

Bringing Stories to Life in 1001 Nights: a Co-creative Text Adventure Game Using a Story Generation Model

Yuqian Sun¹, Xuran Ni², Haozhen Feng³, Ray LC⁴, Chang Hee Lee⁵, and Ali Asadipour¹

¹ Computer Science Research Centre, Royal College of Art, London, UK
{yuqiansun, ali.asadipour}@network.rca.ac.uk

² Unaffiliated, London, UK nxx20110718@gmail.com

³ New Drama xjynewmedia@163.com

⁴ School of Creative Media, City University of Hong Kong, Hong Kong LC@raylc.org

⁵ Affective System and Cognition Lab, KAIST, Daejeon, South Korea
changhee.lee@kaist.ac.kr

Abstract. How can the stories we tell be turned from abstractions in our own minds into concrete elements in a digital environment that we can interact with? To immerse everyday storytelling into 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, mixed-initiative storytelling game using an existing AI creative writing system. In this game, Shahrzad (driven by the player) tells stories through a dialogue interface, while the King (driven by the AI model) continues the player’s story in turn. Text from the story is used in the game mechanics, so that if the player enters keywords such as ‘sword’ and ‘shield’, they are turned into equipment that can be used in battles. Players who are more engaged with the game, measured by the length of their inputs, are rewarded with better achievements. The game aims to facilitate player engagement and creativity through natural language interactions in an empowering setting. This paper presents the game design, a breakdown of the development process and an analysis of user data, including instrumented gameplay data from 2055 players and comments from 422 players. The player feedback indicates that they enjoyed the creative interactions, the game mechanics and the narratives they constructed.

Keywords: Game AI · Intelligent Narrative · Conversational Agent · game design · NCP-Player interactions · Interactive storytelling

1 Introduction

Humans are fundamentally storytellers. From advancing in our careers to making pepperoni pizzas, stories infuse every part of our lives. The ability of machines to

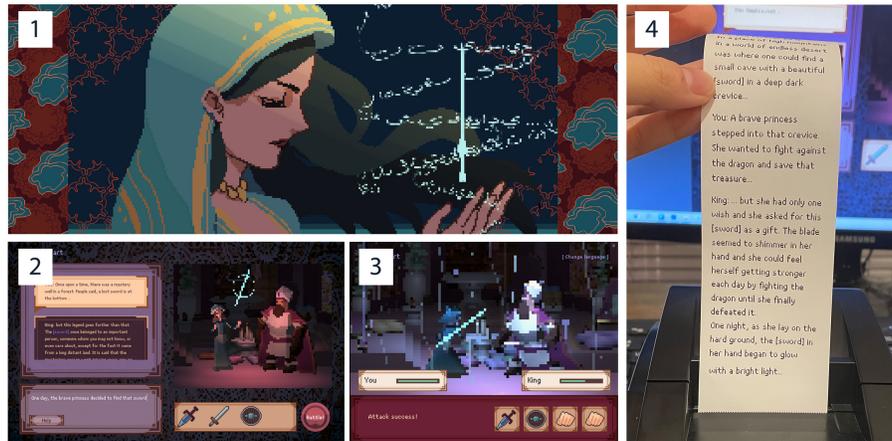


Fig. 1. (1) Shahrzad, the player character, who has a magical ability to turn language into reality. (2) The storytelling phase, where the player writes stories with the King, an AI character. Weapon words like ‘sword’ can be turned into real in-game weapons. (3) The turn-based combat phase, where the player can fight with the king in battles. (4) The printer prints the story when a weapon word is triggered. This matches with the core concept of the game: bringing storytelling into real life.

generate coherent text has allowed stories to be told in new ways by mechanising the writing process, through collaborative writing tools [1,2,3] or even directly talking to fictional characters [11,12]. Can text-based dialogue between human and machine be used as part of the game mechanics? As a metaphor for 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 the classic folklore, we created the game *1001 Nights*⁶ (illustrative screenshots shown in Fig. 1), a co-creative, mixed-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 of a video game setting.

In this game, Shahrzad (controlled by the player) uses a dialogue interface to tell part of a story, and then the King (driven by the AI model) continues the player’s story in turn. 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, creating game mechanics out of the player’s own writing. This leads to an alternative ending of the original

⁶ The game is available for download at: <https://cheesetalk.itch.io/one-thousand-and-one-night>

story: the female storyteller and heroine, Shahrzad, defeats the tyrannical King and puts an end to his heinous crimes.

With this game, we expanded existing creative writing tools to create a playable storytelling experience in a familiar narrative setting. We believe that combining natural language interactions with a classic story can help players to explore and engage more in the game by expressing themselves. The efforts they put into imagination and creativity are rewarded with positive and adaptive content generation by the AI model.

We 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 (measured by the number of their inputs) in the game are rewarded with better achievements, as intended. Comments collected from winning players (n=422) show positive feedback towards various aspects of the game, including the game mechanics and the stories created. Some of their feedback also shows a cultural connection through creative work: some players expressed their own interpretations of characters in the folklore, and were able to include characters and plots from their own cultural backgrounds.

2 Related Work

2.1 Natural Language Processing

Previous studies have investigated the use of natural language processing (NLP) for many different applications, including creative tools [1,2,3]. Some projects have developed collaborative AI writers focused on specific genres, for example Shelly [4], a crowd-sourced horror writer.

