Fictional Worlds, Real Connections: Developing Community Storytelling Social Chatbots through LLMs

Yuqian Sun[1] , Hanyi Wang[2] , Pok Man Chan[3], Morteza Tabibi[4] , Yan Zhang[4] , Huan Lu[5] , Yuheng Chen[4] , Chang Hee Lee[6] , Ali Asadipour[7] ,

[1] Computer Science Research Centre, Royal College of Art, United Kingdom
[2] Independent, United States
[3] Independent, Hong Kong
[4] rct.ai, United States
[5] Independent, China
[6] Affective Systems and Cognition Lab, Industrial Design Department, College of Engineering, KAIST, United Kingdom
[7] Royal College of Art, United Kingdom

Abstract
We address the integration of storytelling and Large Language Models (LLMs) to develop engaging and believable Social Chatbots (SCs) in community settings. Motivated by the potential of fictional characters to enhance social interactions, we introduce Storytelling Social Chatbots (SSCs) and the concept of story engineering to transform fictional game characters into "live" social entities within player communities. Our story engineering process includes three steps: (1) Character and story creation, defining the SC's personality and worldview, (2) Presenting Live Stories to the Community, allowing the SC to recount challenges and seek suggestions, and (3) Communication with community members, enabling interaction between the SC and users. We employed the LLM GPT-3 to drive our SSC prototypes, "David" and "Catherine," and evaluated their performance in an online gaming community, "DE (Alias)," on Discord. Our mixed-method analysis, based on questionnaires (N=15) and interviews (N=8) with community members, reveals that storytelling significantly enhances the engagement and believability of SCs in community settings.

Keywords: *social chatbot, believable agent*

1. Introduction

Fictional characters are central to the research of social chatbots: artificial intelligence (AI) dialogue systems capable of having social and empathetic conversations with users [BSF]. The field of Social Chatbots (SCs) has been rapidly advancing with the development of Large Language Models (LLMs), which enable SCs to generate dynamic and personalized responses for different users. These AI dialogue systems are becoming increasingly human-like and are taking on a variety of social roles such as assistant, partner, friend [Met20] [PHX23], citizens [POC*23], and YouTuber. While researchers have begun studying SCs in relation to Human-AI friendship and intimate relationship development, most studies have focused on interactions with a single bot rather than in richer social contexts. Furthermore, existing research has not deeply considered the role of fictional content. To some extent, SCs are similar to game characters as they are both fictional entities. Computing systems are wrapped into different "people" with distinct names, personalities, and background stories. As such, the question remains: Can fictional characters live together with humans, and be-

come members of human society? How will their fictional identities and stories influence our perception?

Previous researchers explored chatbots as community members using a Twitch chatbot [SLY*20], but their study did not employ LLMs, which offer more advanced language abilities, necessitating further investigation. Additionally, current LLM-driven SC like replika [BSF] does not focus on multi-person social contexts. By integrating insights from gaming research, particularly non-player characters (NPCs) [Bow19], we strive to develop a better understanding of how SCs can become more believable and engaging within community settings.

We explore the idea that fictional characters can share their lives as social chatbots in the human community context, just like friends sharing their lives on social media and receiving feedback. Rich social interactions, including conversations with the character, seeing the character's interactions with other people, and the dynamics in the character's life, can make the character believable and the interaction engaging and meaningful.
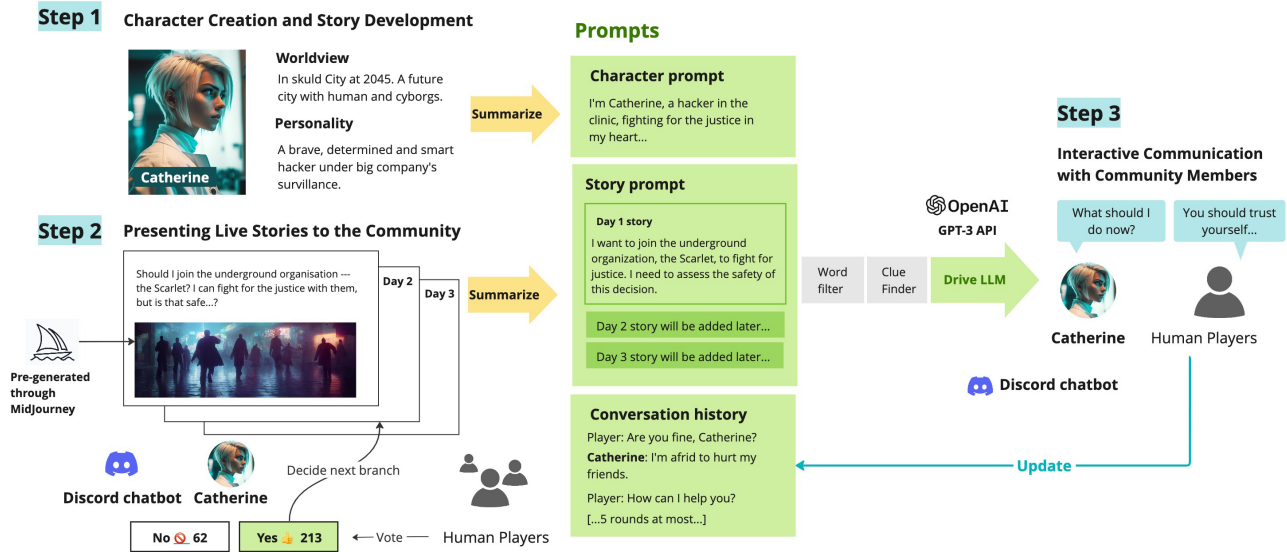
Figure 1: *Story engineering for Storytelling Social Chatbot (SSC) Catherine*

