

# Chatbot-Based Board Game Recommendation System

Sheila Acar

MSDS 453 - Natural Language Processing

Sunday, December 8, 2024

## **Introduction**

This paper introduces the process of developing a chatbot recommendation system for board game players. The dataset contains data collected from BoardGameGeek, the largest online repository of board game information, which consists of data on more than 100,000 games as of February 2021. The collected data includes details such as the name of the game, year of publication, minimum and maximum recommended number of players, playing time, recommended minimum player age, average rating received, BGG rank, average complexity value, number of registered BGG owners, mechanics used in the game, and the board game domains to which the game belongs. The goal of this research is to develop a tool that provides board game recommendations based on factors such as playing time, number of players, or theme. These experiments focus on enhancing the chatbot to ensure it is accessible and useful for players of all levels and serve as a comprehensive guide for both board game enthusiasts and those who are new to board games.

## **Research Design and Modeling Methods**

### Cleaning:

Using Python, I cleaned the dataset by removing columns that were unnecessary for the analysis, such as "ID" and "Users Rated." Next, I removed outliers based on the number of owned users. The 25th percentile for owned users is 146, which I considered a threshold for identifying unpopular games, especially when compared to the mean of 1,408 owned users. I dropped games with fewer than 146 owned users.

Then, I examined the descriptive statistics and noticed that the maximum value for play time was 6,000 minutes. To investigate further, I created a boxplot of the data. Based on the boxplot, I decided to drop rows with play times exceeding 1,000 minutes, as they appeared to be

extreme outliers. Finally, I addressed missing data by dropping rows with null values and removed duplicate entries from the data frame.

To enrich the dataset even further, I added a new column containing board game summaries. Initially, I attempted to scrape summaries from Wikipedia. While this approach provided meaningful text for some games, many entries resulted in "No data found" due to missing or mismatched game pages. To gain more information, I expanded the scraping process to include a fallback to Google Search, which retrieved snippets for games not found on Wikipedia. Despite this enhancement, many rows still lacked sufficient data. Finally, I focused solely on scraping from BoardGameGeek(BGG). This approach provided the most relevant and structured summaries for most of the games and significantly improved the completeness and quality of the dataset. Some rows still contained missing data, but overall, I was satisfied with the results.

After cleaning the original dataset and adding an extra column with more textual data, I used ChatGPT to generate text summaries for each game based on all its columns. For example, the summary for 7 Wonders Duel reads: “The game is called 7 Wonders Duel. It features mechanics such as Card Drafting, Drafting, Set Collection, Sudden Death Ending, and Tug of War. The game belongs to the domains of Strategy Games. It is designed for a playtime of 30 minutes. The recommended minimum age for players is 10 years, and it supports 2 players. The game is ranked 17 on BoardGameGeek”.

#### Chatbot Development:

Using the cleaned dataset, I decided to develop, compare, and contrast two different chatbot models. The first approach uses Llama Index and OpenAI’s GPT API to implement a semantic search-based chatbot. Each board game summary was transformed

into Document objects, which are indexed using Llama Index's VectorStoreIndex. This index supports efficient semantic querying by converting text into vector embeddings that capture the contextual meaning of words. The queries are passed to OpenAI's GPT API, which generates high-quality natural language responses. This cloud-based approach is easy to scale and uses GPT models' advanced training to deliver clear, context-aware responses.

The second approach takes a local deployment strategy and combines text preprocessing techniques from the Natural Language Toolkit (NLTK) with Hugging Face's Llama 2 model for response generation. The dataset undergoes preprocessing, including tokenization, lemmatization, stopword removal, and TF-IDF vectorization to identify relevant sections of text for user queries. A similarity-based retrieval system matches user input with the most relevant summaries, calculated using cosine similarity on TF-IDF vectors. The Llama 2 model, downloaded and deployed locally using the llama-cpp-python library, then generates responses by augmenting the retrieved context.

## **Results**

To evaluate the two approaches, I asked both models five different questions. In terms of detail, ChatGPT provided concise and straightforward responses, but these responses often lacked depth. For example, as shown in Figure 3, ChatGPT's answer to "Tell me about Catan" was significantly shorter compared to Llama 2's more detailed response. Similarly, in Figure 5, ChatGPT's reply to the prompt, "I want to play a cooperative game with 5 players, and we only have one hour," consisted of a single sentence, whereas Llama 2 provided a more comprehensive, paragraph-long explanation.

These succinct responses resulted in ChatGPT offering a single suggestion per query, whereas Llama 2 provided multiple recommendations and a variety of games. For example, in

Figure 1, the prompt “Can you recommend a highly rated family game for 4 people that is less than 30 minutes?” was met with ChatGPT suggesting only one game, while Llama 2 recommended several board game options, including *Ticket to Ride*, *Carcassonne*, and *Sushi Go!*.

