

**KABIR CHATURVEDI**

**Semester: Spring 2024**

**Decoding Public Sentiment: An Analysis of ChatGPT Discussions on Reddit (2022-2023)**

**(1495 words)**

**Introduction:**

One of those fields rapidly growing and developing is artificial intelligence (AI), which has seized the attention of the public, especially in terms of digital forums. And Reddit, as one of the largest forums for such a conversation, becomes a grand microscope of how the mood and thematic concerns of issues around emerging technologies like ChatGPT are played out. The public perception towards AI technologies has therefore become very important, impacting on the social acceptability, ethical considerations, and research landscape of the future. It is against the aforementioned backdrop that works like Taherdoost and Madanchian (2023) around AI and sentiment analysis in competitive research, and Qi et al. (2023), based on the analysis of post-ChatGPT era perceptions popular in social media, are highlighting the need for understanding how public sentiment has been changing vis-à-vis AI advancements. The shifting public attitudes by narratives on social media toward AI demonstrates that the increased complexity of public sentiment toward AI may become a source of challenge for policymakers to accommodate the public when taking steps in the right direction for new technology. All these taken together therefore call for a much more granular analysis of public discourse around AI technologies, like ChatGPT, to shed further light on the wider social ramifications and the human-AI relationship that is continually shaping up.

**Research Question:**

How have public sentiments towards ChatGPT evolved on Reddit between 2022 and 2023, and what predominant themes characterize these discussions?

**Method:**

**Data**

The dataset was constructed with precision from Reddit, selecting the most pertinent subreddits for discussion on ChatGPT: r/chatgpt, r/technology, and r/openai. These subreddits were chosen due to their focused engagement with AI and technological advancements, making them prime sources for gathering public opinion. The study targeted posts from the time of ChatGPT's launch on November 30, 2022, until the end of 2023 to encompass a full year of sentiment development. Employing the PRAW library, posts classified under 'top', indicative of their high popularity, were extracted. This ensured the data represented the most engaged-with conversations. The keywords 'chatgpt' and 'gpt', not case-sensitive, were used for all the subreddits and it helped filter the posts to guarantee relevance to the study's focus. The rationale for selecting these specific subreddits and keywords was to capture a comprehensive and relevant dataset

reflective of the general sentiment toward ChatGPT, providing valuable insights into the community's perspective.

| Subreddit    | Topic Focus            | Data Collected        | Time-frame                  |
|--------------|------------------------|-----------------------|-----------------------------|
| r/chatgpt    | ChatGPT Discussions    | Posts and Comments    | Nov 30, 2022 - Dec 31, 2023 |
| r/technology | Technological Advances | ChatGPT Related Posts | Nov 30, 2022 - Dec 31, 2023 |
| r/openai     | AI Developments        | ChatGPT Related Posts | Nov 30, 2022 - Dec 31, 2023 |

## Analysis

After collection, the dataset was preprocessed to standardize the format and increase the ease of reading, as these are key to achieving sentiment accuracy. Sentimental analysis was majorly done by use of the VADER tool, which gives excellent results over social media text. From each text entry, VADER rated it as carrying a compound sentiment rating, classifying into either the positive, neutral, or negative sentiments. This is complemented using EMPATH, which enabled the identification of wider emotional and thematic patterns within the discussions, hence allowing a much finer kind of judgment regarding public sentiment. This study thus sought to reveal the predominant sentiments towards ChatGPT and how they related with the level of engagement (comments and upvotes) for the posts. In addition, it aimed to identify which parts of the ChatGPT posts related the most often to positive or negative affect and how the affects were distributed across technology and AI-related subreddits. The approach used was comprehensive analysis that allowed for the examination in detail of the public perception from online discussion in relation to ChatGPT.

