/3491102.3501819

[11] John Joon Young Chung and Max Kreminski. 2024. Patchview: LLM-powered Worldbuilding with Generative Dust and Magnet Visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 77, 19 pages. doi:10.1145/3654777.3676352

[12] Nicholas Davis, Jacob Sherson, and Janet Rafner. 2025. The Co-Creative Design Framework for Hybrid Intelligence. In *Proceedings of the 2025 Conference on Creativity and Cognition (CC '25)*. Association for Computing Machinery, New York, NY, USA, 560–572. doi:10.1145/3698061.3726934

[13] DreamGen. 2025. AI Role-Play and Story Generator | DreamGen. https://dreamgen.com/ Accessed: 2026-01-25.

[14] Failbetter Games. 2024. StoryNexus-A Platform for Creating Interactive Worlds and Stories. https://www.failbettergames.com/

[15] Articy Software GmbH and Co. KG. 2024. Articy Draft: Narrative Design for Interactive Projects. https://www.articy.com/en/

[16] Alicia Guo, Shreya Sathyanarayanan, Leijie Wang, Jeffrey Heer, and Amy X. Zhang. 2025. From Pen to Prompt: How Creative Writers Integrate AI into their Writing Practice. In *Proceedings of the 2025 Conference on Creativity and Cognition (CC '25)*. Association for Computing Machinery, New York, NY, USA, 527–545. doi:10.1145/3698061.3726910

[17] N Katherine Hayles. 2001. The Transformation of Narrative and the Materiality of Hypertext. *Narrative* 9, 1 (2001), 21–39.

[18] Jessica He, Stephanie Houde, and Justin D. Weisz. 2025. Which Contributions Deserve Credit? Perceptions of Attribution in Human-AI Co-Creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 540, 18 pages. doi:10.1145/3706598.3713522

[19] David Herman. 2009. *Basic Elements of Narrative*. John Wiley & Sons, online. doi:10.1002/9781444305920

[20] Chengpeng Hu, Yunlong Zhao, and Jialin Liu. 2024. Game Generation via Large Language Models. In *2024 IEEE Conference on Games (CoG)*. IEEE, Milan, Italy, 1–4. doi:10.1109/CoG60054.2024.10645597

[21] Alexis Kennedy I.D. Millington. 2015. Varytale. https://www.ifwiki.org/Varytale

[22] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 3, 20 pages. doi:10.1145/3586183.3606737

[23] Jack Kelly, Alex Calderwood, Noah Wardrip-Fruin, and Michael Mateas. 2023. There and Back Again: Extracting Formal Domains for Controllable Neurosymbolic Story Authoring. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 19. AAAI, Washington, DC, USA, 64–74. doi:10.1609/aiide.v19i1.27502

[24] Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. Cells, Generators, and Lenses: Design Framework for Object-Oriented Interaction with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 4, 18 pages. doi:10.1145/3586183.3606833

[25] Sofia Kitromili, James Jordan, and David E. Millard. 2020. What Authors Think about Hypertext Authoring. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (Virtual Event, USA) *(HT '20)*. Association for Computing Machinery, New York, NY, USA, 9–16. doi:10.1145/3372923.3404798

[26] Chris Klimas. 2025. Twine: An Open-Source Tool for Telling Interactive, Nonlinear Stories. https://twinery.org/

[27] Magda Kochanowska, Weronika Rochacka Gagliardi, and with reference to Jonathan Ball. 2022. *The Double Diamond Model: In Pursuit of Simplicity and Flexibility*. Springer International Publishing, Cham, 19–32. doi:10.1007/978-3-030-79879-6_2

[28] Max Kreminski, Melanie Dickinson, Noah Wardrip-Fruin, and Michael Mateas. 2022. Loose Ends: A Mixed-Initiative Creative Interface for Playful Storytelling. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 18, 1 (Oct. 2022), 120–128. doi:10.1609/aiide.v18i1.21955

[29] Vikram Kumaran, Dan Carpenter, Jonathan Rowe, Bradford Mott, and James Lester. 2023. End-to-End Procedural Level Generation in Educational Games with Natural Language Instruction. In *2023 IEEE Conference on Games (CoG)*. IEEE, Boston, MA, USA, 1–8. doi:10.1109/CoG57401.2023.10333195

