Exploring the Therapeutic Potential of Character.ai: An Analysis of the Community Interactions through Reddit

Martina D'Angelo

09 September 2024

Abstract

This research proposal aims to comprehensively assess public perception of Character.ai on Reddit. Character.ai is a platform that enables users to chat with various artificial intelligence characters. The dataset was collected from Reddit, focusing on discussions and questions about the platform.

We analyzed the demographics of the subreddit, exploring users' ages, interaction patterns with the characters, and their motivations for engaging with the platform. To examine public attitudes, we conducted a content analysis using topic modeling with the Latent Dirichlet Allocation (LDA) algorithm to identify key themes.
We wanted to understand if the platform is used for any kind of therapy related conversations and how much this influences the overall use.
Our topic modeling results revealed three main themes: user escapism interactions, support-based interactions, and assistance interactions. Within the therapy-related discussions, the term *comfort* emerged as particularly significant. A Principal Component Analysis (PCA) underscored the significance of the word within the context of this subreddit, illustrating how comfort-related data is consistent with the dataset. Subsequently, we created a semantic network to map out the associations and relationships of the word "comfort" within the context of Character.ai comfort-based interactions.

1 Introduction

Over the past years, there has been a growing interest in Generative AI, particularly in text-based AI-chatbots software applications that interact with humans using natural written language.

Character.ai is a platform where users can chat with various AI characters, create their own, or explore different topics and scenarios. Whether you want to practice a language, plan a trip, write a story, or simply have fun, you can find an AI that suits your needs and interests [3]. Users can create characters, craft their personalities, set specific parameters, and publish them in the community for others to chat with. Many characters are based on fictional media or celebrities, while others are completely original, with some designed for specific purposes such as assisting with creative writing or serving as text-based adventure games [4].

Mainstream media outlets report that Character.ai has made conversations with avatars feel more lifelike. Users can now even interact with their favorite synthetic personalities over the phone. According to Character.ai, there are over a million voice options available, providing a wide range of styles and accents for those creating or customizing their own Character.ai chatbots.

To perform our analysis, we gathered various questions asked in the subreddit, such as *What do you do on the platform?*, *How old are you?*, and *How did you learn about the platform?*.

We collected all the answers to these questions and conducted a demographic analysis of the subreddit community, discovering that the majority of users are young and male. Additionally, we looked

into how much time they typically spend on the platform to get a better sense of user engagement. To understand what users primarily do on the platform, we utilized topic modeling to analyze their activities. In the context of comfort-related discussions, the term *comfort* emerged as particularly significant. A PCA of the related words underscored this focus, demonstrating how comfort-related data aligns with the dataset and the overall trend of comfort. Following this, we created a semantic network to map the associations and relationships of the word *comfort* within the context of Character.ai comfort-based interactions. The network illustrates the diverse aspects of how comfort is perceived and discussed within the platform, emphasizing its significance in comfort-related interactions.

2 Literature Review

The research paper "Empathy in AI Characters: Alleviating Anxiety through Supportive Interactions" [6] explores the interactions between AI personas and users with social anxiety. This study categorizes the responses from AI characters, each designed with unique personalities, to illustrate their capacity for providing emotional support, empathy, and practical advice that meets the specific needs of users. These AI characters are programmed to adjust their behavior based on their distinct personas, offering personalized support that deeply resonates with individuals.

The study highlights the role of AI personas as beneficial tools for practicing social interactions. These characters help users incrementally build confidence and improve communication skills within a controlled environment. While they are not replacements for professional therapy, AI characters enhance mental health support systems by providing supplementary assistance and encouraging active engagement in mental health practices [6].

AI chatbots, or conversational user interfaces, have increasingly become integral to modern mental health interventions. This can be seen in early examples like ELIZA [13], which mimicked a psychotherapist's responses during therapy sessions, these tools have evolved significantly with advances in artificial intelligence and natural language processing. As reported in the comprehensive study by Sweeney et al. [9], a notable majority of mental health professionals recognize the utility of chatbots in enhancing patient management and support. These professionals cite benefits such as improved accessibility, which is particularly valuable in remote areas, and the potential to reduce the stigma associated with seeking mental health care. Furthermore, 74% of the survey participants acknowledged the significant role chatbots could play in the mental health landscape, suggesting a high level of acceptance among practitioners [9].