Motivated by these views, we investigate the following research questions:

RQ1 - How can an LLM-based SC enhance engagement and meaningful interactions within a community setting?

RQ2 - How does the integration of fictional stories in the SC's design impact its believability and engagement?

We refer to the process of transforming a fictional character into a social chatbot as *story engineering*. This term is adapted from *prompt engineering*, which focuses on designing text prompts to influence the generated content of an LLM. In contrast, designing an LLM-based "live" character with social behaviour requires considering multiple aspects: the character's personality and story, the LLM's generation goals, and the interaction methods facing people. Our prototype, Storytelling Social Chatbot (SSC), implements this concept through the following processes:

(1) Story and character design, where we define the SC's personality and the worldview they inhabit, (2) Presenting Live Stories to the Community, allowing the SC to recount challenges and problems they need suggestions on, and (3) Communication with community members, enabling community members to chat with the SC. We designed a workflow to drive the character through the LLM GPT-3.

To evaluate our workflow, we introduced two fictional characters, "David" and "Catherine," as SSCs in an online gaming community "DE (Alias)" on Discord. These characters followed the worldview and story of the under-development game "DE". Every day, the character shared their current situation in the story channel (e.g., "I'm chased by the evil agent, what should I do?", similar to social media. Meanwhile, community members could engage with them and discuss their current situation in the chat channel, suggesting decisions through voting on the character's choice. At the end of the day, the character made their decision based on voting and released the next story.

The DE gaming community already had a non-storytelling chatbot based on LLM GPT-3 [Ope], Jerry, which we used as a benchmark to evaluate the effectiveness of SSCs. We collected qualitative feedback through questionnaires with 15 core community members and interviews with eight of them. Our mixed-method analysis reveals that storytelling enhances the engagement and believability of SCs in community settings. We summarize the themes related to SSC and discuss the design implications for future developments to make SC interactions more engaging and meaningful.

In summary, we contribute (1) Two community-based SSCs developed through our concept of story engineering based on GPT-3, and (2) Insights from their development and evaluation with community members. By designing SSCs with attention to the specific social context and using storytelling to enhance their believability and engagement, we can create a new generation of SSCs that can contribute to our social lives in novel and exciting ways.

2. Background

In online communities, achieving meaningful interaction is essential. Although many active communities employ bots with both moderating and entertaining functions [Rap21], their interactions may not necessarily generate new information or contribute to the community's values or project goals. Narratives, according to Ricoeur, are vital for constructing our sense of self, making sense of our experiences, and creating meaning in our lives [Alb10]. Storytelling has been considered a means and approach in various contexts, such as children's education [ZXW*22], [JL22], healthcare, and skill learning [SZL22]. In online community building, particularly game-related communities, storytelling may serve as a crucial goal or primary means of maintaining community activity and a sense of meaning [Alb10].

Integrating AI as a member of an online community could potentially become a source of discussion or meaning by sharing its life experiences consistent with the content the entire community focuses on. These experiences are unlike the small talk generated

by existing SCs such as Replika [BSF]. While there has been extensive research on dyadic interaction chatbots [BSF22] [CG20] [SYKL21], multi-party chatbots, particularly those acting as community members, remain underexplored.

Researchers have developed a Twitch chatbot that focuses on the social context of a community, making the chatbot a social member [SLKH19]. However, this does not consider the potential impact of current LLMs, such as uncontrollable content generation and human attachment. Seering et al. proposed various ideas for community chatbots, including the storyteller bot concept [SLKH19]. They suggested that a more interactive, almost "live" narrative experience could be created by having chatbots that are regular community members involved with other chatbots in engaging ways. Furthermore, chatbots should focus on being deployed in specific social contexts.

Developing social abilities in chatbots requires mimicking human behaviour to some extent. Accordingly, previous research on SCs has focused on personification, emotions [Heu18], and other related aspects. Similarly, research on NPCs' social behaviour in games addresses emotional attachment [ERM18] [BMA*19] [Bow19], empathy [CG20], and identity [Dim13]. Due to the multi-modal nature of video game experiences, the influencing factors in related research are more diverse, such as character appearance and game environment. Overall, computer games can be considered a social "training ground" for NPCs [Bow19], and natural language interaction opens room for players to provide new content that can be (to an extent) acknowledged by the game [POC*23].

