Turn story to life in 1001 Nights: A Co-Creative Text Adventure Game Using A Story Generation Model

Anonymous

How can stories we tell be turned from abstractions in our own minds to concrete elements in a digital environment we interact with? To immerse everyday storytelling into real-life contexts in digital interactions, we created a game that turns entities in a story into digital assets that have functional roles. Taking the classic folklore as inspiration, we created *1001 Nights*, a co-creative, mix-initiative storytelling game using an existing AI creative writing system. In this game, Shahrzad (driven by the player) needs to tell stories through a dialogue interface, while the King (driven by the AI model) will continue the player's story in turn. Text from the story is utilised in actual game mechanisms, so that weapon keywords in the game like "sword" or "shield" will turn into equipment that can be used for battle. This also leads to an alternative ending from the original story: Heroine Shahrzad defeats the tyrant. The game aims to facilitate player engagement and creativity through natural language interactions through an empowering story of a female protagonist who tells stories in a cultural context to accomplish her goals. We connected the story background, game mechanics, and AI systems for player engagement, and analysed instrumented gameplay data from 2056 players and comments from 422 players. The result demonstrates that players' engagement, in form of the number of inputs, significantly corresponds with their overall achievements in this game.

CCS CONCEPTS • **General and reference** ~ Cross-computing tools and techniques ~ Design • **Human-centred computing** ~ Interaction design • **Applied computing** ~ Arts and humanities
Additional Keywords and Phrases: Game Design, AI Game, Intelligent Narrative, Conversational Agent
ACM Reference Format:
Not needed on submission.

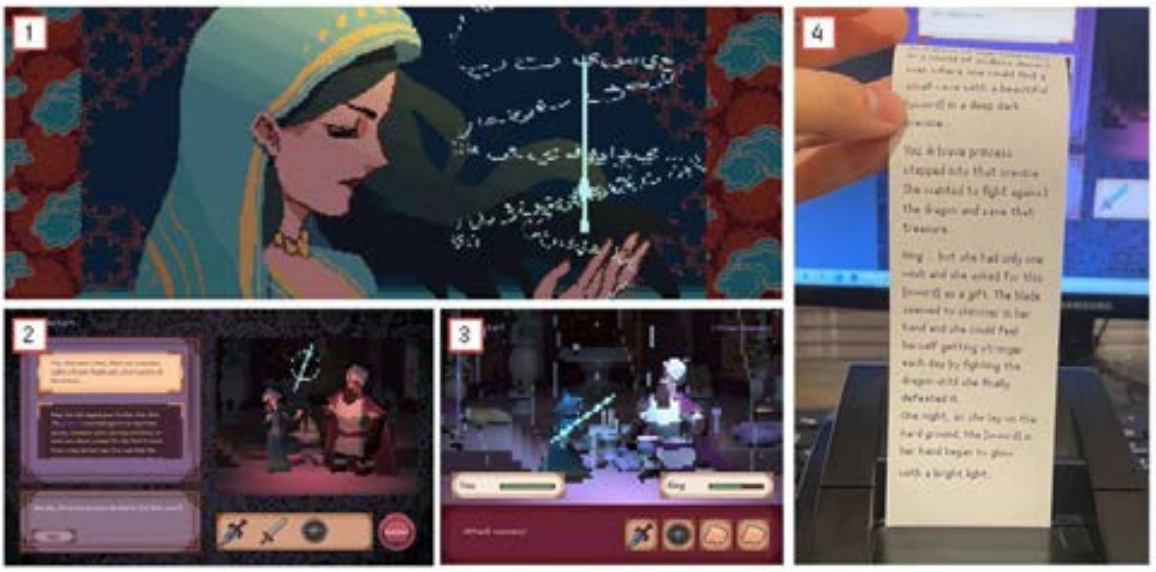

Figure 1: Shahrzad (Player character) with a magical ability to turn "language" into reality **(2)** Storytelling phase, where player writes stories with the King (driven by AI) Weapon words like "sword" can be turned into real weapons in-game **(3)** Turn-based combat phase, where the player can fight with the king in turns **(4)** Printer will print the story when a weapon word is triggered. This matches with the core concept of the game: bringing storytelling to real-life

1 INTRODUCTION

Humans are fundamentally storytellers. From advancing in careers to making a pepperoni pizza, stories infuse every part of our lives. The ability of machines to generate coherent text has changed the way stories can be told by mechanising the writing process. Can this text-based dialogue between human and machine writing be used as part of the mechanism for gameplay? To metaphorize the storytelling process, we turned to the story of Shahrzad, who determined her fate by telling stories in real life. We created a game that uses the conceit of storytelling akin to The thousand and one nights to motivate real-life writing of stories.

Inspired by classic folklore, we created the game 1001 Nights (Figure 1), a co-creative, mix-initiative storytelling game driven by an existing AI creative writing system. The core concept is "bringing storytelling to real-life" in game form: entities in storytelling are not just words and descriptions but can be turned into real assets to change the reality in a video game.

In this game, Shahrzad (driven by the player) needs to tell stories through a dialogue interface, and then the King (driven by the AI model) will continue the player's story in turns. When the King's continuation contains weapon

keywords like “sword”, “knife” or “shield”, Shahrzad can use her special ability to turn words into real weapons and use them to fight with the king, making real mechanisms from the writing of the players themselves. This also leads to an alternative ending from the original story: female storyteller and heroine Shahrzad defeats the tyrant, and averts the King’s heinous crime against herself and more citizens.

With this game, we expanded existing creative writing tools to a playable storytelling experience under a well-known story background. The authors believe combining natural language interactions under a classic story background can help players to explore and engage more in the game by expressing themselves. The efforts they put into imagination and creativity will receive positive and adaptive content generation from the AI model.

The authors showcased a Chinese version of this game in several art exhibitions and received 12030 records of story inputs from 2055 players. This paper aims to investigate if the AI system can encourage players to contribute more collaboratively through engagement. The results demonstrate that those players who are more engaged in the game are intended to reach better achievements. Besides, through comments collected from winning players (n=422), they show positive feedback towards the game from diverse perspectives, including the game mechanics, and the stories they created. Some of their feedback also shows a cultural connection through creative work: express their own interpretations of characters in the folklore, and are supported to include characters and plots from their own cultural backgrounds.

