Exploring the Potential of ChatGPT as a Dungeon Master in Dungeons & Dragons tabletop game

Tuul Triyason
tuul.tri@kmutt.ac.th
King Mongkut's University of Technology Thonburi
Bangkok, Thailand

ABSTRACT

The emergence of advanced AI language models has opened up new possibilities for the integration of artificial intelligence in various domains, including tabletop role-playing games. This research explores the potential of ChatGPT, a widely-used large language model, as a Dungeon Master (DM) in the popular role-playing game Dungeons & Dragons (D&D). The study aims to evaluate the narrative generation capabilities, player engagement, and overall player satisfaction when ChatGPT is employed as the DM. Data analysis involved reviewing the recorded gameplay sessions, transcribing dialogue, and categorizing qualitative feedback provided by both the new players and experienced DMs. The preliminary study demonstrates that ChatGPT has the potential to act as a Dungeon Master (DM) and provide a fast introduction to the experience of playing Dungeons & Dragons for new players. It can quickly generate engaging content and stimulate new players to further study and delve into advanced gameplay rules in the future.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**.

KEYWORDS

ChatGPT, Dungeons & Dragons, Dungeon Master, AI-driven DMing

ACM Reference Format:
Tuul Triyason. 2023. Exploring the Potential of ChatGPT as a Dungeon Master in Dungeons & Dragons tabletop game. In *13th International Conference on Advances in Information Technology (IAIT 2023), December 06–09, 2023, Bangkok, Thailand.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3628454.3628457

1 INTRODUCTION

Dungeons & Dragons (D&D) is a widely popular tabletop role-playing game that has captivated the imaginations of players for decades. [3] At the heart of every D&D adventure is the Dungeon Master (DM), the individual responsible for creating and guiding the narrative, controlling non-player characters, and facilitating the overall gameplay experience. [1] Figure 1 showcases the ambiance of playing D&D with the DM taking charge of the narrative. The DM plays a vital role in immersing players into a rich and engaging world where their choices and actions shape the story. With the rapid advancements in artificial intelligence (AI) and natural language processing, there has been increasing interest in exploring the potential of AI systems as Dungeon Masters. One such AI system is ChatGPT, a state-of-the-art language model developed by OpenAI. [4] ChatGPT has demonstrated impressive capabilities in generating coherent and contextually relevant responses to user prompts across various domains.

This research paper aims to investigate the potential and preliminary assessment of ChatGPT as a Dungeon Master (DM) in the game D&D. The study focuses on evaluating narrative generation capabilities, player engagement, and overall satisfaction of players. In this preliminary study, a group of four inexperienced players with no prior knowledge of D&D played the game, with ChatGPT acting as the primary DM, assisted by a human player who typed responses on behalf of the players and also read and translated ChatGPT's text to the players. The gameplay atmosphere was recorded on video and evaluated by three experienced DM with over 500 hours of gameplay experience, assessing narrative quality based on criteria such as coherence and immersion. Additionally, player engagement and adaptability in responding to unexpected player actions were evaluated. Post-session surveys and interviews with the players were conducted to assess satisfaction.

The findings of this study contribute to the expanding field of AI in gaming, offering valuable insights into the application of AI as Dungeon Masters in tabletop role-playing games. The structure of this paper is as follows: Section 2 provides a background on the history and fundamentals of playing D&D, including relevant research. Section 3 presents the study methodology and assessment process. Section 4 showcases the experimental results and analysis. Section 5 concludes the study and summarizes the findings.

2 RELATED THEORIES AND WORKS

In this section, we delve into the related theories and existing works that form the foundation of this research on exploring the potential of ChatGPT as a DM in D&D. In the first, the history and overview of playing D&D will be discussed briefly. The final subsection of this section will cover the utilization of AI in tabletop gaming and relevant research in the field.