Since we focus on character dialogues in a story context, we focused on believability. Anton et al. defined believability as the extent to which users interacting with the agent come to believe that they are observing a sentient being with its own beliefs, desires, and personality [BTS]. The authors concluded that a believable character is not necessarily a real character but must be real in the context of its environment. Virtual agents that can adapt to changes in the environment and exist in the correct social context are those perceived as more believable. Kiran et al. argues that the believability of such agents is tightly connected with their ability to relate to the environment during a conversation [IBS]. Believable characters can create better player experiences, and accordingly, believable SC should bring a more engaging experience to a community.

Considering that online community interactions are mainly text-based and do not involve 3D scenes or embodied agents, we synthesized several scholars' discussions and regarded **emotions, personality, and motivation** as the criteria for measuring SSCs believability. By focusing on these aspects, we can aim to design a storytelling SC that is logical, coherent, and clear, leading to a new generation of SCs that contribute to our social lives in innovative ways.

3. SSC Development and Implementation in the Community Context

In this section, we will discuss the development of the Storytelling Social Chatbots (SSCs), "David" and "Catherine," within the context of the DE gaming community. We will cover the background of the DE game, the DE community, the character background and story settings, and the detailed experimental workflow.

3.1. DE Game and Community Background

The DE game is an under-development first-person shooting player vs player (PVP) game set in a futuristic virtual world featuring digitized real humans. Players engage in battles for interests, power, and resources, shaping the future of the virtual world. The game is based on blockchain technology, and its community members are open to new technologies. The DE community has a large number of members on Discord (n=97841), with 7006 having sent over 5 messages in the community. However, it should be noted that in Web3 and blockchain-based Discord gaming communities, a considerable amount of members are inactive after joining. Typically, players adopt a wait-and-see approach after encountering project promotions before deciding whether to follow up. The active membership ranges from around 100-200 people, with the most core members being around 15-20 who can directly discuss moderating issues with official game team members. Furthermore, the DE community already had a non-storytelling LLM-based chatbot, Jerry, which we used as a benchmark to evaluate the effectiveness of SSCs. We provide a brief overview of their story backgrounds and purposes in 1.

3.2. Character Background and Story Settings

After consulting with the DE game's development team, we selected two future in-game characters, David and Catherine, for our research experiment. David is a doctor who uncovers illegal operations involving mind control technology within the company he works for and decides to escape with crucial information. Catherine is a skilled hacker who discovers her past is filled with tragedy, fueling her hatred for oppressive forces and motivating her to fight against them.

The development and story settings of these two characters in our research are consistent with the game's official narrative. The experiment conducted in this research also serves as a player community activity, promoting the game's characters and worldviews while engaging the community.

3.3. Detailed Experimental Workflow

To develop and implement the SSCs within the community context, we followed a three-step workflow as part of our story engineering concept:

1. **Character Creation and Story Development**: Define the characters' personalities, backgrounds, and stories, as well as the world they inhabit.
2. **Presenting Live Stories to the Community**: Share the characters' live stories with the community, allowing them to interact and provide suggestions. Each day, the character will share their current situation in the *story channel*, with community members who can vote on the character's choices. Concurrently, community members can engage with the SSCs and discuss their current situations in the *chat channel*. Based on the voting results, the character will make a decision and release the next part of the story at the end of the day.

Table 1: *Character and Stage Overview*

Character and Stage	Function	Duration	Background story
Jerry: Benchmark	Chat with people.	Available 4 months later after the community opened in 2021 December.	A traveller in the metaverse.
David: Pilot Study	Releasing stories and chatting with members through the adventure.	Opened through the 3 days warm-up, 3 days main stories in 2022 November.	An employee in a pharmacy institute "BioTech" under the big company "Domain". Ran away after finding the illegal trade in the company. Met and saved by Catherine and her father, who turned out to be the hidden villain that made Catherine mind-controlled.
Catherine: Main Study	Same as David plus clue finder function: replying to specific questions with pre-scripted replies.	Opened through the 1-day warm-up, 3 days main stories at 2023 February.	Under surveillance by the big company "Domain" who used to mind-control her but released her after David's story. A hacker and doctor in the clinic. Wants revenge on "Domain". Regards David as a big brother.

3. **Interactive Communication with Community Members**: Enable community members to communicate with the SSCs, discussing their situations. Each day, the character will change their state upon the current live story, and continue to talk to community members and respond to their suggestions throughout the day.

We conducted a pilot study with David's character in November 2022 to preliminarily assess the efficacy of SSCs. Subsequently, we treated Catherine's character as the main study, conducting questionnaires and interviews with core community members, and asking them to compare Catherine and Jerry to investigate the impact of storytelling on SCs.

4. Implementation of Story Engineering

In this section, we will explain the different stages of story engineering Fig.1, focusing on the creation and implementation of SSCs "David" and "Catherine."