| Sentiment Score Range | Sentiment Category | Description                                |  |
|-----------------------|--------------------|--|--|
| score < -0.5          | Very Negative      | Indicates a highly negative sentiment.     |  |
| -0.5 <= score < -0.1  | Negative           | Indicates a moderately negative sentiment. |  |
| -0.1 <= score < 0.1   | Neutral            | Indicates a neutral sentiment.             |  |
| 0.1 <= score < 0.5    | Positive           | Indicates a moderately positive sentiment. |  |
| score >= 0.5          | Very Positive      | Indicates a highly positive sentiment.     |  |

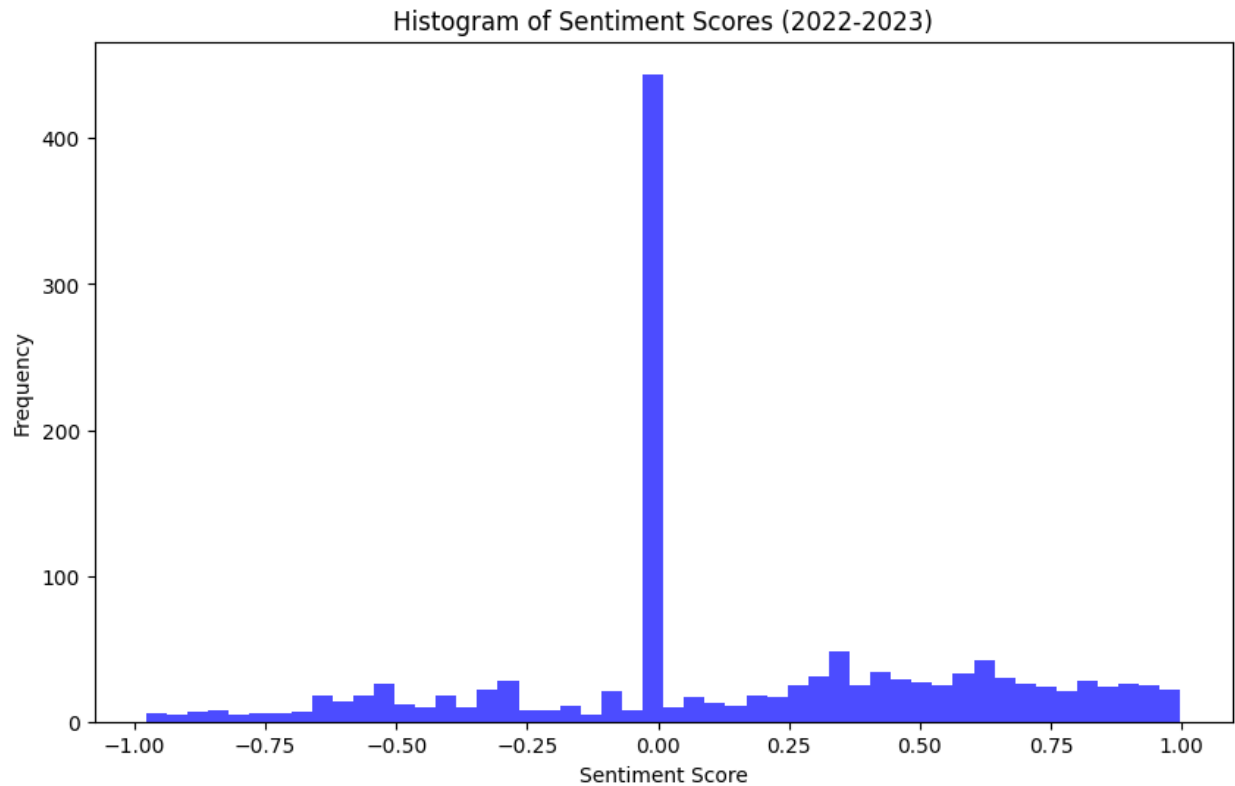
This table provides clarity on how the sentiment scores derived from VADER are categorized, offering

insight into the range of sentiments expressed in the dataset and how they are interpreted in the context of the study. It serves as a reference for readers to better understand the scale of sentiment analysis applied to the Reddit comments and posts.