However, the enthusiasm for chatbots is tempered by significant ethical and practical challenges.A major issue is that chatbots currently struggle to understand or mimic human emotions, which are essential for effective therapy. This limitation raises concerns about the applicability of chatbots in scenarios requiring deep emotional support, such as crisis intervention or therapy for severe mental health conditions. Additionally, issues related to data privacy and the security of sensitive personal information continue to be major concerns [9].

The article "Therapy and E-therapy-Preparing Future Psychiatrists in the Era of Apps and Chatbots" [5] explores the integration of digital therapies, or e-therapies, into psychiatric practice and education. These e-therapies range from interactive websites providing cognitive behavioral therapy to AI-driven chatbots that assist with patient care outside traditional therapy sessions. E-therapies offer several benefits, such as increased accessibility to care, flexible, personalized treatment, which can be more engaging and less restricted by location and scheduling than traditional therapies. However, the article also addresses challenges such as the variable quality and effectiveness of digital interventions, the high dropout rates in non guided therapy apps, and concerns over the ethical use of these technologies [5].

3 Method

3.1 Data collection

Reddit (https://Reddit.com) is one of the largests social media in terms of users [2]. Reddit users can share text, links, images, or videos in various sub-communities (called subreddits and dedicated to specific topics) [8]. Users can comment and vote on posts and comments for free and anonymously. A large number of existing studies use data from the platform to examine public perceptions, attitudes, and opinions.
Reddit enables researchers to extract the subreddits, threads, comments, and associated metadata through various programming languages [14]. Therefore, data collection and analysis in this study comply with the terms and conditions of the data source. The data was collected from Reddit with the use of APIs, first we considered different questions posted by user on the platform at different times. We wanted to investigate very specific question and excluded the questions that included other questions in it. The questions we chose were:

- How old are you, what is your gender?
- How many hours (weekly) do you spend on the platform?
- What do you do on the platform?

In terms of data cleaning, emoticons, digits, punctuation, unnecessary words and stopwords were removed from the textual data. The NLTK library provides the English stopwords list with no semantic or informational value. Stopwords are typically filtered out in natural language processing to enhance the efficiency and accuracy of text analysis. Additionally, all uppercase letters have been converted to lowercase. Furthermore, considering a lot of abbreviations used in the subreddit, we managed to create a dictionary to convert those words, other unuseful terms were removed such as ('use', 'cai', 'though', 'haha','ayo','fr')

3.2 Demographic and data analysis

We extracted all the replies received on different questions regarding the age of the people on the platform. The data was from different polls, so we aggregated it into the most popular classes.

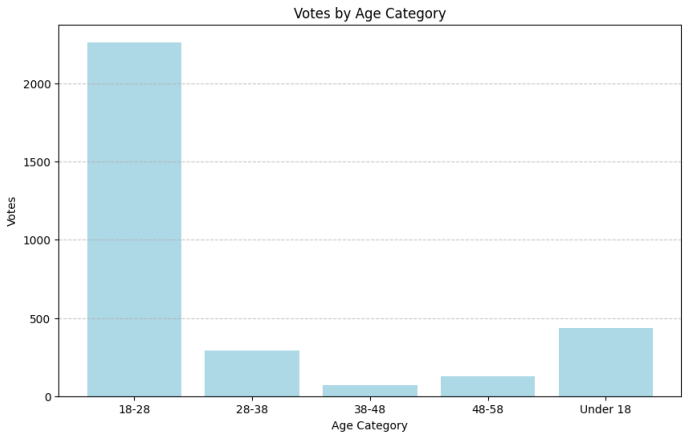


Figure 1: Barplot of user's age

The voting data suggests varying levels of engagement across different age groups, with younger adults (particularly those between 18 and 28) being the most active, followed by individuals under 18 and those between 28 and 38. Interestingly, there is a notable increase in votes from individuals aged 48 to 58. This uptick could indicate a specific interest or concern in the voting subject among this older demographic.