[30] Vikram Kumaran, Jonathan Rowe, Bradford Mott, and James Lester. 2023. Scenecraft: Automating Interactive Narrative Scene Generation in Digital Games with Large Language Models. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 19. AAAI Press, Washington, DC, USA, 86–96. doi:10.1609/aiide.v19i1.27504

[31] Mina Lee and et al. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. doi:10.1145/3613904.3642697

[32] Jialu Li, Yuanzhen Li, Neal Wadhwa, Yael Pritch, David E Jacobs, Michael Rubinstein, Mohit Bansal, and Nataniel Ruiz. 2025. Unbounded: A Generative Infinite Game of Character Life Simulation. In *The Thirteenth International Conference on Learning Representations*, Vol. 2025. OpenReview.net, virtual, 87465–87491.

[33] Max Loh. 2023. AI Roguelite. https://airoguelite.net/ Accessed: 2026-1-25.

[34] Sandy Louchart. 2007. *Emergent Narrative: Towards a Narrative Theory of Virtual Reality*. Ph. D. Dissertation. University of Salford, United Kingdom.

[35] Aditi Mishra, Frederik Brudy, Qian Zhou, George Fitzmaurice, and Fraser Anderson. 2025. WhatIF: Branched Narrative Fiction Visualization for Authoring Emergent Narratives using Large Language Models. In *Proceedings of the 2025 Conference on Creativity and Cognition (CC '25)*. Association for Computing Machinery, New York, NY, USA, 590–605. doi:10.1145/3698061.3726933

[36] Graham Nelson. 2024. Inform 7. https://ganelson.github.io/inform-website/.

[37] Britney Ngaw, Grishma Jena, João Sedoc, and Aline Normoyle. 2024. Towards Authoring Open-Ended Behaviors for Narrative Puzzle Games with Large Language Model Support. In *Proceedings of the 19th International Conference on the Foundations of Digital Games* (Worcester, MA, USA) *(FDG '24)*. Association for Computing Machinery, New York, NY, USA, Article 43, 4 pages. doi:10.1145/3649921.3656975

[38] Jeongyoon Park, Jumin Shin, Gayeon Kim, and Byung-Chull Bae. 2023. Designing a Language Model-Based Authoring Tool Prototype for Interactive Storytelling. In *International Conference on Interactive Digital Storytelling*. Springer, Cham, Cham, Switzerland, 239–245. doi:10.1007/978-3-031-47658-7_22

[39] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. doi:10.1145/3586183.3606763

[40] Stephen C Pepper. 1926. Emergence. *The Journal of Philosophy* 23, 9 (1926), 241–245.

[41] Hua Xuan Qin, Guangzhi Zhu, Mingming Fan, and Pan Hui. 2025. Toward Personalizable AI Node Graph Creative Writing Support: Insights on Preferences for Generative AI Features and Information Presentation Across Story Writing Processes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 897, 30 pages. doi:10.1145/3706598.3713569

[42] Daniel Roßner, Claus Atzenbeck, and Sam Brooker. 2023. SPORE: A Storybreaking Machine. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media* (Rome, Italy) *(HT '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 6 pages. doi:10.1145/3603163.3609075

[43] Marie-Laure Ryan, James Ruppert, and John W Bernet. 2004. *Narrative across media: The languages of storytelling*. University of Nebraska Press, Lincoln, Nebraska.

[44] Martin Schrepp, Jessica Kollmorgen, and Jörg Thomaschewski. 2023. A Comparison of SUS, UMUX-LITE, and UEQ-S. *Journal of User Experience* 18, 2 (jun 2023), 86–104.

[45] Yotam Shibolet, Noam Knoller, and Hartmut Koenitz. 2018. A Framework for Classifying and Describing Authoring Tools for Interactive Digital Narrative. In *Interactive Storytelling*. Springer International Publishing, Cham, 523–533. doi:10.1007/978-3-030-04028-4_61

[46] StoryScape. 2023. StoryScape AI. https://storyscapeai.app/ Accessed: 2026-01-25.