Regarding the accuracy of recommendations, ChatGPT sometimes provided suggestions that did not align with the query's constraints. For instance, in Figure 1, the prompt “Can you recommend a highly rated family game for 4 people that is less than 30 minutes?” was answered with “*Famiglia*, which is a highly rated game but designed for only 2 players”, which failed to meet the specified requirement for 4 players. Another example of inaccurate recommendations is seen in response to the question, “Can you recommend an abstract game with high ratings?” ChatGPT suggested *High-Bid*, a board game with a relatively low rating of 5.78, which is not a strong contender for a highly rated abstract game. Furthermore, it classified the game under the domain of Strategy Games, which does not directly address the query for an abstract game.

Llama 2 provided more relevant and detailed recommendations and offered multiple options along with additional context. For example, in response to the question, “I want to play a cooperative game with 5 players, and we only have one hour,” Llama 2 recommended *Pandemic*. It explained that, “*Pandemic* is a cooperative game where players work together to stop the spread of diseases and save the world. Although it is designed for 2-4 players, it can accommodate 5 players with slight modifications. The game has an estimated play time of 45-60 minutes, which fits the specified time frame.” Furthermore, Llama 2 highlighted that *Pandemic* encourages teamwork and communication, which makes it a great choice for fostering a sense of camaraderie among players.

## **Analysis and Interpretation**

In terms of the depth and detail of the responses, the variety of board game options provided, and the relevance of recommendations, Llama 2 significantly outperformed ChatGPT. Several factors may have contributed to this disparity. First, the difference in response quality suggests that ChatGPT may have gaps or rely on older data in its training corpus. In contrast, Llama 2's training likely included exposure to more recent and comprehensive datasets that encompasses board game ratings, mechanics, and community preferences. This broader knowledge base may have enhanced its ability to provide accurate and contextually appropriate recommendations.

Another contributing factor is Llama 2's compatibility with tools like TF-IDF, which allows it to effectively prioritize relevant information and integrate it into responses. This capability enables Llama 2 to provide multiple recommendations or elaborate on topics when necessary. For instance, in the family game query, Llama 2 not only suggested multiple games but also provided additional details about their gameplay.

Third, Llama 2's architecture is especially well-suited for open-ended and contextual queries since it is designed to process multiple pieces of information and synthesize them into a coherent and detailed narrative. This capability allows it to generate paragraph-long responses that provide substantial depth and context. In contrast, ChatGPT is effective at handling simpler prompts but tends to prioritize speed, which leads to sacrificing depth and relevance in its responses.

Overall, Llama 2's superior performance is attributed to its broader and more recent training data, ability to synthesize detailed and relevant responses, and advanced architecture. It excels in tasks that require contextual understanding, domain expertise, and engaging

explanations. This makes it a more effective tool for generating recommendations tailored to complex user queries.

## **Conclusions**

This research highlights the development and evaluation of two chatbot models designed to recommend board games based on user preferences. Through a detailed comparison, Llama 2 demonstrated superior performance over ChatGPT in providing accurate, detailed, and contextually relevant recommendations. Llama 2's broader and more recent training data, integration with tools like TF-IDF for prioritizing information, and advanced architecture make it better suited for handling complex, open-ended queries. These strengths allow Llama 2 to generate detailed explanations and offer multiple relevant options which makes it a more effective and user-friendly tool for both experienced board game enthusiasts and new players alike.

## **Directions for future work**

Future work should focus on addressing the missing data in board game summaries by developing more robust web scraping techniques. Another potential approach could involve using natural language generation models to create summaries based on available metadata for games with incomplete information. In addition, the data cleaning process can be further refined. While the final dataset contained over 8,000 games, its relevance and usability could be enhanced by filtering out games that are outdated or lack significant popularity. This additional filtering would reduce the dataset size, streamline performance, and improve the overall user experience by ensuring that recommendations focus on contemporary and widely played games. Finally, incorporating advanced preprocessing methods or domain-specific feature engineering could further optimize the chatbot's recommendation capabilities.

## Appendix

### Figure 1.

CHATGPT:

**Question:** Can you recommend a highly rated family game for 4 people that is less than 30 minutes?

**Answer:** 'Famiglia' is a highly rated family game for 2 players that can be played in less than 30 minutes.

LAMMA 2:

**Question:** Can you recommend a highly rated family game for 4 people that is less than 30 minutes?

**Answer:** If you're looking for a highly rated family game that can be played with 4 people in under 30 minutes, I would recommend "Ticket to Ride." It's a popular train-themed game that is easy to learn and has a lot of replay value. The game involves collecting train cards and claiming routes on a map of North America, and it's a great way to spend time with family and friends.

Other highly rated games that are similar in duration and playability include "Sushi Go!" and "Carcassonne." Both of these games are simple to learn and have a lot of depth, making them great options for families with gamers of all ages.