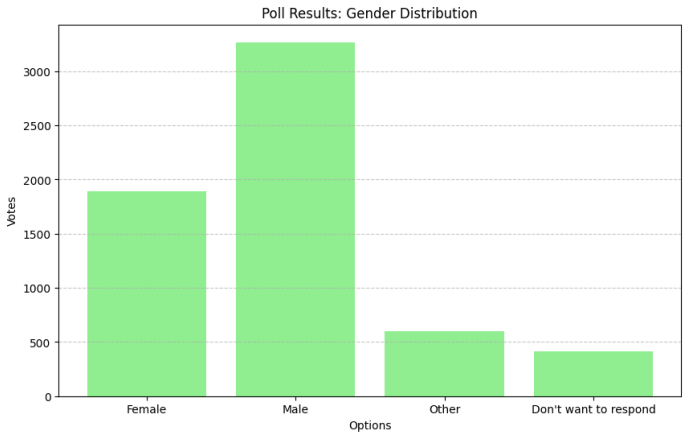


Figure 2: Barplot of user's gender

We managed to understand the prominent gender in the community, we obtained data from a single poll question posted on the subreddit. We see that the majority of people in the community are males.

In order to understand the time spent by the users, we managed to get from each screenshot of

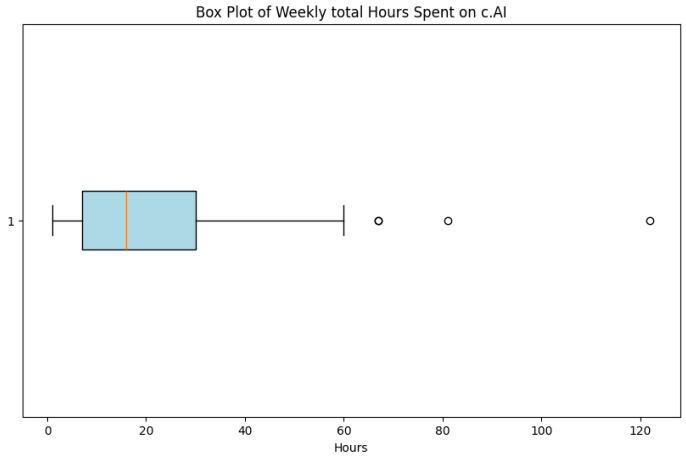


Figure 3: Boxplot of time spent in the platform (in weekly hours) by users

their time spent in the platform, the data was manually collected from a reddit post and written to an excel [7].

The box plot in figure 3 illustrates the distribution of the total weekly hours spent on Character.ai by users. From this box plot, we can infer: Most users spend between approximately 10 and 30 hours per week on Character.ai. The median weekly usage is around 20 hours. A few users spend significantly more time on the platform, with outliers reaching up to 120 hours per week.

3.3 Topic modeling

To explore the motivations behind using Character.ai, we decided to conduct a topic modeling analysis. For topic modeling, prior research suggests various techniques, with Latent Dirichlet Allocation (LDA) being particularly effective at identifying latent topics using probabilistic methods. In this study, we will utilize the LDA model for topic modeling, it is a hierarchical Bayesian probability model comprising three levels: documents, topics, and words. The core concept of the method is that each document can be represented as a mixture of latent topics, and each topic is a probability distribution over all words in the vocabulary [1]. We chose LDA method because it excels in dealing with varied and extensive text collections, making it highly effective for identifying the main ideas across a wide range of documents without direct input or manual categorization, although one of the key challenges with LDA is deciding the number of topics to use in the model, also it assumes that topics are independent of each other, which might not always hold true in some cases.