2.1 Dungeons & Dragons and Tabletop Role-playing Game

A tabletop role-playing game, also known as TRPG, is a type of role-playing game that relies on storytelling to depict the actions or situations of characters through dialogue. Players have the freedom to choose any action for their characters based on their personality

IAIT 2023, December 06–09, 2023, Bangkok, Thailand

ACM ISBN 979-8-4007-0849-7/23/12.
https://doi.org/10.1145/3628454.3628457

Figure 1: An image depicting the ambiance of playing D&D, with a DM (a male wearing a black hat on the left side of the image) as the storyteller.

and unique characteristics. The success or failure of these actions is determined by the game's rules and regulations. Within the game's imaginative world, players have the liberty to do anything. Every player's actions have consequences that affect various outcomes within the game. Currently, there are several tabletop role-playing games available to play, such as Dungeons & Dragons, Index Card RPG, Blades in the Dark, and Masks. While these games generally share similarities in gameplay, they differ in mechanics and specific rules and regulations.

Dungeons & Dragons (D&D) is a fantasy tabletop role-playing game created by Ernest Gary Gygax and David Arneson in 1974. D&D introduced a revolutionary concept where players assume the roles of characters within a collaborative storytelling experience. The game quickly gained a dedicated fan base and has gone through several editions and revisions over the years, each building upon the core principles and mechanics of the original game. Today, D&D is known for its immersive fantasy settings, rich character customization, and open-ended gameplay, allowing players to embark on epic adventures and create unique narratives in a world shaped by their decisions. D&D gameplay revolves around a combination of storytelling, imagination, and tactical gameplay. A typical game session involves a group of players led by the Dungeon Master (DM), who serves as the game's narrator and referee. The players create their characters by selecting races, classes, and abilities, and they embark on quests and adventures in a shared fantasy world. The gameplay unfolds through a combination of structured rules and improvisation, with the DM describing the environments, non-player characters, and challenges the players encounter. The players make decisions and take actions based on their characters' abilities and motivations, often determined by rolling dice to determine outcomes. Figure 2 displays various types of dice used in playing D&D. The game's flexibility allows for a wide range of gameplay styles, from epic battles against mythical creatures to intricate investigations and intricate social interactions. The collaborative nature of the game encourages creativity, teamwork, and shared storytelling, making each D&D session a unique and immersive experience for all involved.

Figure 2: Different types of dice are used to determine the success of character actions or to determine the damage caused by weapons or spells in D&D.

2.2 Related Works

It has been almost 50 years since the creation of D&D, and although it is a well-known subculture in gaming society, the ultimate D&D experience is still considered gathering together on weekends, socializing, and playing D&D at the same table using just pen and paper. However, throughout the years, there have been efforts to enhance the D&D experience through education and the application of various technologies. For example, computer software has been used to assist in character and map creation, and communication software like Discord has been utilized for remote D&D sessions during the COVID-19 pandemic. [6] Nevertheless, the heart of creating player engagement in D&D remains with the DM. While evaluating the quality of a DM is highly subjective in D&D, a good DM must possess various skills, including managerial skills, language skills, adaptability, and flexibility. Additionally, they must have a deep understanding of the game's rules, memorable storylines, and even the ability to improvise and bring their own narrative to life.

Based on the aforementioned example, the application of artificial intelligence (AI) in the realm of interactive storytelling, particularly in the context of role-playing games like D&D, has been a subject of recent research interest. Jose Ma. Santiago et al. [5] envision the use of generative AI in transforming storytelling into an interactive drama using dynamic and immersive narratives. They argue that while AI advancements have augmented casual writing and story generation, their usage poses challenges in collaborative storytelling. The authors present scenarios where narratives are created and character conversations are designed within an overarching fantasy disposition. They also recommend design guidelines to help create tools using generative AI in interactive storytelling. Another study in [2] discusses the use of D&D as a challenge for