Similar approaches have been used in academic research for content generation. Murder mystery generation [5] generated murder mysteries for adventure games, using structured information about 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] presented a series of experiments that connected ancient procedural techniques to modern technologies like language generation models.

Other studies have applied NLP to dialogue systems. Scheherazade’s Tavern [8] and Prom Week [9] tried to develop deeper NPC interactions for a natural social simulation experience. Talk to Ghost [10] adapted Shakespeare’s work to improve high school students’ interest in reading by turning stories into interactive conversations with virtual characters.

In this work, we describe a hybrid experience that sits between creative writing and a game. Some of our main influences are CharacterChat [11] and BanterBot [12], dialogue systems that allow writers to talk with characters they have created. This extends writing assistance to an intelligent agent, turning the context of the interactions into a more familiar social setting. We designed the

dialogue interface in 1001 Nights as a special scenario: two people telling each other stories. Through this, we hope players can easily understand the narrative context, and overcome a barrier to creativity documented by Kreminski and Wardrip-Fruin [13]: the fear of the blank canvas.

2.2 Game Interactions Using NLP

Through the use of NLP, emerging text adventure games give players more control over games compared with traditional games, which give players a limited number of fixed choices. For instance, *Interview With The Whisperer* [14] and *Mystery Of Three Bots* [15] let players explore mysterious stories with natural language text input through Semantic ML, a tool for semantic analysis developed by Google [16]. Fraser et al. [17] developed 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 parts of commercial games such as *KuileiXi* [20], *Event [0]* [21] and *Bot Colony* [22]. These games use a natural dialogue system as the primary means to push the plot forward.

2.3 Natural Language Generation in Stories

Among these NLP-driven projects, only a few have used a natural language generation model, for example personaChat [25] in CharacterChat [11] or OpenAI’s GPT-3 in AI Dungeon [18]; most other projects have implemented NLP for a specific task like parsing player utterances into logical statements [8,26] or finding the closest response from a database [14,15]. The main reason for this is that the use of a natural language generation model risks producing 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 player inputs. Off-topic input will either lead to confusing responses, which are frequently discussed in the player community of *AI Dungeon* [27,28,29], or get limited by the customised module, like in Facade’s Global Context Pool [30], which tried to maintain players’ suspension of disbelief (their belief in the fictional story for the sake of enjoyment) even when their input was out of bounds. Another example is that, in Scheherazade’s Tavern [8], when a player mentions a topic that is not in the knowledge module, the AI agent will repeat a word and then change the topic. These control mechanisms are methods to ensure output quality.

This led us to think about the possibility of letting players decide the game’s plot and generating mechanics corresponding to the narrative framework of their

story. To avoid quality pitfalls faced by other games, we set a clear goal – the player must lead the AI king to mention weapons – to restrict the output and keep it on topic. 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 components of the story (such as equipment and scenes), the full gameplay will become more dynamic, and bring co-authored creative artifact in the same time.

2.4 Dynamic Feedback Beyond Text

In games based on NLP, it is difficult to give dynamic feedback beyond the text. The text can adapt to the player’s natural language input, but the rest of the game cannot. It is very time-consuming for creators to prepare assets, such as character animations and background sceneries, that are synchronous with text output. For instance, in the development of *Facade* [30], two authors spent two years preparing the character reactions and assets for a 20-minute game with a single scene [30]. Some projects have started to use other AI generation models to provide adaptive content, like the GAN-generated images in *AI Dungeon* [18]. Our focus is on the text modality, but our contribution is to map text to another part of the game world: equipment. We were inspired by word-typing [31,32] games, where players must quickly type specific words to release character skills. However, these games do not have semantic relationships between the words and the skills or world environment, while keywords in 1001 Nights will always be part of stories and created through human-AI interactions.

We designed game mechanics to create 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 that they are 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 the game.

The game is made up of two parts: storytelling and battles. Fig. 2 illustrates the game mechanics: the player needs to keep telling stories to lead the King to produce story continuations that mention the important items for battle. In the first phase (Fig. 2 bottom left), Shahrzad (driven by the player) and the King (driven by the AI model) take turns continuing the story. In this game, Shahrzad has a special ability to turn words into reality: when another person utters keywords like ‘sword’, ‘knife’ or ‘shield’, those items materialise and drop to the ground. The player’s goal in this phase is to lead the King to tell more stories that contain keywords and collect weapons.