2 RELATED WORKS

2.1 Natural Language Processing

Previous research studies investigated the use of Natural Language Processing (NLP) with different focuses, including creative tools [1,2,3]. Some projects developed specific genre-focused collaborative AI writers like Shelly [4], a crowd-sourced horror writer.

Similar approaches were employed in academic research for content generation. Murder mystery generation[5] generated murder mysteries for adventure games, using structured information of real-world people mined from Wikipedia articles. Designing for Narrative Influence [6] trained a language model to generate micro-fiction that promotes sustainable public health guidelines. Martin et al [7] present a series of experiments that connect ancient procedural techniques to modern technologies like the language generation models.

Related research studies facilitate different focuses through dialogue systems. Sheharazad Tavern [8] and Prom Week [9] tried to develop deeper NPC interactions for a natural social simulation experience. Talk to Ghost [10] adapted Shakespear's work to improve high school student's interest in reading by turning stories into interactive conversations with virtual characters.

This study is proposing a hybrid experience that sits between creative writing and game. One of our main references is CharacterChat [11] and BanterBot [12], dialogue systems that allow writers to talk with the characters they created. This extends writing assistance to an intelligent agent that turns interaction context into a more familiar social scene. The authors designed the dialogue interface in 1001 Nights as a special scenario: two people telling stories with each other. Through this, we hope players can easily understand the narrative context, and overcome a barrier to creativity documented in [13]: fear of the blank canvas.

2.2 Game Interactions Around NLP

Through the development of NLP, emerging text adventure games give a player more control over games without fixed choices in traditional games. For instance, Interview With The Whisperer [14] and Mystery Of Three Bots [15] set mysterious plots to let players explore the story with natural language text input through Semantic ML, a tool for semantic analysing developed by Google [16]. Fraser et. al. [17] present different studies on open-domain social conversational AI using emotion detection. In recent years, experimental games like AI Dungeon [18] even allow players to fully generate their text adventure with natural language input. The main goal of such games is to enhance the game playing experience by providing an immersive and engaging experience, similar to Sali’s work [19], which has shown that natural language interfaces, while difficult to use, can reward players with high levels of engagement and enjoyment.

By creating conversational interactions through free-input dialogue systems, NLP has also been used in part of commercial games like KuileiXi's drama [20], “Event [0]” [21] and Bot Colony [22]. These games introduce a natural dialogue system as the primary approach to push the storyline forward in the games.

2.3 Cultural Story

The original folklore, *The thousand and one nights*, has a worldwide influence to capture the imagination of numerous audiences. It has never ceased to inspire creators globally [23]. Meanwhile, its history is quite symmetric to the current mode of AI-human collaboration as a combination of different knowledge backgrounds and interpretations. It is the result of a "cultural and ethnic melting process" [24], in which Indian and Persian elements blend, and even including Greek, Egyptian and more.

We got inspiration from *The thousand and one nights* from both the multicultural side and narratology side. It is a classic example of a framed story and embedded narrative: the character narrates a set of tales, and tales contain other

tales. This format inspires the core mechanic of this AI-based game: It is a story capable of integrating other stories whatever their plot may be [24].

Another motivation for us to adapt this story is reinterpreting the previous ending of this female narrator and heroine. In the previous end, the king truly fell in love with her, but we think her power of telling stories can lead to another fate, that is a revolt against the heinous King.

2.4 Natural Language Generation In Stories

Among these NLP-driven projects, only several are using the natural language generation(NLG) model like personaChat [25] in CharacterChat [11] or OpenAI's GPT-3 in AI Dungeon [18], and others implemented NLP for a specific task like parsing player utterances into logical statements [8], [26], or finding the closest response from database[14,15]. The main reason is that the NLG model faces the risk of generating content that is out of topic, and these projects need to find a balance between player freedom and content quality. Accordingly, even when players can use natural language input, these games set very fixed storylines and backgrounds that cannot be intentionally changed by players' inputs. Off-topic input will either lead to confusing responses, which is frequently discussed in the player community of AI Dungeon [27]–[29] or get limited by the customised module, like Facade's Global Context Pool [30], which tried to maintain players' suspension of disbelief(belief in the fictional story for the sake of enjoyment) even when their input is out of domains. Another example is, in Scheherazade tavern [8], when a player mentions a topic that is not in the knowledge module, the AI agent will repeat a word and then transfer the topic. These control mechanisms are methods to ensure output quality.

This led us to think about the possibility of letting players decide the plot inside the game story and making mechanics corresponding to the narrative framework of the game story. We assume the natural language model's creative ability has space to improve under the game environment: when it can generate adaptive content according to player's input, and let players' choices define part of the story directions(equipment, scenes, etc), the full gameplay will become more dynamic, and bring co-authored creative artefact in the same time.

2.5 Dynamic Feedback Beyond Text

Games based on NLP are hard to give dynamic feedback beyond the text. Text content can become adaptive to the player's natural language input, but others are not. It's very time consuming for creators to prepare corresponding assets (character animation, scenes, etc) that are synchronous with text output. For instance, *Facade*[30] took two authors two years to prepare all character reactions and assets for a 20 minutes game in a single scene [30]. There are some projects that started to use other AI generation models to provide adaptive content, like the GAN-generated images in the AI Dungeon [18].

Our focus is still on text modality, but our contribution is to map text to other parts of the game world: equipment. The authors were inspired by word-typing [31,32] games, where players quickly type required words to release character skills. The difference is, that word-typing games do not bring semantic relationships between the words and skills or world environment, and in contrast, keywords in 1001 Nights will always be part of stories and created through human-AI interactions.