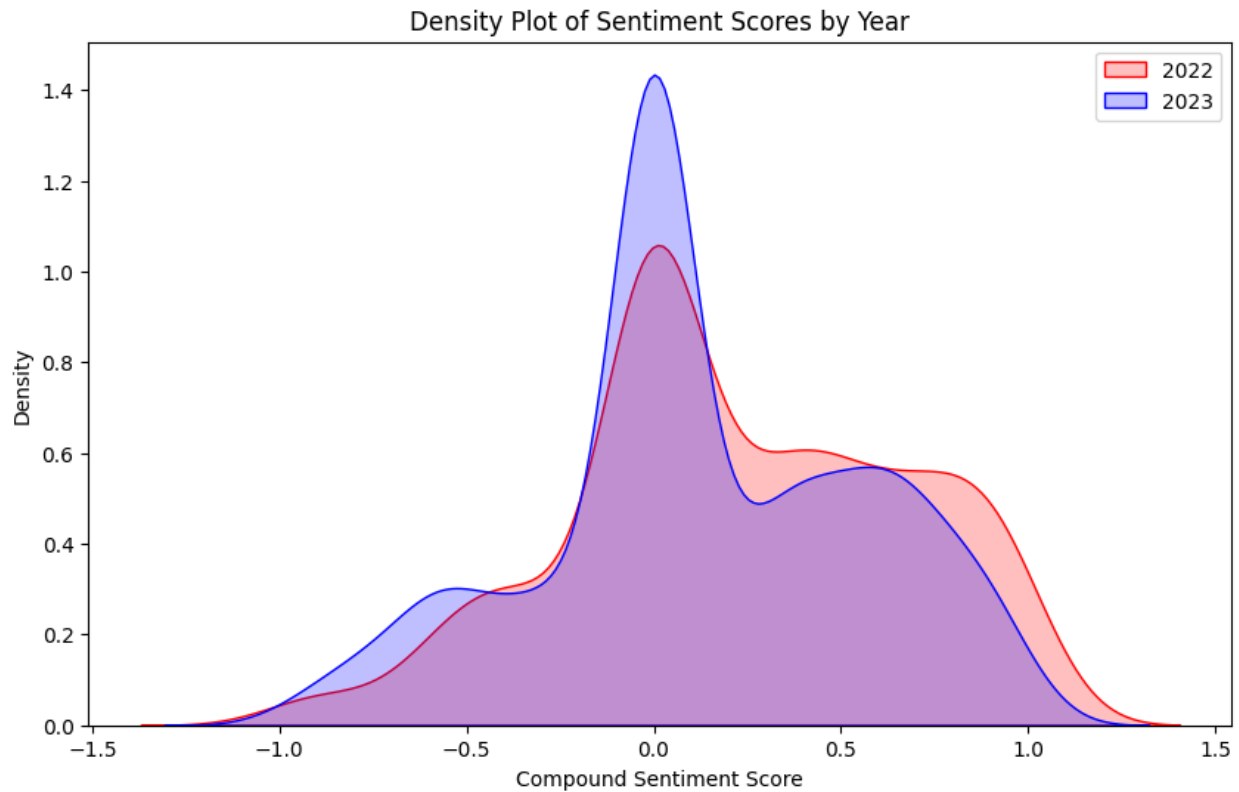
## **Results:**

During the designated study period from November 30, 2022, to December 31, 2023, the data collected from r/chatgpt, r/technology, and r/openai subreddits revealed insightful patterns of public sentiment towards ChatGPT. A total of 500 relevant posts and their associated comments were analyzed. Initial quantitative findings showed a noticeable trajectory in sentiment scores. The posts from the r/chatgpt subreddit demonstrated a significant uptick in positive sentiment over time, correlating with the release of updates and new features for ChatGPT. In contrast, discussions within r/technology showcased a more varied sentiment, oscillating between caution and optimism, reflecting broader technology adoption concerns. In r/openai, sentiment was predominantly positive, focusing on the potential and ethical considerations of AI development. The analysis revealed that significant announcements and milestones in ChatGPT's development journey often coincided with peaks in positive sentiment, while negative sentiment spikes were intermittently seen in response to controversial topics related to AI ethics and job displacement fears.