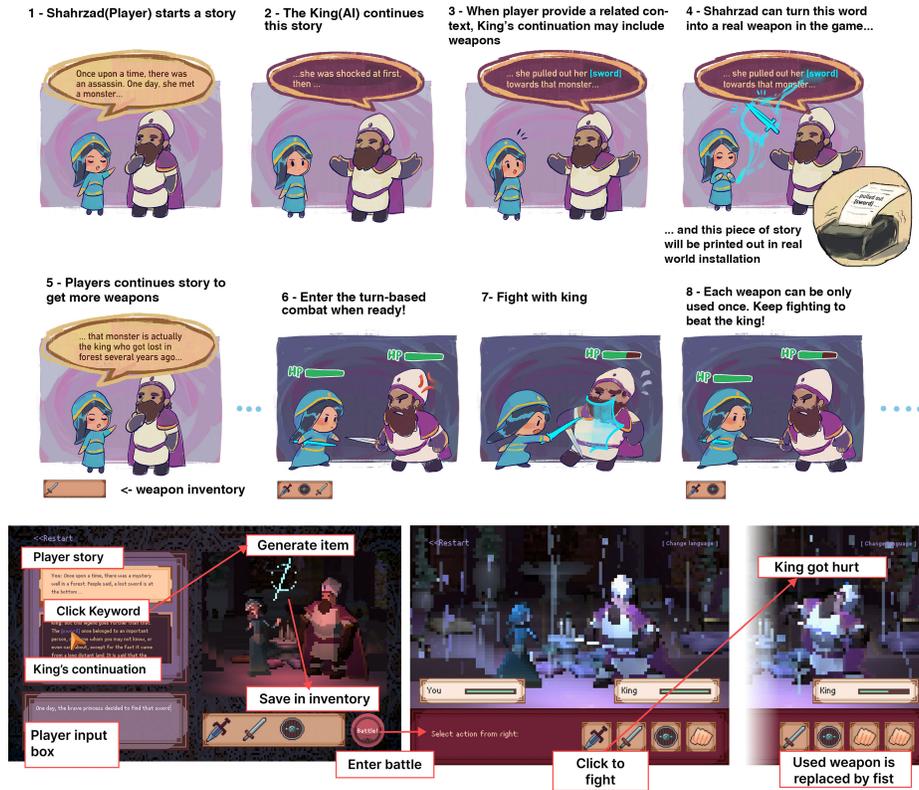


Fig. 2. (Up) Storyboard of the gameplay. (Bottom) Game process

After collecting enough weapons and pieces of armour, the player can enter the turn-based battle phase (Fig. 2 bottom right) to fight with the King. In this phase, the player can use the weapons collected during the last phase to fight the King. The Player’s goal in this phase is to beat the King and free Shahrzad. This is a different ending from the original folklore.

We attempt to combine all components of the game into a coherent experience. The story background links to the mechanics: Shahrzad needs to create stories to survive. The AI system allows the player to be creative and explore different parts of the story.

3.1 Game Art

To encourage players to focus on the gameplay, we use a pixel art visual style. Most pixel art games, like *Terraria* [34] and *Red String Club* [35], use 2D hand-drawn images for all game assets. However, to save time while keeping the fluent aesthetic, we used 3D-to-2D techniques as in *A Short Hike* [36]. With this method, there is no need to manually draw character animations; instead, we used existing action animations from open source libraries like Mixamo [37]. The

resulting game art is still in the traditional 2D pixel art style but with a reduced development time. The low-resolution art that blurs character faces also matches the feel of the ancient story and leaves room for the players' imagination.

3.2 Battle System

The aims of the battle system (illustrated in Fig. 3) are (1) to make the game interesting and challenging and encourage players to write more stories; and (2) to balance the difficulty – a player should not feel that it is too easy or too hard to win. For these purposes, we decided to require players to write at least two stories that trigger valid responses to win the game, that is, a player needs two attack weapons to win.

However, if two attacks are enough to defeat the King, a player will only click twice to win the battle. Hence, the time spent in this phase will be too short. The battle phase needs to be exciting for players, and they may meet failures before the victory. Thus, we 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 attacking, but it deals much less damage. Shahrzad can only attack with a punch when she has no weapon in her hands, as in real life.
2. Some equipment is for defence rather than attack, such as 'shield' and 'armour'. They cannot hurt the king but enrich the game experience by preventing king from dealing damage during his turn. This effect is also realistic.
3. The maximum inventory space is four. Equipment exceeding this limit is not saved for battle but is recorded in the backend to confirm the player's achievement of obtaining weapons. If the inventory is not full, the empty spaces are 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 then punch 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 method of attack is a punch.



Fig. 3. Battle in game

3.3 AI Development

Demo 1 and Keywords We designed four offensive weapons – the sword, knife, blade and dagger – and two defensive weapons – the shield and armour. These are keywords that yielded over 50 matching paragraphs when we searched for them in the Gutenberg Short Story collection [38], a dataset extracted from the portal of Project Gutenberg, a library of over 60,000 eBooks. Meanwhile, these are also weapon words that match with the ancient narrative setting of 1001 Nights. Within the original 94 MB Gutenberg short story dataset [38], we located each paragraph with a matching keyword and extracted the previous five paragraphs, yielding a smaller 16.24 MB dataset. In the first prototype, the King’s AI model was trained on the extracted dataset using the GPT-2 medium model [39] with 6000 steps, so that it would be familiar with common tropes in various adventure stories. The text generation temperature was set to 0.8. Here are several 200-word examples of the results:

Example 1. Player: The princess tried to find her magic sword in the forest

AI: ...and at last through sword at once hid it in her breast; whereupon the beast – a common creature – pretended to understand all she said, and then proceeded to strangle her for a good while. The princess fought with it, till the beast relaxed, took its sword, and set off home. ‘They are the daughters of Snaxen,’ said the young man, passing by to his friends.

Player: ‘Do you know them?’ The princess asked.

AI: ‘I knew them,’ answered Javelin. ‘I was called about twenty years ago, and the King of Thessaly picked me up for the defense of the south against the armies of the Chilkatrachels. He gave me orders, and I carried them, and soon learned to toss my sword, and to wait for the return of the Sun, so that I might meet him face to face in the courtyard of my honor...’