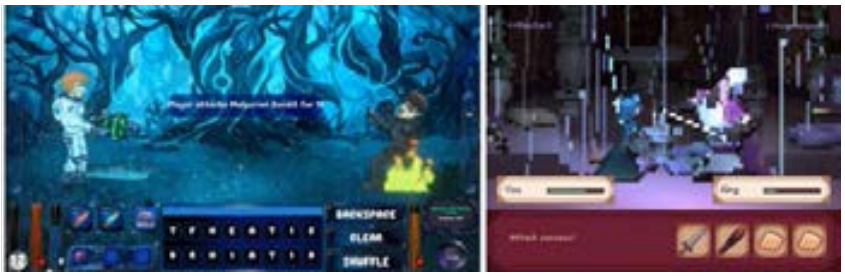

Figure 2 - Left: Screenshot of Orbi's Chronicle. Players need to type words to release skills, Right: Screenshot of 1001 Nights. Players' weapons came from the stories they wrote together with the King

We designed game mechanics to bring a rich AI system: understanding the player's actions and responding intelligently, through which a player can attempt many different strategies in the game and find them equally supported by the system [33]. Accordingly, with weapon words as the main target, we can map infinite creations from players to limited instances, and then we can provide dynamic interactions with prepared assets, including 3D models and visual effects when a keyword is triggered.

3 GAME DESIGN

This section presents the design and development of this game.

Figure 3: Storyboard of the gameplay.

Figure 4: Gameplay process.

The game includes two parts: storytelling and battle. Figure 1 illustrates the game mechanics: the player needs to keep telling stories to lead the King to produce story continuations that include the important items for battle. In the first phase, Shahrzad (driven by the player) needs to tell stories, and then the King (driven by the AI model) will continue the player's story in turns. In this game, Shahrzad has a special ability to turn words into reality: When other's words contain keywords like “sword”, “knife” or “shield”, those items will materialise and drop. The Player's target in this phase is to lead Sasanian King to tell more stories that contain keywords and collect weapons through them.

After collecting enough weapons and armours, the player can enter the turn-based battle part to fight with the King. In the second turn-based combat(battle) phase, players can use the collected weapons from the last phase to fight the King. The Player's target in this phase is to beat the flagitious King and free Shahrzad. This is a different ending from the original folklore.

In summary, we try to put all components into a coherent experience. The story background links to the mechanisms: Shahrzad needs to create stories to survive. Mechanics are centred on, also realised by the AI system that makes the story explorable.

3.1 Game Art

To make players focused on the gameplay itself, we apply a pixel-art style. Most pixel art games, like Terraria [34] and Red String Club[35] use 2D hand-drawing for all game assets. However, to save time while keeping the fluent aesthetic, the authors used 3D-to-2D techniques like A Short Hike [36]. With this method, there is no need to manually draw character animations, but instead, open-sourced action animation from open source libraries like Mixamo [37] was utilised. The final game art is still in the traditional 2D pixel-game style but reduced the development time. The low-resolution art that blurs character faces also matches the ancient story background and will leave imagination space for players.

3.2 Battle system

The aims of the battle system are: (1) To make the game interesting and challenging and encourage players to write more stories. (2) To balance the difficulty. A player should not feel too easy or too hard to win. For these purposes, the authors decided to demand players to write at least two stories that trigger valid responses to win this game, which means to win this game, a player needs two attack weapons.

However, if two attacks are enough to defeat the King, a player will only click twice to win the battle. Hence, the time of this phase will be too short. The battle mode is designed to be exciting for players, and they may meet failures before the victory. Thus the authors implemented the following rules:

1. Each weapon collected in phase one can only be used once. Once used, this weapon will be replaced by a fist (punch) icon. The fist can also be used for attack, but it has much less hurt. It's reasonable that Shahrzad can only attack by a punch when she has no weapon in her hands.
2. Some equipment is for defence rather than attack, like "shield", and "armour". They cannot hurt the king but will rich the game experience by preventing hurt from the king in his turn. The effect also corresponds to their actual use of them in reality.
3. The max of inventory space is four. Equipment exceeding inventory will not be saved for battle but will be recorded in the backend to confirm this player's achievement of getting weapons. If the inventory is not full, the blank spots will be replaced by the first in the battle phase.

By these rules, once players collect two attacking weapons (like "sword" and "dagger"), they can use them and punch them with their fist twice. These combinations will lead to victory. If players only have defending equipment, they can not defeat the King, since the only attack method is a punch.

Figure 5: Weapon inventory

3.3 AI DEVELOPMENT

3.3.1 Demo 1 and keywords

We designed four attack weapons: sword, knife, blade, dagger and two defending weapons: shield and armour. These are keywords that raise more than 50 matching paragraphs when we searched the word in the Guttenburg Short Story collection[38], a dataset extracted from the portal of Project Guttenberg, a library of over 60,000 eBooks. Meanwhile, these are also weapons words that match with 1001 Nights' ancient narrative background.

Within the original Guttenberg short story dataset(94MB)[38], we located each paragraph with a matched keyword and extracted the previous five paragraphs, then we get a subset dataset of keywords in 16.24MB. In the first prototype, the King's AI model is trained on the extracted dataset through GPT-2 medium model[39] with six thousand steps, so it is familiar with common tropes in various adventure stories. The text generation temperature is set to 0.8. Here is an example of the result in two hundred words:

You: The princess tried to find her magic sword,
AI: ... but finding it in the lap of a child, the headmaster led her back into the stable, where she found that it had been stolen, for the little thief was the child of the schoolmaster.

This model is not effective for a game, so the authors switch to another solution.

3.3.2 English and Chinese Demo 2

Although we trained a model in demo 1, the story-continuation ability is not flexible for modern players. Since the original story collection includes many vintage books, when players living in 21 century are asked to write a story, they may do it in other styles, and get less chance to be "understood" by the AI model. This will result in low-quality and confusing stories. To fix that, we need a new AI model that is inclusive of the current writing style.

We used Dreamily.ai[2] to reinforce our game design. Dreamily.ai is a creative writing platform using a modified transformer (a self-attention multi-layer neural network) model trained with quality fiction. Both English and Chinese dataset consists of open access fan-fiction and ebooks on the web, and both sizes are about 100GB. This platform has over one million Chinese version users and 200 thousand English version users. Although their model is not ubiquitous for multiple tasks(e.g. writing official documents or codes), it's very suitable for story generation.