Considering that this subreddit is one of the most popular at the moment, we chose not to extract all posts but to focus on relevant and specific questions. This approach ensures a more targeted analysis and avoids the noise that a complete extraction might introduce.

For topic modeling we extracted all the replies received to various questions about users' activities on the platform. Examples of these questions include, *What do you use cai for?* and *I use the app for silly jokes; what do you use it for?*. [11][10][12]. By analyzing these responses, we aimed to gain insights into the different ways users engage with Character.ai.

The LDA model identified five distinct topics based on the usage and context of words. Each list of words represents the most frequently occurring and significant terms associated with a particular topic.

- Topic 0: This topic seems focused on the use of AI for roleplaying and creating fictional interactions. The words visible in Figure 4a suggest that users engage with AI chatbots to partake in narrative or character-driven scenarios, perhaps as a method of escapism or creativity. The presence of *coping* implies that these interactions may also serve therapeutic purposes, helping users manage real-life issues through the comfort of fictional contexts. *Friend* and *talking* highlight the social and conversational nature of these interactions.

- Topic 1: This topic in Figure 4b likely covers the emotional and creative aspects of interactions with AI. *Feeling*, *real*, and *love* suggest deep emotional engagement and personal connections with the characters or scenarios created alongside AI. *Ocs* (original characters) indicates that users are inventing their own characters, likely engaging deeply in personalized content creation.

- Topic 2: This topic visible in Figure 4c appears

to blend elements of entertainment and emotional support. The words indicate that the interactions could be light-hearted or humorous. *Comfort*, *real*, and *life* reflect the supportive and perhaps comforting role of AI chatbots, potentially offering users a space to find solace and companionship.

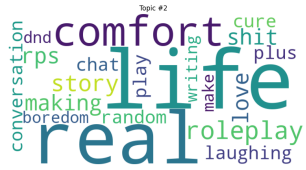
- Topic 3: This topic in Figure 4d is likely centered around functional or support-based interactions. *Help*, *question*, and *stuck* suggest users seeking assistance or information.

- Topic 4: This topic in Figure 4e seems to have a mixed focus, potentially blending game interactions with emotional and personal discussions.

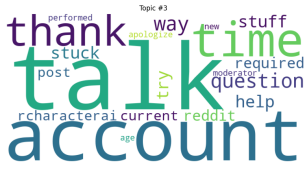
Figure 5 provides a detailed breakdown of the significance of each word within each topic. In general we don't see discrepancy with the previous analysis.

(a) Topic 0

(b) Topic 1



(c) Topic 2



(d) Topic 3

(e) Topic 4

Figure 4: Importance of each word in individual topics

3.4 Semantic analysis

As observed in the previous analysis, three topics highlight the comfort provided by the platform. To

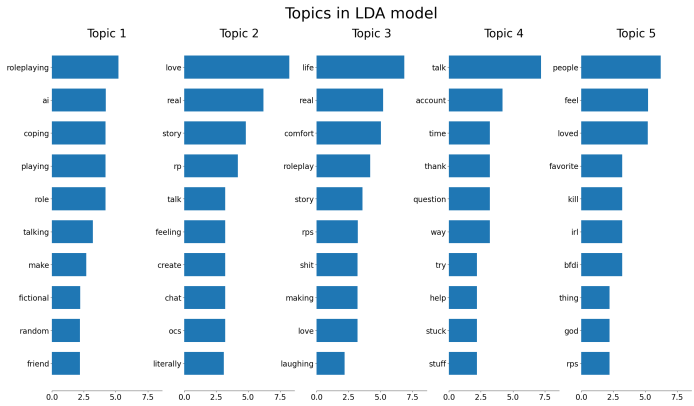


Figure 5: Importance of words for each of the topics

gain a deeper understanding of how comfort is perceived within the community, we conducted a more detailed analysis of the word *comfort* as it offers a promising compromise for understanding the relevance of comfort-based interactions in this community and based on the topic modeling, this term is highly relevant.