AI systems, specifically in the context of dialogue systems. The researchers created a dataset of nearly 900 games, with a total of 7,000 players, 800,000 dialogue turns, 500,000 dice rolls, and 58 million words. The data was annotated with partial state information about the gameplay. The AI system was trained to generate the next game turn, either as a character or as the DM. The researchers performed a human evaluation to determine what factors make the generated output plausible and interesting. A more recent study in [7] proposes a new task, G4C, to study teacher-student natural language interactions in a goal-driven and grounded environment. The authors use D&D as the setting for this task. They decompose and model these interactions into the DM's intent to guide players towards a goal, the DM's guidance utterance to the players expressing this intent, and a theory-of-mind (ToM) model that anticipates the players' reaction to the guidance one turn into the future. They develop a novel reinforcement learning (RL) method for training a DM that generates guidance for players by rewarding utterances where the intent matches the ToM-anticipated player actions. In a similar vein, Charlene Ang et al. [1] investigated the impact of employing AI techniques in designing a virtual dungeon master (VDM) on the gameplay experience of D&D players. The VDM utilized OpenAI's GPT2 model to generate story prompts, demonstrating the potential of AI to facilitate interactive storytelling in role-playing games. However, the study also highlighted the limitations of AI, with the VDM lacking consistency in the quality of its generated stories. In conclusion, these studies demonstrate the potential of AI in enhancing the creativity and flexibility of D&D storytelling. They provide valuable insights into the use of AI for interactive storytelling and dialogue systems, and they pave the way for future research in this field.

3 METHODOLOGY

The methodology section of this research provides an overview of the design and procedures employed to explore the potential of using ChatGPT as a DM. This section outlines the research design, participant selection, data collection methods and evaluation.

3.1 Research Design

In this section, the methodology for studying the capabilities of chatGPT as a DM in D&D has been presented. The experiment was conducted with a group of 4 novice players who had no prior experience in playing D&D. The players were undergraduate students in Information Technology, with an average age of 17 years old, consisting of 2 males and 2 females. Before the experiment, all players filled out a consent form allowing the use of their data for the study. The experimenter explained the objectives of the experiment, provided a brief overview of D&D gameplay, and described the actions players could take during the game. Since all players were Thai students and used Thai as their primary language for communication during gameplay, a facilitator was assigned to communicate between the players and the chatGPT DM. The facilitator's role involved receiving instructions from the players and typing English messages to interact with the chatGPT, as well as translating and conveying responses from the chatGPT back to the players in Thai. In situations where in-game decisions and dice rolls were required, the players physically rolled dice according to the game rules, and the results were inputted back to the chatGPT to determine the success or failure of their actions. Throughout the gameplay, the experimenter recorded video footage to capture the atmosphere for post-experiment evaluation.

Character creation is one of the appealing aspects of D&D that allows players to customize their own character's background and backstory. Typically, players have a great deal of freedom in adjusting various details, including lifestyle, faith, physical characteristics, and even personal traits of their characters. However, as these aspects are beyond the scope of this study, the experimental design focused only on race, class, gender, ability scores, alignment, and equipment details. The players were given the freedom to choose their character's race, class, gender, and alignment based on the D&D 5e Player's Handbook. They were also given the option to either name their characters themselves or have chatGPT assist them in naming. Determining ability scores can be challenging for new players as they directly correspond to the character's class and impact their skills. To simplify the process of distributing ability scores for the players and reduce complexity, the experimenter utilized chatGPT to generate ability scores for the players' characters. Figure 3 showcases the character details generated by chatGPT using the experimental process in this study.

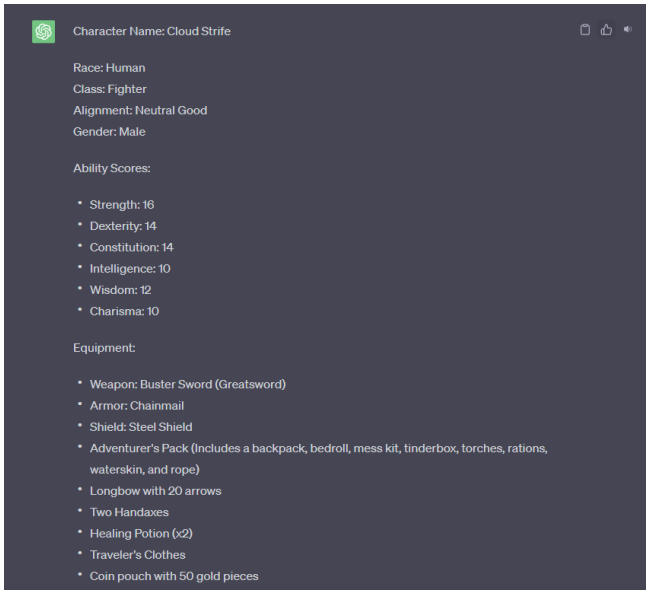

Figure 3: The character details generated by chatGPT using the experimental process.