To use the model, we only need to call Application Programming Interface (API) and send the title and prompt to dreamily.ai's server. We did prompt engineering work to make sure the King actually "knows" what age and world he lives in, and enhance the possibility to let him mention weapon words. Here is the prompt we used:

> *Fantasy: Ancient sword and shield*
> *This is the start of an ancient mystery middle eastern story. A story about a sword, shield, knife, blade, dagger, armour and magic wand.*

With this design, dreamily is easy to produce flexible stories that closely correspond to players' input. The past five inputs will be added to the prompt to ensure fluency.

It's allowed to mention keywords(like a sword) in the input phase. It may enhance the chance to get the stories containing corresponding weapons in some way, but it will not ensure that:

> *You: Betty is the best assassin in the world. but people barely know her. One day, she was attacked, then she pulled out her* ***sword*** *...*
> *King: ...and killed three enemies in a row. That's how she got her name; a legend among the common folk. Betty Cooper, the best in the land...*

In contrast, a good input that creates suitable context without any weapon word can also lead to a valid response that contains keywords:

> *You: Betty is the best assassin in this world. Each time after a battle, she would look at her inventory...*
> *King: The things in her inventory were her weapons, magic or magic-related.The most powerful weapon that could be used against her was the* ***[sword]*** *called shaled, its* ***[blade]*** *is made of steel with a gem embedded to form...*

This game takes the record of the past five turns as part of the prompt to keep stories fluent. Records will be refreshed when players start a new game or go to the battle. When King's response doesn't include a keyword, Shrazad will send a notification message to hint the player to tell a more related story. This helps to relieve confusion and the Tale-Spin effect[40], in which a system makes people feel less intelligent than it is due to the insufficient explanation of the underlying processes.

4 EVALUATION STUDY

4.1 Experimental setup

The authors were invited to showcase their work at three exhibition sites in China, to research different players to collect various feedback. All three are in different cities in China. All of them shared the same digital and analogue setups, including a vintage monitor (to match the ancient story background), a printer, and a workstation using Windows 10 system. A tutorial leaflet (screenshot of the help page in the game) is on the table for players.

Figure 6: Left: Exhibition setup Right: Tutorial leaflet for players

A mini-printer was used during a two-month off-line exhibition in Beijing to emphasise the concept of "invading language" and to improve public engagement. Each time when a keyword is triggered, the current piece of the story will be printed out. In this game, the keywords are the materialised language that becomes part of the "reality", and to players, the printed text are stories that are tangible output from the game to the real world. This function encouraged people to spend more time at the exhibition since they can keep a hard copy of their stories.

4.2 Opening and tutorial

All players are informed about data collection for research use before they enter the tutorial. In the tutorial, players will be informed about the game dynamics, for instance, the click and collect mechanism using displayed keywords: sword and shield. Not all valid keywords were revealed to players. One reason is to let them focus on a more specific instruction: write about sword and shield. Another reason is to encourage exploration. Trying out valid keywords is also part of the gameplay.

Figure 7: Left: Opening screen Right: Screenshot of tutorial

5 RESULTS AND FINDINGS

Collaborative storytelling is a product of adequate interactions and the quality of the contributions among authors. In this game, the time spent between the player (Shahrzad) and AI agent (King) in developing a meaningful outcome is identified by the number of interactions (inputs from a player) and conversation quality (thematic analysis) of stories exchanged. This paper focuses on the impact of the former as a measure of engagement on achievements made by players via a between group design. The gameplay is not limited to one round, hence, players' average engagement per play was calculated. Three groups were identified based on the achievements made as follows: non-winner (G1, n= 299), journeyman (G2, n=1106), and winner (G3, n=650). The achievements made by players cannot be predicted or controlled, hence, the groups vary in size with less players in the non-winner category compared with the other two groups. These groups are independent of each other. Players who defeated the king in one or more games are categorised as winners. Journeymen got at least one weapon but didn't win, so they are in the middle stage towards victory. Non-winners have not collected any weapons or won the game. Table 1 highlights the differences between groups.

Table 1: grouping players by achievements made

Definition	G1	G2	G3

Collected at least one weapon	No	Yes	Yes
Defeated the King more than once	No	No	Yes

All players had a chance to familiarise themselves with the game with printed (screenshots) and integrated tutorials prior to the game with identical experimental setup in all three locations. The number of players playing this game and the duration of each round was not limited to observing extreme cases and creating a real-world experience of piloting games in exhibitions. For instance, on a busy day players had less time to play while some had enough time to exceed the expected average number of inputs per play in the winners and journeyman groups (G2 and G3) as highlighted in Fig. 10.

5.1 Analysis

This study aims to investigate the impacts of engagement in storytelling on overall achievements made by players and to understand any potential trends between the groups. Hence, a Levene test is used to check the homogeneity of variances among groups, $F\ (2, 2052) = 32.02$, $p < .05$. The authors believe that players that make little progress on average are more likely to experience frustration, meaning that the distribution of inputs per play in G1 ($M = 2.59$, $SD = 2.05$) is more influenced by players' propensity to lose patience. Meanwhile, the distribution of inputs per play in G2 ($M = 4.51$, $SD = 3.61$) and G3 ($M = 6.30$, $SD = 3.85$) may also be influenced by players' luck in finding the right inputs. This distinction may explain the difference in variances. Figure 8 highlights the distributions in each group.