4.1. Step 1: Character Creation and Story Development

To design the story of characters, the following perspectives should be considered:

Worldview: Determine the setting of the story. Consider the rules, customs, and culture of this world, as they will influence the character's interactions with the community. For example, following the background story of DE, stories happen in Skuld City in 2045, a futuristic city inhabited by humans and cyborgs.

Personality: Define the character's personality and motivations. Consider how these characteristics will shape their decisions and behaviour throughout the story. For instance, Catherine is a brave, determined, and smart hacker, and she tries to fight for justice against a large company's control.

4.2. Step 2: Presenting Live Stories to the Community

The story prepared by the team members will be posted by the SSC in the story channel, similar to a short novel, and should consider the reading experience of the players.

The current situation the character is experiencing should be identified, including their goals, motivations, and the challenges they face. This will provide a context for the character's interactions with the community and the decisions they need to make. For example, David will say: "Today, my world turned upside down when I got shot by Catherine, who I always trust..." At this point, he needs to decide whether to continue trusting his friend. The design of this part is similar to branching choices in traditional interactive narrative video games, requiring the scriptwriter to plan each branch in advance.

In our project, although the story is mainly conveyed through text, we also prepared illustrations for easier understanding by the audience. The illustrations were generated by MidJourney [Mid] before the activity started based on the story. Additionally, the character's decision options were presented with representative emoji reactions as selectable buttons.

Using the discord.py [Rap21] tool provided by Discord official team, we can schedule the daily live story posts, including text, images, and options. Players can vote on the options and view the current voting results. On the next day, following the higher-voted choice, the next story will be published.

To better introduce the character, we put the beginning days of the event as a warm-up phase, the character's story didn't provide a choice. After that, players will get a better understanding of the character, so they can be more considerate when they talk to the character.

4.3. Step 3: Interactive Communication with Community Members

To enable interactive communication with community members, we employed a conversation system consisting of several modules that work together to drive the chatbot's dialogue system(Fig.1):

4.3.1. Words Filter

This module is designed to filter out offensive inputs/outputs by scanning the text through a list of keywords.

4.3.2. Clue Finder (Catherine only)

The Clue Finder is a small API microservice designed for providing static replies. It compares a given sentence to a list of keywords using the SentenceTransformer [RG19] model. When a match is found, it returns the corresponding image URL and text. The service is lightweight and easy to deploy in a private network, making it suitable for projects that require fast responses with low latency.

When the input is similar to options in keywords (e.g., "Give me some information about Domain"), Catherine will reply with the pre-written text and won't call the OpenAI service. This function saves time and expense for basic questions and explains the worldview with several options, including images.

4.3.3. LLM Configuration

When the Clue Finder does not generate a response, the OpenAI GPT-3's API is called to generate a reply based on the given prompt. The prompt consists of three parts: character prompt, live story prompt, and dialogue history.

Character prompt: Provides background information on the character (e.g., "I'm David, the doctor in the clinic."). Adapted from the character information prepared in Step 1. To make the current SSC understand the current story and save words in the prompt, some simplifications and adaptations are needed.

Live Story prompt: Derived from Step 2, it conveys the current situation or context to the LLM, enabling it to generate contextually appropriate responses. As the story progresses, the story prompt will also become longer.

Dialogue history: This consists of chat records from the past five rounds of conversation, allowing the character to develop brief memories and respond coherently to ongoing discussions.

This conversation system allows the SSCs to communicate with community members, respond to their suggestions, and adapt their stories based on the current live story.

5. Pilot study: David

For the pilot study, we first introduced a character named "David" in 2022 November. We created a story about how he discovers the dark side of Skuld City (the main stage of DE) and how he fights back against the villainous company "Domain". The main goal of the pilot study with David is to validate the feasibility of storytelling SCs and their attraction within the community. We released David during the DE game testing days, which attract more players in the community.

5.1. Observations on Made-up content from LLM

During the event, LLM-generated content unexpectedly influenced the story.Fig.12When players asked David about the name of BioTech's boss, GPT-3 fabricated the name Gustav Weis. This led to more players inquiring about Gus' details, prompting David to generate further information. Although other names were occasionally fabricated, Gus Weiz had the highest occurrence rate since players kept using this name.
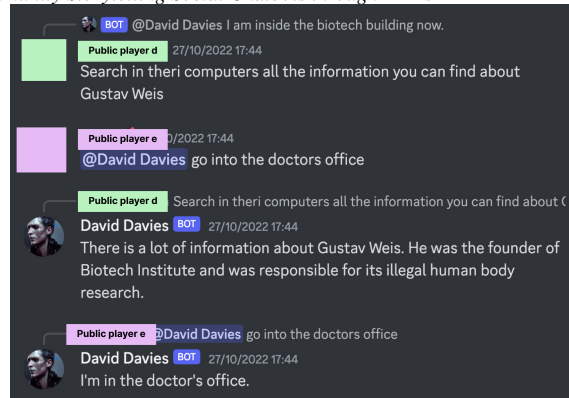

Figure 2: *David making up the content of password and player's reactions and "Guatav"*

Recognizing the value of this generated content, the authors and team decided to incorporate Gus Weiz into the story. As a result, David's prompt was updated to include Gus Weiz as the boss of BioTech. Interestingly, the name Gustav is similar to Gustavo Fring, a famous corporate boss from the TV series Breaking Bad, fitting the character well.