Stories generated in this manner are full of leaps and bounds, and develop logical holes as the story goes on. Although the training model might be improved using a larger dataset, this process would be time-consuming. Even if the model were improved, the story-continuation ability may not be flexible enough for modern players. Since the original story collection includes many old books, the model may struggle to understand players when they write in a more modern style. This may result in low-quality and confusing stories. To fix this problem, we sought alternative solutions.

English and Chinese Demo 2 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 high-quality fiction. Both its English and Chinese datasets consist of open access fan fiction and ebooks on the web, and both datasets are about 100 GB in size. This platform has over one million users of the Chinese version and 200 thousand users of the English version. Although the model is not suitable for all tasks (e.g. writing official documents or code), it is well suited to story generation. To use the model,

it is only necessary to call the application programming interface and send the title and prompt to dreamily.ai’s server[45].

This generation model with a large dataset was able to produce results similar to Demo 1, except for the keywords part. To implement our game mechanics, we designed the structure of requests for the model as in Fig. 4. With this design, dreamily.ai produces flexible stories that closely correspond to player input. The past five inputs are added to the prompt to ensure fluency. Records are refreshed when the player starts a new game or moves to the battle phase. When King’s response does not include a keyword, Shahrzad sends a notification message to provide a hint to the player that they should tell a more relevant story. This helps to relieve confusion and the Tale-Spin effect [40], in which a system makes people feel it is less intelligent than it actually is due to insufficient explanation of the underlying processes.

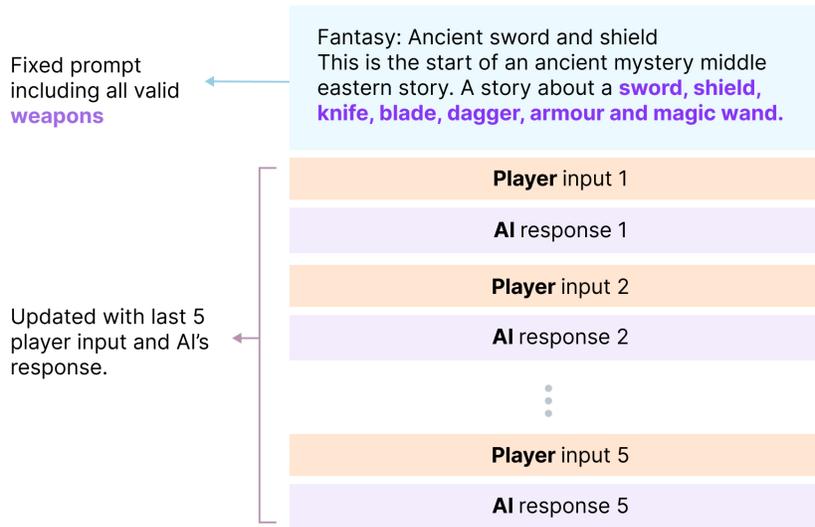


Fig. 4. Requests for the AI model

The player is allowed to mention keywords (like ‘sword’) in the input phase. This may increase the chance of obtaining stories that mention the corresponding weapons in some way, but it will not guarantee it. In contrast, an input that creates a suitable context without a weapon word can still lead to a valid response that contains keywords. Fig. 5 shows an example of this. However, as illustrated in Fig. 6, due to the randomness of the AI system and the limited word list, a player may mention weapon words repeatedly, but may still fail to trigger a weapon. Some players may ask questions in conversations with the king, as shown in Fig. 7, but the king will still give adaptive responses. This can sometimes make the stories appear out of context.

 Player	A fallen chief knight of an army of knights wanders through the Dark Forest with his cherished sword ...	Mentioned valid weapon word: sword
 King	He was not happy that his army had been wiped out by the enemy, but his weapons were pretty much destroyed. "Holy knight lord, I would join your order, please take me in, if you will take me in then I would give anything just to be kept alive." ...	No trigger keyword
 Player	The knight was cold and hungry; he had not had decent food for a week. Then a huge crow flew in and asked, "What good things do you have in your hand, human?"	No mentioned weapon
 King	The raven smelt him and put its beak close to the tip. The raven took a breath and found his weapon appetising and said, "This is a good lance, a rare weapon with blades , although it's broken ...	Trigger keyword [blades]

Fig. 5. Player record with a triggered weapon

 Player	The little dog in the green suit rushed out with a red spear	Mentioned invalid weapon word: spear
 King	The little children were terrified and fled in all directions while the other two rushed over to help, but they were still a step too late and only two of them managed to survive. One of them, a young boy, looked to be about ten years old, his face covered in beads of sweat, but with a look of stubbornness in his eyes. ...	No trigger keyword
 Player	He had heard a legend about magical swords, that if you grew a year older, you would gain a different sword every day	Mentioned valid weapon word: sword
 King	and every day there would be a potion that would change his physique, so he wanted to have it all. The other little girl, who seemed to be very weak, not yet an adult, but whose eyes revealed a resolute look, did not seem to care about anything ...	No trigger keyword
 Player	He had heard a legend about magical swords , that if you grew a year older, you would gain a different sword every day	Mentioned valid weapon word: sword
 King	The girl, however, was not at all afraid of the boy, but looked at him calmly, knowing that he must be her future husband, so she did not show any signs of panic. The two of them supported each other to the depths of the valley, under an ancient tree, when a figure came out from the side ...	No trigger keyword