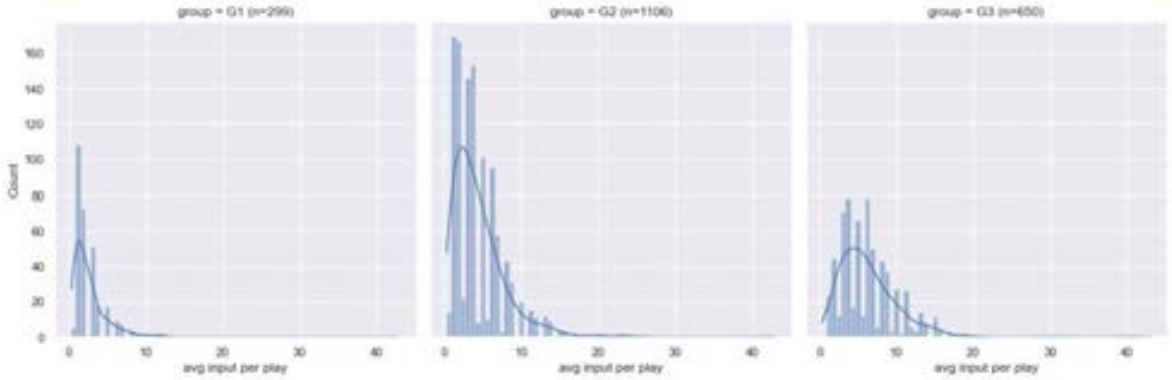

Figure 8: average input per play distributions per group

A non-parametric Kruskal-Wallis test as an alternative to one-way ANOVA is used to evaluate the impact of engagement in collaborative writings via conversations on overall achievements. Players' achievements in this game were reported to be affected, $H(2) = 328.295$, $p < .05$. The results show it is likely to achieve a better outcome by more contributions to the storyline. Also, a positive trend is observed and reported by Jonckheere-Terpstra. Since the shape and variability assumption is violated, the obtained Welch's adjusted F ratio was used $F(2, 1012.54) = 191.85$, $p < .001$. Hence, the authors can conclude that at least two of the three groups differ significantly on their overall achievements in this game.

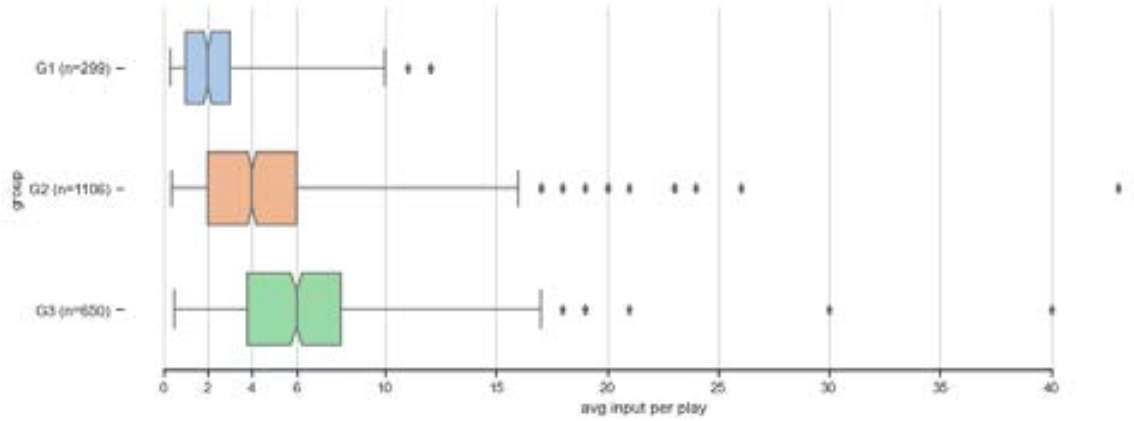

Figure 9: Positive trend in overall achievements by average contributions made

5.2 Comments from winners

Since this game was only exhibited in China, although it received some feedback in English, the following section will only focus on Chinese ones. Only players in G3 (winners) can leave feedback after victory for the following reasons:

1. Players play this game during an exhibition, so not all of them have enough interest to reach the end. Sometimes there was a queue for playing this game.
2. We want to encourage players who were defeated to try again until they achieve victory, which means they went through the full gameplay. If we show this page to all players, including ones who were defeated by the King, they may regard it as an ending and leave.
3. The authors assume players who were patient enough to win gained a deeper experience in this game, which will be helpful for us to identify our weaknesses.

We received winners' records (n=650) and removed those who didn't leave comments (n=226). We also removed two records from players who met technical difficulties during the experience (the printer is not working). Finally, we manually categorised the rest of the remaining feedback (n=422) into nine categories.

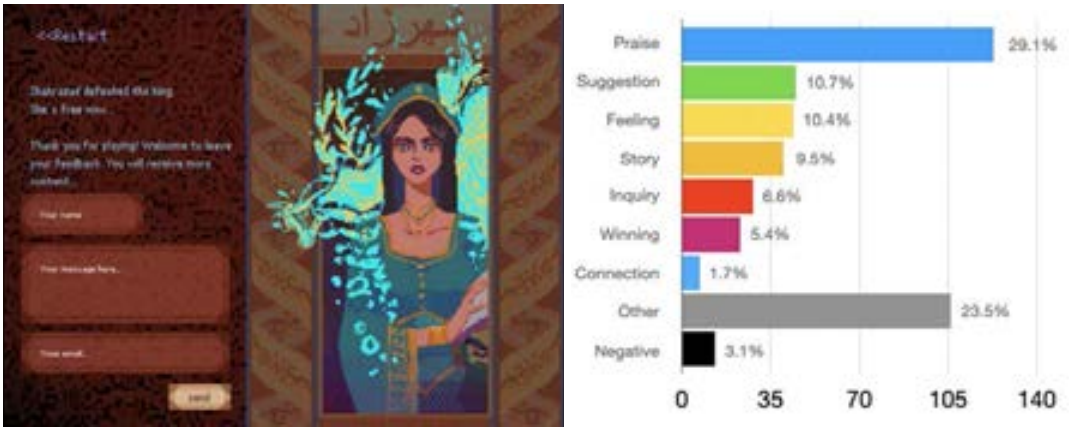

Figure 10: Feedback page appears when a player achieves victory (left), categorised feedback (right)