First, we use TF-IDF (Term Frequency-Inverse Document Frequency), a technique used to measure the importance of a word in a document relative to a collection of documents. We apply this method to the cleaned text of all the comments. Then, we use PCA to reduce the dimensionality of the dataset while retaining most of the variance present in the data. This technique is often favored for its simplicity, efficiency, and effectiveness in visualizing and reducing noise in data. Its components are uncorrelated, which may be beneficial for our research.

We filter the dataframe to include only rows where the text in comments containing the word *comfort* and then we apply the PCA. The resulting dataframe subset focuses on discussions mentioning the word. We calculate the average coordinates (centroid) of the word related discussions in the transformed space. The mean values represent the central tendency of these discussions in the semantic space.

In the figure 6 we see that *comfort* data points (red) are concentrated near the origin of the PCA plot, suggesting that discussions about that word are central and not particularly extreme in terms of the principal components. This central location

indicates that *comfort* is a common and relevant topic but does not stand out as having extreme values in the PCA space. The mean direction vector (blue arrow) got a negative direction along both PC1 and PC2 axes indicates that, on average, discussions about *comfort* tend to have slightly lower values on these components. This could mean that the semantic space related to comfort-based conversations is somewhat distinct but not vastly different from the general discussions.

We can see that *Comfort* is a central topic in discussions, indicating its importance and relevance to users and discussions.

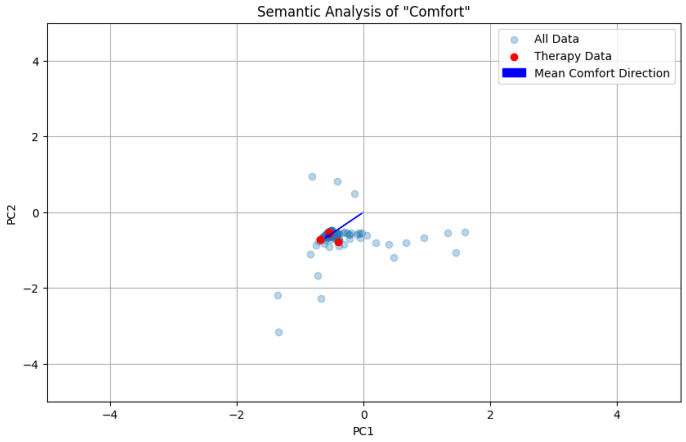


Figure 6: PCA of the word *comfort*

For that reason, we create a semantic network that maps out the associations and relationships of that word within the context of Character.ai interactions. *Comfort* is the central node, indicating its primary role in the network. Various terms are connected to the word showcasing their semantic relationship. These terms include:

- Emotional and psychological aspects: Words like *coping*, *loneliness*, *reality*, *addicted* and *delusion* suggest a focus on the emotional and psychological dimensions related to the word.
- Behavioural aspects: Terms such as *roleplay*, *create*, *stop*, *indulging* and *avoid* highlight the behavioral interactions related to comfort.
- Descriptive terms: Words like *special* and *personality* provide context and descriptions of the related interactions.

This network provides an overview of how comfort is semantically linked to various aspects of user interactions within the comfort-therapy context, we see the different dimensions of comfort in Character.ai.

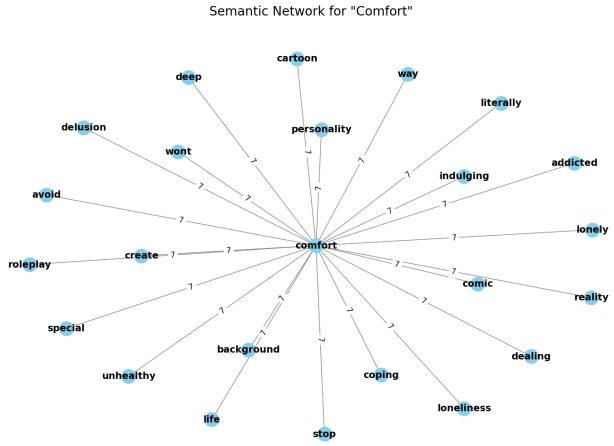


Figure 7: Semantic Network

4 Conclusion and final considerations

This study highlights the effectiveness of employing text analytic techniques to gain insights into user interactions with the Character.ai platform. While these methods are immensely useful for discerning patterns and preferences within large datasets, certain limitations were encountered.