Ultimately, the best game for you will depend on your family's interests and playstyle, but I hope these recommendations are helpful in your search for a fun and engaging game to play with 4 people in under 30 minutes.

### Figure 2.

CHATGPT:

**Question:** I have 6 players, and we want to play a party game.

**Answer:** Junior' would be a suitable party game for 6 players.

LAMMA 2:

**Question:** I have 6 players, and we want to play a party game.

**Answer:** Based on the information provided, the best game from the Penguin Party collection for 6 players to play at a party would be "Penguin Party" itself! This game is designed for 2-6 players and is estimated to take around 15 minutes to play, making it a great option for a quick and fun party game.

Figure 3.

CHATGPT:

**Question:** Tell me about Catan

**Answer:** Catan is a board game designed for 3-4 players, with an estimated play time of 120 minutes. It is recommended for ages 10 and up. The game has an average rating of 7.15 and is ranked #395 on BGG. It has a complexity rating of 2.32. With 154531 registered owners, this game features mechanics such as Dice Rolling, Hexagon Grid, Income, Modular Board, Network and Route Building, Race, Random Production, Trading, Variable Set-up and belongs to the domain(s) Family Games, Strategy Games. Summary: Collect and trade resources to build up the island of Catan in this modern classic.

LAMMA 2:

**Question:** Tell me about Catan

**Answer:** Sure! Catan, also known as The Settlers of Catan, is a popular board game designed by Klaus Teuber and first released in 1995. The game is set on an island where players collect resources, build settlements, and trade with each other to build the best kingdom. The game has evolved over time through various expansions and editions, each adding new mechanics and elements to the game. Here are some key ways in which Catan has evolved over time: 1. Original Game (1995): The original game focused on collecting resources such as brick, wood, sheep, wheat, and ore. Players used these resources to build settlements, roads, and cities, and the player with the most points at the end of the game was declared the winner. 2. 5th Edition (2006): This edition introduced the concept of development cards, which gave players additional abilities and bonuses. The game also introduced a new resource, gold, which could be used to buy development cards or other resources. 3. Seafarers Expansion (2004): This expansion added ships and maritime trade to the game, allowing players to trade resources with other players at sea. 4. Cities and Knights Expansion (2015): This expansion added new mechanics such as city improvements, knight cards, and the ability to build cities on water. 5. Catan: Seas of Titans (2014): This edition introduced a new map and gameplay elements such as sea battles, naval ports, and the ability to build on water. 6. Catan: Junior (2013): This game is a simplified version of Catan designed for younger players. It has fewer resources and a shorter gameplay time. 7. Catan: The Card Game (2017): This deck-building game allows players to collect and trade resources, build settlements, and battle each other. 8. Catan: Rise of the Robbers (2020): This game introduces new mechanics such as robber.

Figure 4.

CHATGPT:

**Question:** Can you recommend an abstract game with high ratings?

**Answer:** 'High-Bid' is a board game with a relatively high rating of 5.78 and is classified under the domain of Strategy Games.

LAMMA 2:

**Question:** Can you recommend an abstract game with high ratings?

**Answer:**

- A) Azul
- B) Sushi Go!
- C) Carcassonne
- D) Ticket to Ride

Figure 5.

CHATGPT:

**Question:** I want to play a cooperative game with 5 players, and we only have one hour.

**Answer:** You can play '5-Minute Dungeon' with 5 players in one hour.

LAMMA 2:

**Question:** I want to play a cooperative game with 5 players, and we only have one hour.

**Answer:** Pandemic is a cooperative game that requires players to work together to stop the spread of diseases and save the world. It is designed for 2-4 players but can be played with as few as 3 players. The game has an estimated play time of 45-60 minutes, which is well within the 1-hour time frame you mentioned. Additionally, Pandemic is a game that encourages teamwork and communication, which can help foster a sense of camaraderie among players.

## References:

BoardGameGeek. "Search: Game." Accessed December 8,

2024. <https://boardgamegeek.com/geeksearch.php?action=search&objecttype=boardgame&q=game>.

Dilini Samarasinghe. "BoardGameGeek Dataset on Board Games." doi: 10.21227/9g61-bs59

Hugging Face. 2024. *Llama 2 Model by Meta (Version used: Llama 2-7B Chat GGUF)*.

Accessed December 8, 2024. <https://huggingface.co/TheBloke/Llama-2-7B-Chat-GGUF>.

Lane, H., C. Howard, and H. M. Hapke 2019. *Natural Language Processing in Action:*

*Understanding, Analyzing, and Generating Text with Python*. Shelter Island, N.Y.: Manning.

OpenAI. 2024. *ChatGPT (December 8 version)*. Accessed December 8,

2024. <https://chat.openai.com>.

Tunstall L., Werra L., and Wolf T. 2022. *Natural Language Processing with Transformers*.