Table 2: Comments categories

Category	Description	Example	Number
Praise	General positive comments toward the game	Great game!/ I enjoy it	123 (29.1%)
Suggestion	Suggestions about gameplay	I think it should include more weapons/ hope it to be easier	45 (10.7%)
Feeling	Specific impression and feeling towards the gameplay, character, etc	This king is smarter than me/ It's an adult version of the folklore/ Feels like I'm teaching this AI	44(10.4%)
Story	Retell or comment specifically on the stories they created	The king ate too much deer meat and died in A's sword / A brings B to beat the king and get revenge!	40 (9.5%)
Inquiries	Ask questions about the game	Will it be published on Steam(a game platform)? / How can I get more weapons?	28 (6.6%)
Winning	Comments on their victory	I'm a king among kings!/ So easy, nobody can be my rival!	23 (5.4%)
Connection	Connected personal experiences outside the game	I hope I can become Shahrzad in real life and fight for freedom/ It reminds me of my memory of writing stories with friends when I was in high school.	7 (1.7%)
Criticism	Points out unsatisfied points	This is a bad game/ The king is too stupid.	13 (3.1%)
Others	Other unclassifiable comments	Oh yeah!!/ But he indeed lost language skills	99(23.5%)

5.2.1 General

Positive feedback like "*Good game* (P10)" or "*Interesting* (P60)" are in the praise category. This type takes up the largest amount of results (29.1%, n=123). Feedback in other categories shows various focuses. 10.7% (n= 45) of players leave their suggestions. Together with 6.6%(n=28) of players who leave inquiries, 8 of them express willingness for further development and publishing on a commercial game platform like Steam. These comments made us notice several

perspectives that we ignored before. About half of the players(n=25) who leave suggestions, and some from the inquiry category (n=7) asked to include more weapons. Although some of them(n=5) mentioned weapons like guns that do not fit into the story background, we do realise the requirement to improve weapon choices and better player guides.

Several players (n=13, 3.1%) are unsatisfied with or disliked this game, and it's reasonable for this early stage game.

5.2.2 Immerse in the game and story

10.4%(n=44)players shared personal feelings towards gameplay, like "*we should always believe in love and magic* (P326)", "*It's interesting and immersive. Players are invited to save a character and feel strength* (P511)" and "*we choose free rather than love* (P261)". Some also express thoughts about AI: "*Humans reach consensus with AI*(P586)"," *Humans are those who think beyond AI* (P561)" These suggest our game can provoke their reflection about freedom, strength and AI's agency. P624 gave a good summary that matches our motivation: "*It's interesting. The game mechanics of hidden triggers also brought 'freedom' to players, not only to Shahrzad.*"

The more interesting fact is, nearly one-third of this group(n=14) express their impression about the king, like "*This king sounds like a gastronome* (P564)" and "*The king can become a good writer in his next life* (P542)". A player even said, "*There is not only betrayal and injury but also warmth and protection, in the hope that the defeated King in prison can understand what he has, treasure what he has, do not ask the past* (P491)". Even though we didn't add any personal line to the king, some players show empathy toward this character. To some degree, this feedback shows the potential attraction of intelligent characters driven by NLP technology. A character can give reasonable responses even without detailed design work, and the player's interpretation can fill the gap in related stories.

Players' feedback is more varied than our assumption. We expected to receive general praise and suggestions, and the most surprising result is that 9.5% (n=40) of players talked specifically about the stories they created. Most of them (n=29) mentioned the character they include in their stories and described the plot in detail, like" *Summon the beasts' success! ! The black cat is turned to the witch, and it turns into a magic hat* (P395)" and "*Princess Li finally defeated the evil emperor with high ideal* (P230)". This provides evidence that many players are highly indulged in the stories they created.

While some players are immersed in the stories they created, some players receive more pleasure from their victory. 5.4% (n=23) of players show highly positive feedback describing their feelings of victory, like "*AI cannot defeat human Shahrzad!* (P30)" or "*I am very smart! I'm the smartest princess*(P138)". In general, this feedback suggests our game can bring both an entertaining experience and creative collaboration between humans and AI.

5.2.3 Cultural connection

Among players who comments on the story they created, many of them show inspiring component from personal interests that match with the mysterious background, like"*I want to lit the fire of renaissance in the darkness*"(P234), "*No matter what, Sword Soul, Shield Sprite and Gun Sprite will always be good friends!* (P637)" and "*Mountain·Boots Puss and Iliad, Hit, the three live together forever and inherit the throne of Snow Mountain.*（p540）"

Since the testing was done in exhibitions in China, some of the players put aspects of their cultural background into their stories, which became creative artefacts through human-AI collaboration that show possibilities in culture blends. Like "*Awesome! How to play the sequel? I want to chat more with the old ancestor Ye who fought with the shovel in Luoyang and the witcher who fought with the lich. . . What happened to the Prince*？"(P350) "Shovel in Luoyang" here is one of the most important tools in Chinese archaeology, and usually mentioned in grave robbing fictions. Another player(P148) put a character from pop culture in the story: "*A Liang, the youth who left the factory, can beat the king* ." This character "A Liang" comes from the pop song "About Life" by the "Wutiaoren"[41], a popular band in China that famous for their attention to the current situation of Chinese rural youth and strong humanistic feelings in their music.

5.2.4 Reflection on reality

It was observed by the authors that a few number of players (n=7, 1.7%) even connect this game to their experiences and feeling in real life, like someone who feel the encouragement from Shahrzad: "*I am a student, I also want to be free, be as brave as the heroine in-game once* (P451)" and "*I love this world, I also want to create valuable works.*"(P46)

We also received very detailed feedback about previous personal experiences: "*Thank you for reminding me of my favourite game I played with my friends in class during my reading time. In those days we used to write a story on a large piece of scratch paper, one at a time. No one knows what will happen next, and we tend to avoid stories that fall into a rut, creating more and more mysterious adventures for it. Good memories. That's a good game.* (P11)" This feedback suggests this game may have the potential ability as an educational game for story writing.