Fig. 6. Player record with no triggered weapon

 Player	I will fight you to the death, and I drew	No mentioned weapon
 King	...the lance that was stuck in the ground, and then, we began to duel. It is a legendary weapon of very high quality, and in my hand it is called a divine weapon. There are three, six, nine levels of legendary weapons, and the top of the list is the divine weapon.	No trigger keyword
 Player	What weapons do you have?	No mentioned weapon
 King	The legendary weapons are divided into: divine weapons,... How did you acquire your Legendary weapons? The questions are all answered, don't think that's easy, in fact, they have a lot of connotations, but I'm not a professional, I can only choose the most intuitive way to answer these questions, after all ... The sorcerer is killed by his own magic.	No trigger keyword
 Player	What is this magic?	No mentioned weapon
 King	The Divine Soldier is a mysterious long [sword] that looks very much like the weapons used by humans. It is made of dark magic stone, a very rare and scarce magic stone, which has a powerful energy and is the best configuration for alchemy.	Trigger keyword [sword]

Fig. 7. Player record with out-of-context inputs

All these examples, translated from Chinese, are taken from play test records during exhibitions, which are discussed in the next section.

4 Evaluation Study

4.1 Experimental Setup

We were invited to showcase our work at three exhibition sites in China, to research a range of players and collect feedback. All three were in different cities, but they all shared the same digital and analogue setup, including a vintage monitor (to match the ancient setting of the story), a printer, and a workstation using the Windows 10 operating system. A tutorial leaflet (a screenshot of the help page in the game) was on the table for players to read.

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



Fig. 8. Left: Exhibition setup Right: Tutorial leaflet for players

4.2 Opening and Tutorial

All players were informed about data collection for research use before they entered the tutorial. In the tutorial, players were 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 was to let them focus on a more specific instruction: to write about the sword and shield. Another reason was to encourage exploration – finding the valid keywords is also part of the gameplay.



Fig. 9. Left: Opening screen Right: Screenshot of tutorial

5 Results and Findings

Since the core system in the game is the story generation model, this inevitably adds randomness to the results, which cannot be fully limited by rules. When analysing the player data, we aimed to confirm if the AI system can encourage players to engage and contribute more collaboratively in play.

Following the aim of this game – collaborative storytelling with clear goals – ‘engagement’ can be regarded as how much time and energy players want to spend, and ‘contribution’ means the quality of their inputs: whether the sentences make sense or not and whether they are directed towards obtaining weapons. Accordingly, the AI system should give positive feedback to players in the form of responses containing weapon keywords, which become the items that lead to success. If the AI system works effectively, then when a player engages and contributes more in collaborative storytelling, they should receive more weapons, making them more likely to win the game.

When evaluating player contribution, we met some difficulties. We had large amounts of player data (2065 players with 12030 inputs), so it was not feasible to evaluate the quality of all story content. Additionally, since the data were collected during exhibitions, the playing time might have been influenced by unpredictable factors, like the queue length or the number of visitors on that day. Thus, we decided to use average inputs per play to evaluate player engagement, and compare this with their achievements in the game. To be specific, ‘play’ is defined as reaching the end: success in defeating the king, failure to defeat the king, restarting or ending the game. Further evaluation like thematic analysis of stories and play tests without time limitations are left for future works.

To evaluate the level of achievements, we categorised players into three groups, as shown in Table 1: non-winner (G1, n=299), journeyman (G2, n=1106), and winner (G3, n=650). These groups are independent of each other, but the level of progress raised from G1 to G3: non-winners (G1) did not collect any weapons or win the game, journeymen (G2) acquired at least one weapon but did not win, and winners (G3) defeated the king in one or more plays. All players had a chance to familiarise themselves with the game with printed screenshots and integrated tutorials prior to the game, with an identical experimental setup in all three locations.

Table 1. Grouping players by their achievements

Definition	G1	G2	G3
Collected at least one weapon	No	Yes	Yes
Defeated the King at least once	No	No	Yes

5.1 Analysis

This study aims to investigate the impact of engagement in storytelling (average inputs per play) on overall achievement level (from G1 to G3) 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 engagement of each group, $F(2, 2055) = 32.02, p < .05$. We 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 the 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 the players' luck in finding the right inputs. This distinction may explain the difference in variances. Fig. 10 shows the distributions in each group.

A non-parametric Kruskal-Wallis test, as an alternative to one-way ANOVA, is used to evaluate the correlation between player engagement and overall achievements. Players' achievements in this game were reported to be affected by engagement, $H(2) = 328.295$, $p < .05$. The results show that players are more likely to achieve a better outcome by making more contributions to the storyline. Also, a positive trend (shown in Fig. 11) is observed and reported by the Jonckheere-Terpstra test. 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, we can conclude that at least two of the three groups differed significantly in their overall achievements in this game.

In general, the randomness of story generation sometimes influences the gameplay: high engagement (more inputs per play) does not guarantee victory, and fewer inputs may also lead to enough valid keywords for the player to win. However, as shown by the previous analysis, this randomness does not impact the overall performance of game design. In conclusion, the game encourages players to engage in storytelling: the more they engaged, the better achievements they would reach in the game.