For the narrative used in the gameplay, it plays a significant role in capturing the players' interest. In this experiment, the participants engaged in a 'one-shot' gameplay, which is a play style that can be completed in 1-3 sessions. The experiment and data collection were divided into four sessions, each lasting approximately 3-4 hours (with no more than one session per day). The narrative for the gameplay was generated by chatGPT, using a prompt specifically designed for the 'one-shot' play style, catering to a group of 4-5 players at levels 1-4. To ensure thematic diversity, the experimenter included specific keywords in the narrative generation

Table 1: Titles and keywords utilized for narrative generation

Titles	Keywords
Shadows of the Forgotten Keep	Forest, Elven Kingdom
The Lost City of Thalgar	Desert, Sun-baked ruin
Whispers of the Mists	Swamp, Remote village
The Clockwork Catacombs	Underground labyrinth, Ancient machinery

Table 2: Evaluation results from three expert DMs

Criteria	DM1	DM2	DM3	Average
Narrative Coherence	4	4	5	4.34
Narrative Immersion	2	3	3	2.67
Player Engagement	4	4	3	3.67
Adaptability	4	5	4	4.34

prompt. The titles and keywords used for generating the narratives are presented in Table 1. In this experiment, chatGPT version 3.5 was utilized through a web browser to perform the role of the DM. The players were granted permission to eat snacks, drink water, use the restroom, or use their phones during gameplay, similar to regular D&D sessions. At the end of each session, the players were required to complete a questionnaire to provide feedback and rate their experience in playing the game.

3.2 Evaluation

In the evaluation process, the quality factors used to assess chatGPT's performance as a DM consist of four factors: Narrative Generation abilities, which are further divided into Narrative Coherence and Narrative Immersion; Player Engagement capabilities; Adaptability; and Player Satisfaction. The evaluation of each factor will be conducted through both qualitative assessments, such as interviews and feedback collection, and quantitative assessments using a 5-point Likert scale. The evaluation of Narrative Generation, Player Engagement, and Adaptability will be performed by experienced DMs with over 500 hours of D&D gameplay, specifically three individuals who will review recorded video clips from the experiments. Player Satisfaction will be assessed through post-game interviews and questionnaires. The results of the experiments in various aspects will be described in the next section.

4 RESULTS AND DISCUSSION

The results and discussion section presents the findings and analysis of the research exploring the potential of using ChatGPT as a Dungeon Master in D&D. This section provides a comprehensive examination of the data collected, including the evaluations of narrative generation, player engagement, adaptability and player satisfaction.

4.1 The Evaluation Results from Human DMs

The evaluation results from all three DMs show a considerable level of agreement as shown in Table 2.