Figure 11: Exhibition photos. Left: A 11-year old boy kept playing half an hour Right: Official photo from exhibition

6 DISCUSSION

This study investigates how the AI system can reward players to motivate collaboration in writing stories. The results show a significant difference between at least two groups' overall achievements based on the level of contributions made in storytelling. - players with higher engagement are more likely to reach much improved achievements in this game. This is aligned with the hypothesis that our game design did encourage players to explore and collaboratively interact in the game, and the AI system can reward them with creative feedback.

Besides, even when we did not ask players to rank their experience, in comments from players(n=422) who won the game, players show a high level of enjoyment and interest throughout the game, where they contribute their own stories to be part of it. Supported by the AI system, the same character and interface may bring different stories and experiences based on the players' personal choices, and they can immerse themselves in the game, exploring their own interests. Similar to previous research studies[42], the unexpected but logical text generated by AI may make the story more exciting than the player's intention.

Players express their own interpretations of characters in the folklore and are supported to include characters and plots from their own cultural backgrounds. For them, the king can be a coward, a peace lover or a gastronome, and these are reflections through the creation, rather than the line the creators set. They can introduce a character from a pop song in their story, or link the game to current social news. This suggests a potential chance to alleviate the creator's burden to develop games. Players' autonomy and imagination may fill in the gap that developers leave blank. Similar to the finding from Aljammaz et al.[8] that a player may view the repeated response as the NPC's own personality.

However, the creators faced common barriers in developing 1001 Night. Like similar studies in dialogue interface, we found open dialogue systems are a double edge sword. It contributes to a sense of freedom, but still, faces the risk of out-of-track, and to limit that requires a large amount of authoring and design work. Even though we selected a group of weapon keywords that work fine in the current scenario, we can not predict a player's input to ensure an ideal response from the AI system. However, since weapons are fundamental components of game mechanics for this role-playing type of interaction, current responses may capture the most fundamental expected interactions for players in this genre of games. The creators cannot go through all possible weapons in the world, and some players may lose interest in playing this game if they are not able to collect all weapons that can be triggered, like "spears".

Furthermore, the current AI system is not able to evaluate the quality of player input.Consequently, we can only evaluate player performance through time spent (number of inputs) and text length, but the quality of inputs and responses received from the King are important parameters to be analysed in future. The authors aim to explore the text contents using sentiment analysis to understand how players performed in the story writing. Meanwhile, when player input includes some components that do not fit well in *The thousand and one nights* background (like "computer" or "riffle gun"), the king can still continue the story, which may make him out of immersion in character by AI, since an ancient king should not know about modern technologies.

Overall, as a game in its early stage, we received encouraging results. We started with the concept of "bringing storytelling to real life", and surprisingly to see that many players can naturally blend their own life into this game. Authors also suggest future investigation on using NLP models like OpenAI's GPT-3 [43] in more storytelling games.

8 CONCLUSION AND FUTURE WORKS

To extend storytelling to real-life contexts beyond the language, we created game *1001 Nights*, a co-creative storytelling game that leverages story writing into actual game mechanisms, based on a current story generation model. This paper presents how *1001 Nights* facilitate player engagement and creativity through natural language interactions under a

well-known folklore background. Our data suggest higher player engagement generally leads to better achievements in the game which demands further investigations.

Future works include semantic analysis based on the data to see what and how would people write with an AI agent under a narrative context, and ask them to evaluate their engagement and interest. This is key to evaluating the quality of inputs via a hybrid approach using NLP algorithms like text perplexity and domain experts.

REFERENCE

[1] R. Swanson and A. S. Gordon, 'Say Anything: A Massively Collaborative Open Domain Story Writing Companion', in *Interactive Storytelling*, Berlin, Heidelberg, 2008, pp. 32–40. doi: 10.1007/978-3-540-89454-4_5.

[2] 'Dreamily-beta'. https://dreamily.ai/#/ (accessed Oct. 09, 2021).

[3] A. Coenen, L. Davis, D. Ippolito, E. Reif, and A. Yuan, 'Wordcraft: a Human-AI Collaborative Editor for Story Writing', *ArXiv210707430 Cs*, Jul. 2021, Accessed: Nov. 02, 2021. [Online]. Available: http://arxiv.org/abs/2107.07430

[4] P. Yanardag, M. Cebrian, and I. Rahwan, 'Shelley: A Crowd-sourced Collaborative Horror Writer', in *Creativity and Cognition*, New York, NY, USA, Jun. 2021, pp. 1–8. doi: 10.1145/3450741.3465251.

[5] G. A. B. Barros, A. Liapis, and J. Togelius, 'Murder Mystery Generation from Open Data', p. 8.

[6] R. Lc and D. Mizuno, 'Designing for Narrative Influence:: Speculative Storytelling for Social Good in Times of Public Health and Climate Crises', in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan, May 2021, pp. 1–13. doi: 10.1145/3411763.3450373.

[7] M. Pichlmair and C. Putney, 'Procedural Generation for Divination and Inspiration', in *International Conference on the Foundations of Digital Games*, New York, NY, USA, 2020, pp. 1–7. doi: 10.1145/3402942.3402950.

[8] R. Aljammaz, E. Oliver, J. Whitehead, and M. Mateas, 'Scheherazade’s Tavern: A Prototype For Deeper NPC Interactions', in *International Conference on the Foundations of Digital Games*, New York, NY, USA, Sep. 2020, pp. 1–9. doi: 10.1145/3402942.3402984.

[9] J. McCoy, M. Treanor, B. Samuel, A. A. Reed, M. Mateas, and N. Wardrip-Fruin, 'Prom Week: Designing past the game/story dilemma', p. 8.

[10] D. Jackson and A. Latham, 'Talk to The Ghost: The Storybox methodology for faster development of storytelling chatbots', *Expert Syst. Appl.*, vol. 190, p. 116223, Mar. 2022, doi: 10.1016/j.eswa.2021.116223.

[11] O. Schmitt and D. Buschek, 'CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation with a Chatbot', in *Creativity and Cognition*, Virtual Event Italy, Jun. 2021, pp. 1–10. doi: 10.1145/3450741.3465253.