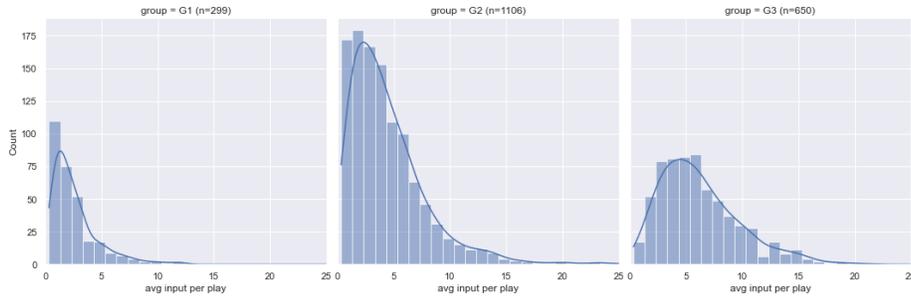


Fig. 10. Average input per play distributions per group

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 feedback in Chinese. Only players in G3 (winners) were allowed to leave feedback after victory for the following reasons:

1. Players played this game during an exhibition, so not all of them had enough interest to reach the end. Sometimes there was a queue to play this game.

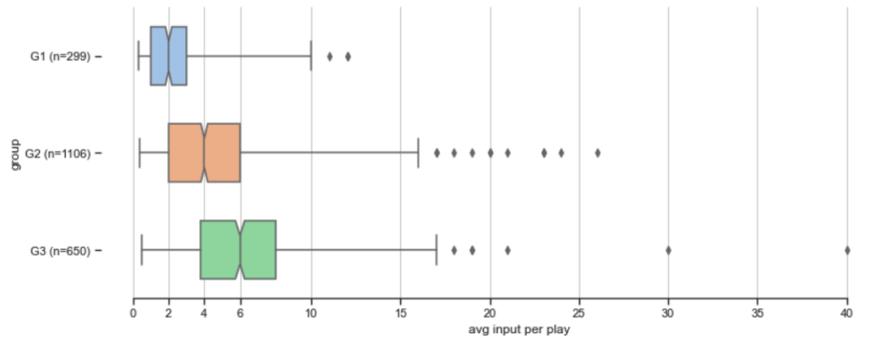


Fig. 11. Positive trend in overall achievements by average contributions made

2. We wanted to encourage players who were defeated to try again until they achieved victory, so that they went through the full gameplay. If we showed the ending page (Fig. 12 left) to all players, including ones who were defeated by the King, they may have regarded it as an ending and left.
3. We assume that players who were patient enough to win gained a deeper experience in this game, which is helpful for us to identify its weaknesses.

We received winners' records (n=650) and removed those who did not leave comments (n=226). We also removed two records from players who met technical difficulties during the experience (the printer was not working).

Finally, with thematic analysis, one of us developed a set of initial codes. After discussion with colleagues, the rest of the remaining feedback (n=422) was identified and classified into nine categories (shown in Fig. 12 right). In future work, we hope to include multiple coders and inter-rater reliability calculations.

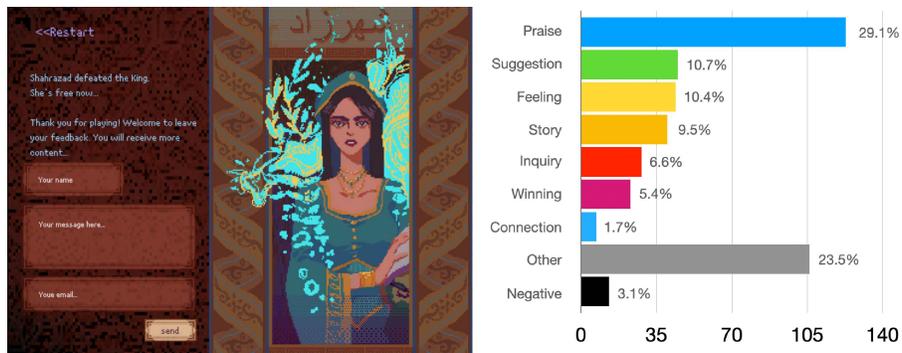


Fig. 12. Feedback page appears when a player achieves victory (left), categorised feedback (right)

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%)

Fig. 13. Comments categories

General Positive feedback like ‘Good game (P10)’ or ‘Interesting (P60)’ are in the praise category. This type makes up the largest share of results (29.1%, n=123). Feedback in other categories shows various focuses. 10.7% (n=45) of players made suggestions. Together with 6.6%(n=28) of players who left inquiries, 8 of them expressed 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 left suggestions, and some from the inquiry category (n=7) asked for the inclusion of more weapons. Although some of them (n=5) mentioned weapons like guns that do not fit into the narrative setting, we do recognise the benefits of improving weapon choices and better player guides. Several players (n=13, 3.1%) were unsatisfied with or disliked this game, which is expected for a game at an early stage of development.

Immersion 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 expressed thoughts about AI: ‘Humans reach consensus with AI (P586)’ and ‘Humans are those who think beyond AI (P561)’. These suggest our game can provoke reflections about freedom, strength and the agency of AI. 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 that nearly one-third of this group (n=14) shared their impressions 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 did not add any per-

sonal lines to the king, some players showed 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 interpretations can fill the gap in the story. The players’ feedback was more varied than expected. We expected to receive general praise and suggestions – 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 characters they included 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 engaged in the stories they created.