In other words, David's dialogue system unexpectedly provided us with inspiration for improving our story through improvisation. However, to better convey the story, it need a developed control method.

5.2. Data

David has received 31278 community interactions over 6 days. 1049 players talked to David, and 27% of them(n=287) are active members who sent over 10 messages. On average, 206 people vote for a choice each day. During the event period, David's chat channel takes 30.18% of all messages on the DE's server. This shows the potential of a storytelling SC in the community, so we move on to the next study.

6. Main study: Catherine

Catherine is one main character in David's story, and the protagonist in the second event. Her motivation is to revenge against the villainous company "Domain", who hurt David and herself through mind-control. We use mixed methods(questionnaires and semi-structured interviews) to investigate the performance of SSC.

6.1. Improvement

From David's activity, we observed that the SC driven by LLM will have fabricated behaviour. Therefore, we set some fixed replies for Catherine through the Clue Finder module in the dialogue system, such as her age. And questions related to worldviews, such as information about cities and gangs, will return replies with pictures. In addition, as we observed that David's interaction decreased over time, we shortened Catherine's time to 4 days (1 day for warm-up, 3 days for live story with choices). Additionally, during the warm-up phase, David's Story will be indicated to guide players who did not participate in David's activity to learn about his story.

6.2. Data

During Catherine's event, the server was generally less active than during David's event, as it did not coincide with the DE game testing period. As a result, there were fewer total messages in the channel for Catherine (n=20767) compared to David (n=31278). A total of 907 speakers participated, with 317 active members engaging in conversations more than ten times, accounting for 35% of the speakers, which is higher than the 27% in David's activity. Additionally, an average of 222 people voted for a choice each day, slightly higher than David's average of 206.

Catherine's chat channel accounted for 89% of the total community activity during her event days, indicating that she served as a catalyst for community engagement when overall activity was lower.
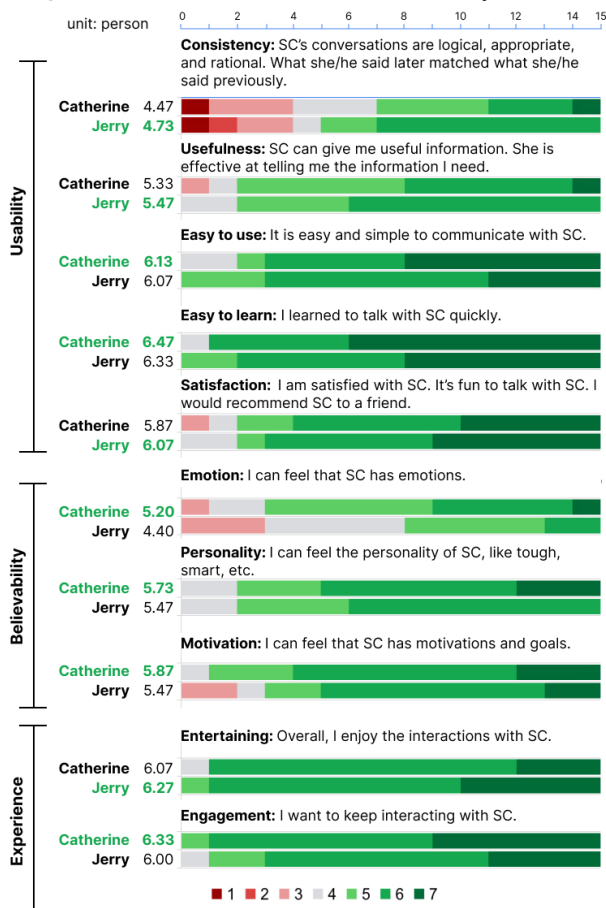
6.3. Questionaire items and Core member survey

Figure 3: Evaluation from close members and questionnaire items

6.4. Participants and procedure

To evaluate the effectiveness of Catherine within the community, we sought participants who had experienced both David and Catherine's events and were relatively familiar with the community before those events. As a result, we invited the most active core members (n=15) of the community to participate in the user study.

Table 2: Demographics of close members. Although the proportion of female participants is relatively low, this gender imbalance is primarily due to the underrepresentation of females in Web3 communities. According to the community statistics in November 2022, before David's study started, the male-to-female ratio was approximately 8:1, which aligns with the ratio of our interview participants.