[12] 'Banter Bot by Google Creative Lab - Experiments with Google'. https://experiments.withgoogle.com/banter-bot (accessed Mar. 27, 2022).

[13] M. Kreminski and N. Wardrip-Fruin, 'Generative games as storytelling partners', in *Proceedings of the 14th International Conference on the Foundations of Digital Games*, San Luis Obispo California USA, Aug. 2019, pp. 1–8. doi: 10.1145/3337722.3341861.

[14] 'Interview with the Whisperer by Deconstructeam', *itch.io*. https://deconstructeam.itch.io/interview-with-the-whisperer (accessed Oct. 08, 2021).

[15] 'The Mystery of the Three Bots'. https://google.github.io/mysteryofthreebots/ (accessed Nov. 03, 2020).

[16] 'making_with_ml/semantic_ml at master · google/making_with_ml', *GitHub*. https://github.com/google/making_with_ml (accessed Oct. 21, 2021).

[17] J. Fraser, I. Papaioannou, and O. Lemon, 'Spoken Conversational AI in Video Games: Emotional Dialogue Management Increases User Engagement', in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, New York, NY, USA, Nov. 2018, pp. 179–184. doi: 10.1145/3267851.3267896.

[18] 'AI Dungeon'. https://play.aidungeon.io/main/home (accessed Oct. 21, 2021).

[19] S. Sali *et al.*, 'Playing with words: from intuition to evaluation of game dialogue interfaces', Jun. 2010, pp. 179–186. doi: 10.1145/1822348.1822372.

[20] Y. Xi *et al.*, 'KuiLeiXi: a Chinese Open-Ended Text Adventure Game', in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, Online, Aug. 2021, pp. 175–184. doi: 10.18653/v1/2021.acl-demo.21.

[21] 'Event[0] on Steam'. https://store.steampowered.com/app/470260/Event0/ (accessed Oct. 27, 2020).

[22] 'Steam 上的 Bot Colony'. https://store.steampowered.com/app/263040/Bot_Colony/ (accessed Mar. 27, 2022).

[23] K. Kobzošová, 'The Changing Value of The Thousand and one Nights and its Influence on Modern and Contemporary Arabic Literature', p. 16.

[24] S. Enderwitz, 'Shahrazâd Is One of Us: Practical Narrative, Theoretical Discussion, and Feminist Discourse', *Marvels Tales*, vol. 18, no. 2, pp. 187–200, 2004, Accessed: Apr. 06, 2022. [Online]. Available: https://www.jstor.org/stable/41388707

[25] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, 'Personalizing Dialogue Agents: I have a dog, do you have pets too?', *ArXiv180107243 Cs*, Sep. 2018, Accessed: Apr. 07, 2022. [Online]. Available: http://arxiv.org/abs/1801.07243

[26] S. O. Villar, *SHRDLU*. 2022. Accessed: Mar. 27, 2022. [Online]. Available: https://github.com/santiontanon/SHRDLU

[27] Realistic_Net_1799, 'No comment, just some goofy AI stuff.', *r/AIDungeon*, Apr. 04, 2022. www.reddit.com/r/AIDungeon/comments/twbpvw/no_comment_just_some_goofy_ai_stuff/ (accessed Apr. 04, 2022).

[28] TrovianIcyLucario, ‘A.I Dungeon nonsense I’ve collected: Part 2.’, *r/AIDungeon*, Apr. 03, 2021. www.reddit.com/r/AIDungeon/comments/mj11bt/ai_dungeon_nonsense_ive_collected_part_2/ (accessed Apr. 04, 2022).
[29] Wank_my_Butt, ‘New player. All of my stories derail into nonsense almost immediately. Is this normal?’, *r/AIDungeon*, Jan. 16, 2022. www.reddit.com/r/AIDungeon/comments/s53ywr/new_player_all_of_my_stories_derail_into_nonsense/ (accessed Apr. 04, 2022).
[30] M. Mateas and A. Stern, ‘Façade: An Experiment in Building a Fully-Realized Interactive Drama’, p. 24.
[31] ‘Steam 上的 God of Word’. https://store.steampowered.com/app/467320/God_of_Word/ (accessed Apr. 04, 2022).
[32] ‘Steam 上的 Orbi’s chronicles’. https://store.steampowered.com/app/1492190/Orbis_chronicles/ (accessed Apr. 04, 2022).
[33] M. P. Eladhari, A. Sullivan, G. Smith, and J. McCoy, ‘AI-Based Game Design: Enabling New Playable Experiences’, p. 13.
[34] ‘Steam 上的 Terraria’. https://store.steampowered.com/app/105600/Terraria/ (accessed Apr. 03, 2022).
[35] ‘Steam 上的 The Red Strings Club’. https://store.steampowered.com/app/589780/The_Red_Strings_Club/ (accessed Apr. 03, 2022).
[36] ‘Steam 上的 A Short Hike’. https://store.steampowered.com/app/1055540/A_Short_Hike/ (accessed Apr. 03, 2022).
[37] ‘Mixamo’. https://www.mixamo.com/#/ (accessed Apr. 06, 2022).
[38] ‘1002 short stories from project guttenberg | Kaggle’. https://www.kaggle.com/shubchat/1002-short-stories-from-project-guttenberg (accessed Jan. 18, 2021).
[39] *openai/gpt-2*. OpenAI, 2020. Accessed: May 01, 2020. [Online]. Available: https://github.com/openai/gpt-2
[40] ‘Expressive Processing: Digital Fictions, Computer Games, and Software Studies | MIT Press eBooks | IEEE Xplore’. https://ieeexplore.ieee.org/book/6267550 (accessed Mar. 27, 2022).
[41] ‘About Wutiaoren’. https://wutiaoren.info/ (accessed Apr. 04, 2022).
[42] D. Yang, Y. Zhou, Z. Zhang, and T. J.-J. Li, ‘AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing’, p. 10.
[43] ‘GPT-3 Powers the Next Generation of Apps’, *OpenAI*, Mar. 25, 2021. https://openai.com/blog/gpt-3-apps/ (accessed Apr. 06, 2022).