[47] Yuqian Sun, Hanyi Wang, Pok Man Chan, Morteza Tabibi, Yan Zhang, Huan Lu, Yuheng Chen, Chang Hee Lee, and Ali Asadipour. 2023. FictionalWorlds, Real Connections: Developing Community Storytelling Social Chatbots through LLMs. In *International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET)*. The Eurographics Association, Digital Library, 1–14. doi:10.2312/imet.20231261

[48] Nicolas Szilas, Ruud de Jong, and Mariët Theune. 2020. ExpReal: a Writing Language and System for Authoring Texts in Interactive Narrative. In *Proceedings of the 15th International Conference on the Foundations of Digital Games* (Bugibba, Malta) *(FDG '20)*. Association for Computing Machinery, New York, NY, USA, Article 12, 4 pages. doi:10.1145/3402942.3402949

[49] Kent Tessman. 2020. The Hugo Book: An Interactive Fiction Design System. https://tajmone.github.io/hugo-book/

[50] Graham Todd, Sam Earle, Muhammad Umair Nasir, Michael Cerny Green, and Julian Togelius. 2023. Level Generation Through Large Language Models. In *Proceedings of the 18th International Conference on the Foundations of Digital Games* (Lisbon, Portugal) *(FDG '23)*. Association for Computing Machinery, New York, NY, USA, Article 70, 8 pages. doi:10.1145/3582437.3587211

[51] Nicholas Sloss Treynor and Joshua McCoy. 2024. College Ruled: A Pathfinding Approach to Generative Storytelling. In *Proceedings of the 19th International Conference on the Foundations of Digital Games* (Worcester, MA, USA) *(FDG '24)*. Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages. doi:10.1145/3649921.3649994

[52] Nick Walton. 2023. AI Dungeon. https://play.aidungeon.io/

[53] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2024. Voyager: An Open-Ended Embodied Agent with Large Language Models. *Transactions on Machine Learning Research* 2024 (2024), 1–44. https://openreview.net/forum?id=ehfRiF0R3a

[54] Yi Wang, Qian Zhou, and David Ledo. 2024. StoryVerse: Towards Co-authoring Dynamic Plot with LLM-based Character Simulation via Narrative Planning. In *Proceedings of the 19th International Conference on the Foundations of Digital Games* (Worcester, MA, USA) *(FDG '24)*. Association for Computing Machinery, New York, NY, USA, Article 55, 4 pages. doi:10.1145/3649921.3656987

[55] Zhen Wu, Serkan Kumyol, Shing Yin Wong, Xiaozhu Hu, Xin Tong, and Tristan Braud. 2025. Orchid: A Creative Approach for Authoring LLM-Driven Interactive Narratives. In *Proceedings of the 2025 Conference on Creativity and Cognition (CC '25)*. Association for Computing Machinery, New York, NY, USA, 774–791. doi:10.1145/3698061.3726906

[56] Zhen Wu, Ruoyu Wen, Marcel Zaes Sagesser, Sum Yi Ma, Jiashuo Xian, Wenda Wu, and Tristan Braud. 2024. I Light U Up: Exploring a New Emergent Narrative Paradigm through Physical Data Participation in AI Generative Experiences. In *SIGGRAPH Asia 2024 Art Papers (SA '24)*. Association for Computing Machinery, New York, NY, USA, Article 22, 9 pages. doi:10.1145/3680530.3695446

[57] Zihan Yan, Chunxu Yang, Qihao Liang, and Xiang 'Anthony' Chen. 2023. XCreation: A Graph-based Crossmodal Generative Creativity Support Tool. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 48, 15 pages. doi:10.1145/3586183.3606826

[58] Qing Ru Yong and Alex Mitchell. 2023. From Playing the Story to Gaming the System: Repeat Experiences of a Large Language Model-Based Interactive Story. In *Interactive Storytelling*. Springer Nature Switzerland, Cham, 395–409. doi:10.1007/978-3-031-47655-6_24

[59] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. doi:10.1145/3490099.3511105

[60] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 30 pages. doi:10.1145/3586183.3606800