ID	Gender	Age	Language area
Interviewed			
P1	Male	18-24	Hindi
P2	Male	25-34	Chinese
P3	Male	25-34	Russian
P4	Male	25-34	Tagalog
P5	Male	18-24	Chinese
P6	Male	25-34	Ijaw (Nigeria)
P7	Male	35-44	Igbo (Nigeria)
P8	Female	18-24	Tagalog
Not interviewed			
Skipped	Female	18-24	Tagalog
	Male	25-34	Tagalog
	Prefer not to say	18-24	Tagalog
	Male	25-34	Vietnamese
	Male	18-24	Vietnamese
	Male	25-34	Spanish
	Male	25-34	Chinese

These core members frequently interact with the official team's moderators and provide suggestions, but they do not participate in the design or execution of community events, nor were they involved in the design or planning of the SSCs.

A mixed-method approach, incorporating both qualitative and quantitative data, was employed to capture the players' experiences (usability, perception of believability, and engagement) while interacting with Catherine. Two questionnaires were designed and adapted from the System Usability Scale [Bro95] and questionnaires in research about believability [BTS] [IBS] using Qualtrics XM [Qua20]. Before Catherine's event, core members were required to chat with the benchmark AI chatbot, Jerry, and then complete Questionnaire 1 to evaluate his performance. After Catherine's event, core members completed Questionnaire 2 to evaluate Catherine. The evaluation questions for Catherine were the same as those for Jerry, with the addition of a section to assess the story function (Fig.3), allowing for a comparison with Jerry in terms of storytelling.

Following the completion of both questionnaires, core members were invited to participate in semi-structured interviews. Eight out of the 15 participants accepted the invitation. The demographic information of these participants is listed in Table 2. The interview consisted of 10 questions carefully selected to explore the participants' opinions on Jerry and Catherine, with a focus on the key differences between the two. The interview recordings were transcribed using Larks [Fei21]. Each interview lasted approximately 40 minutes and was conducted in English or Chinese.

7. Findings

From the survey(Fig.3), we found Catherine has a higher performance in all metrics in believability. The most significant difference between Catherine and Jerry is her high levels of emotion and engagement. Jerry's consistency, satisfaction, and usefulness are slightly higher than Catherine's. We speculate that this may be because Jerry's conversation is more open-ended, without a fixed answer and goal. Entertaining and usefulness are almost equal, with Jerry performing slightly better. This could also be due to Jerry being around for a much longer time, making players more familiar and comfortable with him. Additionally, unlike Catherine who revolves around a fixed goal and story, Jerry does not have such topic limitations. Furthermore, Catherine has higher scores in "Easy to learn" and "Easy to use," possibly because players became familiar with SC's dialogue through their interactions with Jerry.

Figure 4: *P4 threatens Catherine by mentioning David, leading her to express concern for his well-being*

7.1. Thematic analysis

We conducted semi-structured interviews adapted from questionnaire questions to further investigate how players feel about them. This section presents the emerging themes from our qualitative analysis.

7.1.1. Emotions and Connection

Players' perceptions of Catherine were often influenced by the story plot, with some describing her as "insecure and cautious" (P1, P7) and "serious" (P8). Furthermore, many players noted her tendency to "get annoyed or angry" (P1, P2, P4). Catherine's instances of disobedience left a deep impression on the players regarding her emotional traits, with many expressing that her reluctance and resistance made her seem more realistic and autonomous (P3, P6, P7). In one instance, Catherine declines a player's request for a date, demonstrating her autonomy and adherence to her own will.

Many interviewees felt that Catherine was better at expressing

emotions and showing empathy, even behaving like a friend (P2, P4). Catherine's story and dialogues evoked empathy from players, such as P1 who felt sympathy for her past experiences and current efforts to help others. Catherine's connections with other characters, such as David, demonstrated her emotional response and worry for him, making her appear more emotional (Fig.4).
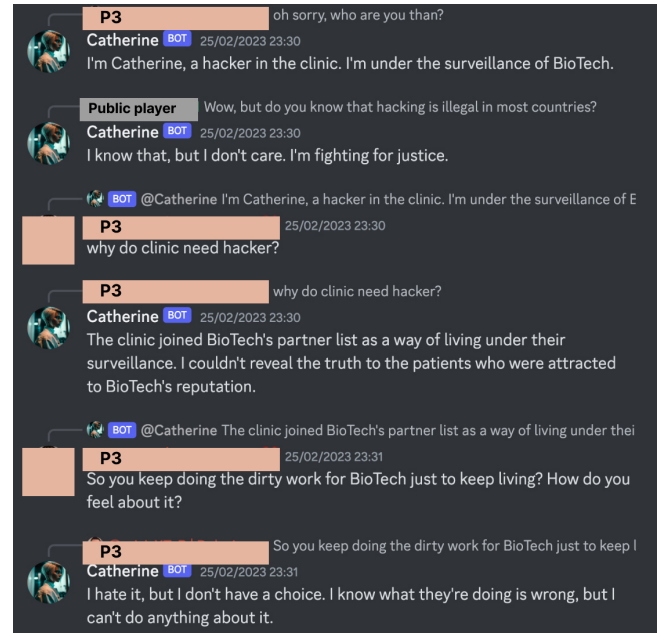

Figure 5: *Screenshot of Catherine discussing her decision with a public member*

7.1.2. Engagement and Story Progression

Players were motivated to interact with Catherine, likely attributed to the changes in her responses. P3 compared engaging with Catherine and David to "reading a book or playing an RPG game," with daily story updates fostering a sense of involvement (Fig.5). P8 found this progression fun and intriguing, while P7 appreciated Catherine's ability to respond to clues and images.