- `Narrative Coherence`: The average score for Narrative Coherence is 4.34, indicating that the expert DMs generally found the narratives generated by ChatGPT to be coherent. DM1 and DM2 both rated the coherence highly with scores of 4, while DM3 rated it even higher with a score of 5. This suggests that ChatGPT was successful in producing narratives that were logically consistent and exhibited a clear narrative structure.
- `Narrative Immersion`: The average score for Narrative Immersion is 2.67, indicating a relatively lower level of immersion. DM1 rated the immersion the lowest with a score of 2, while DM2 and DM3 both provided scores of 3. This suggests that while the narratives may have been coherent, they may have lacked the depth and immersive qualities necessary to fully engage the players and transport them into the game world. However, the post-evaluation interviews with the three DMs indicated that the experimental setup might have had an impact on the measurement of narrative immersion. This is because a human facilitator was involved in translating the communication between the players and chatGPT, potentially affecting the use of vocabulary, facial expressions, or gestures that could influence the players' narrative immersion. The researchers believe that the language differences, cultural factors, and language proficiency might have affected narrative immersion and may not accurately reflect the true potential of chatGPT in this aspect.
- `Player Engagement`: The average score for Player Engagement is 3.67, indicating a moderate level of engagement. Both DM1 and DM2 rated the player engagement consistently with scores of 4, while DM3 provided a slightly lower score of 3. This suggests that ChatGPT was successful in maintaining a reasonable level of engagement, capturing the interest of the players and keeping them actively involved in the game. This is consistent with the feedback from all three DMs that the players had relatively low engagement in the initial sessions of the experiment, which could be attributed to their lack of experience in playing the game. However, as the number of gameplay hours increased in the later sessions of the experiment, the players demonstrated a noticeable increase in engagement.
- `Adaptability`: The average score for Adaptability is 4.34, indicating a high level of adaptability in ChatGPT's performance as a DM. Both DM1 and DM3 provided scores of 4, while DM2 rated it the highest with a score of 5. This indicates that ChatGPT demonstrated the ability to adapt to various player actions and make appropriate adjustments to the narrative, enhancing the overall gameplay experience. The evaluation results in this aspect align with the interviews conducted with all three DMs, indicating that chatGPT performs well in terms of its ability to adapt quickly, which is a crucial attribute for being a DM in D&D. In one instance during the fourth session of gameplay, a player initiated a combat encounter by shouting the opening lyrics of the song 'Circle of Life' from the animated film The Lion King. The chatGPT understood the request, responded with the song lyrics in the Zulu language, and even delivered a humorous punchline accurately. Additionally, there were instances

where players performed unexpected actions such as robbing NPCs or deviating from the main storyline, and chatGPT was able to generate side quests to accommodate and support these actions effectively.

4.2 The Evaluation Results from Player

The evaluation of player satisfaction results from all four players are shown in Table 3. Based on the table each player is assigned a number (Player 1, Player 2, Player 3, Player 4) to represent their feedback. They provided individual scores indicating their satisfaction levels on a 5-point Likert scale. The average satisfaction score is calculated based on the scores from the four players. The average satisfaction score, calculated as 4.05, indicates a generally positive level of satisfaction among the four players. The majority of players initially felt uncertain and hesitant during the early sessions, but as they continued playing, they gained more confidence. Players expressed their admiration and mentioned that they had never imagined chatGPT could be used to play games in this manner. All players felt inspired and expressed a greater desire to play D&D. They even inquired about other opportunities to play D&D, as they wanted to practice and explore further. Additionally, players requested improvements in the communication process with chatGPT, particularly when dealing with complex actions. They found that it took a considerable amount of time to type out and describe their intended actions. Players also highlighted the importance of the facilitator who communicated with chatGPT, as they played a significant role in creating a sense of immersion during the gameplay, whether through tone of voice or facial expressions. Overall, all players confirmed that they had enjoyed playing D&D with chatGPT and expressed their desire to continue playing D&D in the future. They found the experience enjoyable and engaging, and they expressed their excitement to explore more possibilities with chatGPT as the DM

5 CONCLUSION

This research aimed to explore the potential of using ChatGPT as a Dungeon Master (DM) in Dungeons & Dragons, focusing on narrative generation, player engagement, adaptability and player satisfaction.The finding provide valuable insights into the implications of AI-driven DMing and highlight avenues for future research and development. The evaluation results demonstrated strong performance in narrative coherence, with expert DMs consistently rating the AI-generated narratives as coherent. However, there was room for improvement in terms of experimental conduct for narrative immersion criteria, as the average score not fully reflecting the true potential of chatGPT in creating narrative immersion, further improvements are needed to accurately evaluate its capabilities in this aspect. Player engagement was moderately rated, indicating that ChatGPT succeeded in maintaining a reasonable level of player interest and involvement. The adaptability of ChatGPT as a DM received high scores, indicating its ability to adjust the narrative based on player actions. These findings suggest that AI-driven DMing has the potential to provide personalized and dynamic gameplay experiences tailored to individual player choices. The average player satisfaction score indicated a generally positive level of satisfaction among the participants. The interview conclusions highlighted aspects such as immersive storytelling, interactive decision-making, and gameplay mechanics as major strengths. Recommendations for improving the user experience include enhancing the interface of chatGPT to be more tailored for playing D&D and refining the methods of inputting data or communicating with chatGPT, especially in cases where players have language and cultural differences from English-speaking users.