While some players were immersed in the stories they created, some players receive more pleasure from their victory. 5.4% (n=23) of players gave 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.

Cultural Connection Among players who commented on the story they created, many of them were inspired by personal interests that matched 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 performed in exhibitions in China, some of the players put aspects of their cultural background into their stories, which became creative artifacts through human-AI collaboration that show possibilities in cultural blends. For example, ‘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 is usually mentioned in grave robbing stories. 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 who are famous for their attention to the current situation of Chinese rural youth and for the strong humanistic feelings in their music.

Reflection on Reality It was observed that a few players (n=7, 1.7%) even connected this game to their experiences and feelings in real life, like someone who felt 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 potential ability as an educational game for story writing.



Fig. 14. Exhibition photos. Left: A 11-year old boy kept playing for 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 engagement in storytelling. Players with higher engagement were more likely to reach improved achievements in the 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 could reward them with creative feedback.

Even when we did not ask players to rate their experience, in comments from players (n=422) who won the game, players showed a high level of enjoyment and interest throughout the game, where they contributed 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 studies [42], the unexpected but logical text generated by AI may make the story more exciting than the player’s intention.

Players expressed their own interpretations of characters in the folklore and were able to include characters and plots from their own cultural backgrounds. For them, the king could be a coward, a peace lover or a gastronome, and these

are reflections through the creation, rather than the line the creators set. They could 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 gaps that developers leave blank. This is similar to the finding of Aljammaz et al. [8] that a player may view the repeated responses as the NPC’s own personality.

7 Limitations

The creators faced common barriers in developing 1001 Nights. Like similar studies on dialogue interfaces, we found that open dialogue systems are a double-edged sword. They contribute to a sense of freedom but face the risk of going off track, and to limit that requires a large amount of authoring and design work.

Current weapon keywords are specific and limited. In future work, we plan to use semantic similarity detection to extend the range of valid keywords. For instance, in WebVectors [44], the sword is similar to the scimitar, rapier and broadsword. Less similar but closely related words, like hilt and scabbard, may become fragments that can be used to form actual weapons later.

Furthermore, due to the large playtest data, the current AI system was not able to evaluate the quality of player input. Consequently, we could only evaluate player performance through engagement (number of inputs per play). In future we hope to analyse the quality of inputs and responses received from the King. To see the impact of the game environment, we also hope to analyse how the results differ for players who directly use the story generation model and those who play the game in a version without keywords. In future work, participants will be able to download and play the game in their preferred environment, without the time pressure of physical exhibitions.

Meanwhile, when the player input includes some components that do not fit well in the setting of The Thousand and One Nights (like ‘computer’ or ‘rifle gun’), the king can still continue the story, which may reduce player immersion, since an ancient king should not know about modern technologies. This could be improved by future enhancements, like keyword detection or neural classifiers.

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

8 Conclusion and Future Work

To extend storytelling to real-life contexts beyond the language, we created the game 1001 Nights, a co-creative storytelling game that leverages story writing into actual game mechanics, based on an existing story generation model. We have shown that 1001 Nights facilitates player engagement and creativity through natural language interactions in a well-known folklore setting. Our data

suggest that higher player engagement generally leads to better achievements in the game, which demands further investigations.

Potential extensions of this research include multiple coder thematic analyses of players' stories and comments, and asking them to evaluate their engagement and interest on a Likert scale. This is key to evaluating the quality of inputs via a hybrid approach using NLP algorithms like text perplexity and domain experts. The game design could be improved to enhance replay value. For example, a weapon index could be added that documents the stories behind the triggered items. Furthermore, with emerging text-to-image technologies like DALL-E[46], it may even be possible to change the appearance of a weapon or character based on the corresponding sentences or change the background scenery over the course of the story. Beyond 1001 Nights, supported by text generation models, similar mechanics that set clear goals for players could be extended to more games, like suggesting a non-player character to spill a secret location that hides treasure.

References

1. Swanson R, Gordon AS: Say anything: A massively collaborative open domain story writing companion. In: SpringerLink (2006) https://link.springer.com/chapter/10.1007/978-3-540-89454-4_5
2. ColorfulClouds Tech, Dreamily.ai. <https://dreamily.ai/> Accessed 16 Oct 2022
3. Coenen A, Davis L, Ippolito D, et al: Wordcraft: A human-ai collaborative editor for story writing. In: arXiv.org(2021) <https://arxiv.org/abs/2107.07430>
4. Yanardag P, Cebrian M, Rahwan I: Shelley: A crowd-sourced collaborative horror writer. *Creativity and Cognition*(2021) <https://doi.org/10.1145/3450741.3465251>
5. Barros GAB, Liapis A, Togelius J: Murder mystery generation from open data. In: L-Università ta' Malta (2016) <https://www.um.edu.mt/library/oar/handle/123456789/47474>
6. LC RAY, Mizuno D: Designing for narrative influence: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems(2021) <https://doi.org/10.1145/3411763.3450373>
7. Pichlmair M, Putney C: Procedural generation for divination and inspiration. *International Conference on the Foundations of Digital Games*(2020) <https://doi.org/10.1145/3402942.3402950>
8. Aljammaz R, Oliver E, Whitehead J, Mateas M: Scheherazade's Tavern: A prototype for deeper NPC interactions. *International Conference on the Foundations of Digital Games*(2020) <https://doi.org/10.1145/3402942.3402984>
9. McCoy J, Treanor M, Samuel B, et al: Prom week: Designing past the game/story dilemma . In: *Foundation of Digital Games (FDG)*(2013)
10. Jackson D, Latham A: Talk to the ghost: The Storybox methodology for faster development of storytelling chatbots. *Expert Systems with Applications* 190:116223. (2022) <https://doi.org/10.1016/j.eswa.2021.116223>
11. Schmitt O, Buschek D: CharacterChat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot. *Creativity and Cognition*(2021)<https://doi.org/10.1145/3450741.3465253>
12. Banter bot by Google Creative Lab - experiments with google. In: Google. <https://experiments.withgoogle.com/banter-bot> Accessed 16 Oct 2022