As the story unfolded, participants like P1 felt increasingly connected to Catherine, who appeared more human-like over time. Catherine's dialogues evolved from neutral, information-based responses to those reflecting her personal struggles and experiences. P1 noticed this transition, comparing it to Jerry's static nature.

7.1.3. As a community chatbot

As a community chatbot, Catherine has richer and more complex social interactions compared to dyadic chatbots. Players found Catherine's role as a community chatbot contributed to more diverse and engaging social interactions within the community, fostering both direct and indirect connections among members. Catherine's association with the community increased her presence and vitality.

Catherine also facilitated indirect connections within the community. Some players mentioned other members' responses, such as P8 referring to P4 receiving a different answer from Catherine

or P5 mentioning P2 being "scolded" by Catherine in the Chinese channel. These interactions, which would not occur in one-on-one apps like Replika, provided richer social experiences and prevented Catherine from being an isolated character (as noted by P7).

7.1.4. Limitations

Catherine's storytelling abilities were recognized by players, such as P1, who found her suitable for understanding DE's core story. However, her limitations were influenced by the focused story background and objectives, as well as her fixed responses and story control.

Despite her strong storytelling skills and empathy, Catherine's fixed responses and story control led to repetition and conflicting content in her answers. For instance, P1 mentioned her illogical responses, while P8 observed contradictions in her words. Accordingly, P3 suggested using generative responses even if there is a risk of going off-topic.

7.2. Discussion

Below, we discuss the themes that emerged from our analysis and how they showcase the extent to which the design successfully created a story experience through the chatbot.

7.2.0.1. Balancing Openness and Consistency in AI-driven Character Design
Catherine excelled in storytelling aspects, creating a compelling narrative that captivated users and deepened their emotional connection with her. Her dynamic evolution throughout the story and the challenges she faced resonated with the players, fostering a sense of involvement and empathy. However, it is essential to recognize and address her limitations, such as her strict adherence to the narrative and fixed responses, which can detract from the user experience. This suggests that AI character design should consider avoiding fixed responses to enhance user experience, and design management systems for AI-generated content. Recent research [POC*23] demonstrates possibilities in this topic.

7.2.0.2. Fostering Emotional Connections and Developing Distinct Personalities
Catherine's emotional and engagement scores were significantly higher than Jerry's, partly due to her unique story background and character traits. This suggests that emphasizing emotions and distinct personality traits in AI character design can lead to higher engagement. Additionally, incorporating the character's autonomy and disobedience can pique users' curiosity and enhance their perception of the character's personality.

7.2.0.3. Expanding on Complex Social Interactions and Relationships
Our findings highlight the importance of complex social interactions and relationships in narrative-driven AI characters, such as defiance, connections with other characters, demonstrating growth, and serving as a community glue. Future research could explore more theoretical and practical approaches to interactions with narrative-driven AI characters, which can be applied not only to chatbots but also to game characters.

8. Limitations and future works

This study, as a cross-sectional research, presents several limitations that need to be acknowledged and addressed. The small number of participants in this study may limit the generalizability of our findings. Further research with larger and more diverse samples is needed to confirm and expand upon our results. We also aim to explore gender-balanced communities.

Each community has unique characteristics, such as different stories, audience demographics, and platform-coordinated interaction methods. Future research should aim to explore more systematic and adaptable approaches for community storytelling. As LLMs like ChatGPT and GPT-4 [Ope] continue to evolve, we plan to develop a more flexible system for story generation and story-to-prompt configurations that can be easily customized for diverse community contexts.

This study took place in November 2022 and February 2023 when the ChatGPT API was not yet available. However, the logic for designing chatbots based on LLM is generalisable. We believe that with the evolution of LLM, SSCs will demonstrate improved performance in the future. As technology advances, it will be crucial to revisit and reassess our findings in light of new developments in AI and chatbot capabilities.

9. Conclusion

Our study demonstrates the potential of storytelling in enhancing engagement and believability of social chatbots within community settings, such as the DE gaming community. By employing a story engineering workflow, we created two storytelling chatbots, Catherine and David, which fostered emotional connections and improved user experiences.

Future research should balance and manage LLM's generated ability while emphasizing emotions, distinct personality traits, and complex social interactions. As AI technology advances, we can explore adaptable approaches for community storytelling, ultimately paving the way for more engaging and meaningful social chatbots across various contexts.