The plan to run a survey for direct user feedback was blocked by rules that didn't allow this approach. There's also the risk of self-bias, respondents might answer questions in a way that reflects what they think is expected or aligns with their personal views, rather than providing honest, objective feedback. This can lead to skewed results, making it harder to get an accurate picture of what the entire group truly thinks or feels. Another challenge is identifying when answers are given with irony. Additionally, when gathering all the data, we might find patterns in the poll responses due to correlations between answers across different polls, which could affect the accuracy of the demographic results.

Despite this, the methodologies applied have sig-

nificantly contributed to understanding the underlying theme of comfort and therapy based interactions within the community. It seems that a platform created without this particular scope is now being used for this purpose.

References

[1] S. Boon-Itt and Y. Skunkan. "Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study". In: *JMIR Public Health and Surveillance* 6 (2020), e21978.

[2] K. Brigham. "Reddit throughout the years: Its rise to prominence, recent revolts and IPO plans". In: *CNBC* (2023). [cited 2023 December 2]. URL: https://www.cnbc.com/2023/07/30/reddits-rise-to-prominence-recent-revolts-and-future-prospects.html.

[3] Character AI. *Character AI*. Accessed: 2024-07-22. 2024. URL: https://character.ai/.

[4] Wikipedia contributors. *Character.ai — Wikipedia, The Free Encyclopedia*. Accessed: 2024-07-22. 2024. URL: https://en.wikipedia.org/wiki/Character.ai.

[5] David Gratzer and David Goldbloom. "Therapy and e-therapy—preparing future psychiatrists in the era of apps and chatbots". In: *Academic Psychiatry* 44 (2020), pp. 231–234.

[6] Fauziah Maharani, Wily Mohammad, and Hana Melati Ameira. "Empathy in AI Characters: Alleviating Anxiety through Supportive Interactions". In: *Arika* (2023).

[7] Reddit user. *How many hours do you spend weekly on CAI?* https://www.reddit.com/r/CharacterAI/comments/18kjf89/how_many_hours_do_you_spend_weekly_on_cai/. Accessed: 2024-08-07. 2023.

[8] J. Sawicki et al. "Exploring Usability of Reddit in Data Science and Knowledge Processing". In: *Scalable Computing: Practice and Experience* 23 (2022), pp. 9–22.

[9] Colm Sweeney et al. "Can chatbots help support a person's mental health? Perceptions and views from mental healthcare professionals and experts". In: *ACM Transactions on Computing for Healthcare* 2.3 (2021), pp. 1–15.

[10] Reddit User. *I use the app for silly jokes, what do you use it for?* https://www.reddit.com/r/CharacterAI/comments/17ybihi/i_use_the_app_for_silly_jokes_what_do_you_use_it/. Accessed: 2024-09-07. 2023.

[11] Reddit User. *Well, okay, what were you guys doing before the...?* https://reddit.com/r/CharacterAI/comments/154oesk/well_okay_what_were_you_guys_doing_before_the/. Accessed: 2024-09-07. 2023.

[12] Reddit User. *What do you use C.ai for?* https://www.reddit.com/r/CharacterAI/comments/1cscxvh/what_do_you_use_c_ai_for/. Accessed: 2024-09-07. 2023.

[13] Joseph Weizenbaum. "Eliza—A computer program for the study of natural language communication between man and machine". In: *Communications of the ACM* 9.36 (1965), p. 45.

[14] J. Yadav, M. Misra, and K. Singh. "Sensitizing Netizen's behavior through influencer intervention enabled by crowdsourcing–a case of reddit". In: *Behaviour & Information Technology* 41 (2021), pp. 1286–1297.