13. Kreminski M, Wardrip-Fruin N: Generative Games as Storytelling Partners. Proceedings of the 14th International Conference on the Foundations of Digital Games(2019) <https://doi.org/10.1145/3337722.3341861>
14. deconstructeam, Interview with the Whisperer . In: itch.io. <https://deconstructeam.itch.io/interview-with-the-whisperer> Accessed 16 Oct 2022
15. (2020) The Mystery of the Three Bots. <https://google.github.io/mysteryofthreebots/>. Accessed 16 Oct 2022
16. (2021) Semantic ML. In: Google. https://github.com/google/making_with_ml Accessed 16 Oct 2022
17. Fraser J, Papaioannou I, Lemon O: Spoken conversational AI in video games. Proceedings of the 18th International Conference on Intelligent Virtual Agents(2018) <https://doi.org/10.1145/3267851.3267896>
18. ‘AI Dungeon’. <https://play.aidungeon.io/main/home> (accessed Oct. 21, 2021).
19. Sali S, Wardrip-Fruin N, Dow S, et al: Playing with words. Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10(2010) <https://doi.org/10.1145/1822348.1822372>
20. Xi Y, Mao X, Li L, et al:Kuileixi: A Chinese open-ended text adventure game. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations(2021) <https://doi.org/10.18653/v1/2021.acl-demo.21>
21. ‘Event[0] on Steam’.
<https://store.steampowered.com/app/470260/Event0/> (accessed Oct. 27, 2020).
22. ‘Bot Colony in Steam’.
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. (2012)
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) <https://www.jstor.org/stable/41388707>
25. Zhang S, Dinan E, Urbanek J, et al: Personalizing dialogue agents: I have a dog, do you have pets too? (2018) <https://arxiv.org/abs/1801.07243>
26. S. O. Villar, SHRDLU. 2022. Accessed: Mar. 27, 2022. <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’(2003)
31. ‘Steam - God of Word’. https://store.steampowered.com/app/467320/God_of_Word/ (accessed Apr. 04, 2022).

32. ‘Steam - Orbi’s chronicles’. https://store.steampowered.com/app/1492190/Orbis_chronicles/ (accessed Apr. 04, 2022).
33. Eladhar MP, Sullivan A, Smith G, McCoy J: AI-based Game Design: Enabling new playable experiences (2011)
34. ‘Steam Terraria’. <https://store.steampowered.com/app/105600/Terraria/> (accessed Apr. 03, 2022).
35. ‘Steam - The Red Strings Club’. https://store.steampowered.com/app/589780/The_Red_Strings_Club/ (accessed Apr. 03, 2022).
36. ‘Steam - 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. Chatterjee S :1002 short stories from project Gutenberg. In: Kaggle(2020) <https://www.kaggle.com/shubchat/1002-short-stories-from-project-gutenberg> Accessed 16 Oct 2022
39. openai/gpt-2. OpenAI, 2020. Accessed: May 01, 2020. <https://github.com/openai/gpt-2>
40. Wardrip-Fruin, N.:Three Play Effects–Eliza, Tale-Spin, and Sim City. *Digital Humanities*, 1-2. (2007)
41. ‘About Wutiaoren’. <https://wutiaoren.info/> (accessed Apr. 04, 2022).
42. Yang, D., Zhou, Y., Zhang, Z., Li, T. J. J., and LC, R.: AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops 2022* (Vol. 10). (2022)
43. ‘GPT-3 Powers the Next Generation of Apps’, OpenAI, Mar. 25, 2021. <https://openai.com/blog/gpt-3-apps/> (accessed Apr. 06, 2022).
44. M. Fares, A. Kutuzov, S. Oepen, and E. Velldal, ‘Word vectors, reuse, and replicability: Towards a community repository of large-text resources’, in Jörg Tiedemann (ed.), *Proceedings of the 21st Nordic Conference on Computational Linguistics, NoDaLiDa*, Linköping University Electronic Press (2017)
45. Dreamily API. In: Dreamily API - CaiyunWiki. https://open.caiyunapp.com/Dreamily_API Accessed 17 Oct 2022
46. OpenAI Dall-E 2. In: OpenAI. (2022) <https://openai.com/dall-e-2/> Accessed 17 Oct 2022