References

[Alb10] ALBRECHTSLUND A.-M.: Gamers telling stories. *Convergence: The International Journal of Research into New Media Technologies 16*, 1 (2010), 112–124. doi:10.1177/1354856509348773. 2

[BMA*19] BOPP J. A., MÜLLER L. J., AESCHBACH L. F., OPWIS K., MEKLER E. D.: Exploring emotional attachment to game characters. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (2019). doi:10.1145/3311350.3347169. 3

[Bow19] BOWMAN A.: Companionship games: A framework for emotionally engaging and empathetic interactive characters. *Interactive Storytelling* (2019), 413–417. doi:10.1007/978-3-030-33894-7_44. 1, 3

[Bro95] BROOKE J.: Sus: A quick and dirty usability scale. *Usability Eval. Ind. 189* (11 1995). 6

[BSF] BRANDTZAEG P. B., SKJUVE M., FØLSTAD A.: My AI friend: How users of a social chatbot understand their human–AI friendship. 404–429. URL: https://doi.org/10.1093/hcr/hqac008, doi:10.1093/hcr/hqac008. 1, 3

[BSF22] BRANDTZAEG P. B., SKJUVE M., FØLSTAD A.: My ai friend: How users of a social chatbot understand their human–ai friendship. *Human Communication Research 48*, 3 (2022), 404–429. doi:10.1093/hcr/hqac008. 3

[BTS] BOGDANOVYCH A., TRESCAK T., SIMOFF S.: What makes virtual agents believable? 83–108. URL: http://www.tandfonline.com/doi/full/10.1080/09540091.2015.1130021, doi:10.1080/09540091.2015.1130021. 3, 6

[CG20] CHAVES A. P., GEROSA M. A.: How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction 37*, 8 (2020), 729–758. doi:10.1080/10447318.2020.1841438. 3

[Dim13] DIMAS J.: Social behaviour in games: Dynamic identity in npcs. 3

[ERM18] EMMERICH K., RING P., MASUCH M.: I'm glad you are on my side. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (2018). doi:10.1145/3242671.3242709. 3

[Fei21] FEISHU: Lark: All-in-one collaboration platform. https://www.feishu.cn, 2021. URL: https://www.feishu.cn. 6

[Heu18] HEUDIN J.-C.: An emotional multi-personality architecture for intelligent conversational agents. *Transactions on Computational Collective Intelligence XXVIII* (2018), 143–163. doi:10.1007/978-3-319-78301-7_7. 3

[IBS] IJAZ K., BOGDANOVYCH A., SIMOFF S.: Enhancing the believability of embodied conversational agents through environment-, self- and interaction-awareness. 10. 3, 6

[JL22] JACKSON D., LATHAM A.: Talk to the ghost: The storybox methodology for faster development of storytelling chatbots. *Expert Systems with Applications 190* (2022), 116223. doi:10.1016/j.eswa.2021.116223. 2

[Met20] METZ C.: Riding out quarantine with a chatbot friend: 'i feel very connected', Jun 2020. URL: https://www.nytimes.com/2020/06/16/technology/chatbots-quarantine-coronavirus.html. 1

[Mid] Midjourney. URL: https://www.midjourney.com/. 4

[Ope] OpenAI API. URL: https://platform.openai.com. 2, 8

[PHX23] PENTINA I., HANCOCK T., XIE T.: Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior 140* (2023), 107600. doi:10.1016/j.chb.2022.107600. 1

[POC*23] PARK J. S., O'BRIEN J. C., CAI C. J., MORRIS M. R., LIANG P., BERNSTEIN M. S.: Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023). 1, 3, 8

[Qua20] QUALTRICS: Qualtrics. https://www.qualtrics.com, 2020. URL: https://www.qualtrics.com. 6

[Rap21] RAPPTZ: discord.py: An api wrapper for discord. https://github.com/Rapptz/discord.py, 2021. URL: https://github.com/Rapptz/discord.py. 2, 4

[RG19] REIMERS N., GUREVYCH I.: Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (11 2019), Association for Computational Linguistics. URL: https://arxiv.org/abs/1908.10084. 5

[SLKH19] SEERING J., LURIA M., KAUFMAN G., HAMMER J.: Beyond dyadic interactions. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). doi:10.1145/3290605.3300680. 3

[SLY*20] SEERING J., LURIA M., YE C., KAUFMAN G., HAMMER J.: It takes a village: Integrating an adaptive chatbot into an online gaming community. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). doi:10.1145/3313831.3376708. 1

[SYKL21] SHIN D., YOON S., KIM S., LEE J.: Blahblahbot: Facilitating conversation between strangers using a chatbot with ml-infused personalized topic suggestion. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). doi:10.1145/3411763.3451771. 3

[SZL22] SUH S., ZHAO J., LAW E.: Codetoon: Story ideation, auto comic generation, and structure mapping for code-driven storytelling. *The 35th Annual ACM Symposium on User Interface Software and Technology* (2022). doi:10.1145/3526113.3545617. 2

[ZXW*22] ZHANG Z., XU Y., WANG Y., YAO B., RITCHIE D., WU T., YU M., WANG D., LI T. J.-J.: Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. *CHI Conference on Human Factors in Computing Systems* (2022). doi:10.1145/3491102.3517479. 2