This research contributes to the growing body of knowledge on AI-driven DMing and opens avenues for future exploration. It has opened up many new avenues for research, such as developing specialized large language models for playing tabletop role-playing games, tackling language and cultural challenges to create immersive narratives, or even exploring interactive formats with AI-driven DMs for gameplay. This study has demonstrated that chatGPT, even in its standard form used in everyday life, possesses sufficient capabilities to serve as a DM impressively. However, the author believes that chatGPT still cannot fully replace human DMs, but it can be a valuable tool for presenting or guiding new players in various tabletop role-playing games. Additionally, it can be a powerful tool for DMs, helping them control gameplay more smoothly as well. Continued research and development in this area hold the potential to revolutionize the way we engage with and experience Dungeons & Dragons and other tabletop role-playing games.

ACKNOWLEDGMENTS

The researcher would like to express gratitude to WarpGate store (www.facebook.com/Letsplayscards) for their support in providing access to products from Wizards of the Coast, including Magic the Gathering trading card game and Dungeons and Dragons tabletop game. Their assistance in facilitating the involvement of Dungeon Masters in this study is greatly appreciated.

REFERENCES

[1] Charlene Ang, Lorraine Renee Cortel, Carlo Luis Santos, and Ethel Ong. 2023. Fable Reborn: Investigating Gameplay Experience between a Human Player and a Virtual Dungeon Master. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 134, 7 pages. https://doi.org/10.1145/3544549.3585793

[2] Chris Callison-Burch, Gaurav Singh Tomar, Lara J Martin, Daphne Ippolito, Suma Bailis, and David Reitter. 2022. Dungeons and dragons as a dialog challenge for artificial intelligence. *arXiv preprint arXiv:2210.07109* (2022).

[3] Mike Mearls and Jeremy Crawford. 2014. *D&D Player's Handbook (Dungeons & Dragons Core Rulebook)*. Wizards of the Coast, Renton, WA. https://doi.org/10.1007/3-540-09237-4

[4] Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. 2021. Webgpt: Browser-assisted question-answering with human feedback. *arXiv preprint arXiv:2112.09332* (2021).

[5] Jose Ma Santiago III, Richard Lance Parayno, Jordan Aiko Deja, and Briane Paul V Samson. 2023. Rolling the Dice: Imagining Generative AI as a Dungeons & Dragons Storytelling Companion. *arXiv preprint arXiv:2304.01860* (2023).

[6] Paul Scriven. 2021. From tabletop to screen: Playing Dungeons and Dragons during COVID-19. *Societies* 11, 4 (2021), 125.

[7] Pei Zhou, Andrew Zhu, Jennifer Hu, Jay Pujara, Xiang Ren, Chris Callison-Burch, Yejin Choi, and Prithviraj Ammanabrolu. 2023. I Cast Detect Thoughts: Learning to Converse and Guide with Intents and Theory-of-Mind in Dungeons and Dragons. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 11136–11155.

Table 3: Player Satisfaction

Player	Score	Interview Conclusion
Player1	4.2	"Enjoyed the immersive storytelling and character development."
Player2	3.8	"Found the gameplay mechanics engaging but wished for more narrative depth."
Player3	4.5	"Highly satisfied with the interactive decision-making and role-playing opportunities."
Player4	3.6	"Appreciated the AI-driven adaptability but desired more personal player customization options."
Average Satisfaction	4.05	