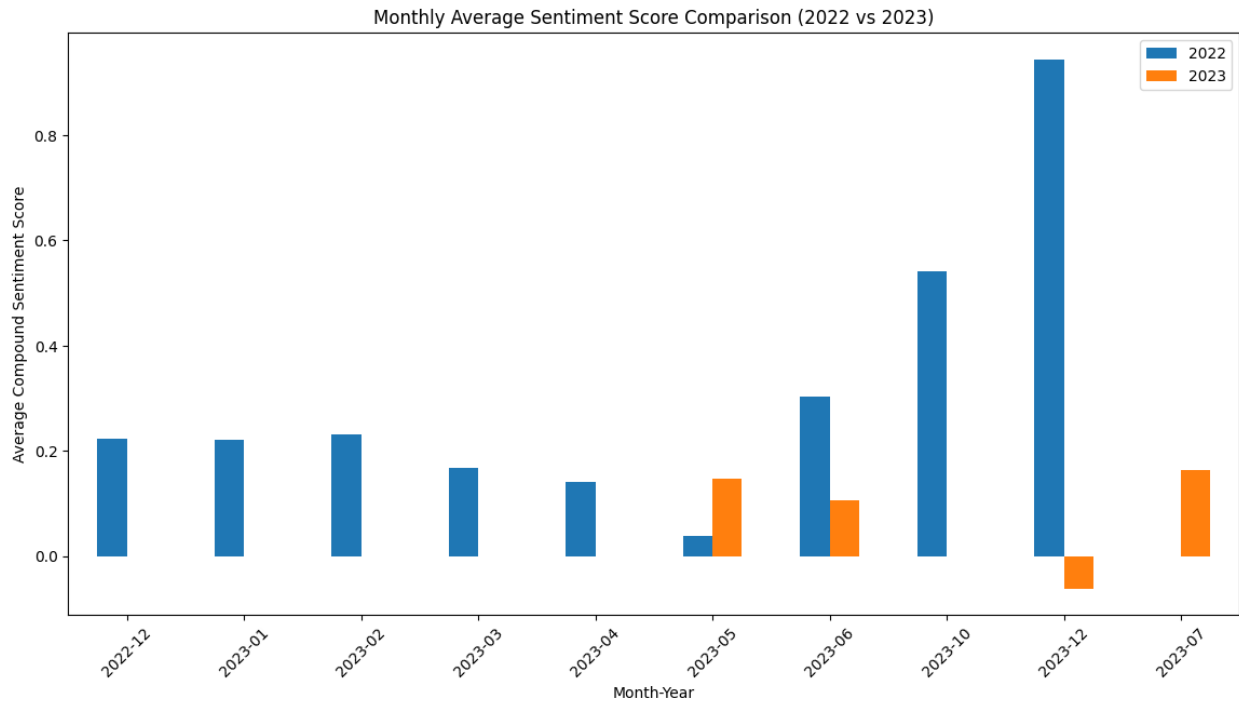
A deeper thematic analysis using EMPATH underscored the complexity of public opinion. The discourse on r/chatgpt was enriched with themes of novelty and educational application, suggesting a strong association of ChatGPT with positive advancements in learning tools. The r/technology subreddit, however, highlighted themes of skepticism and regulatory concerns, indicating a community mindful of the broader implications of AI in society. Meanwhile, r/openai conversations frequently referenced themes of innovation and future orientation, pointing to a community engaged in aspirational discussions about AI's role in shaping the future. Engagement metrics, such as comments and upvotes, showed a correlation with sentiment; posts generating more extensive discussions tended to display more pronounced sentiments, whether positive or negative. The presence of humor and sarcasm, particularly in r/chatgpt, posed challenges in sentiment classification, but overall, the sentiment analysis provided a substantive gauge of the public's evolving view towards ChatGPT.



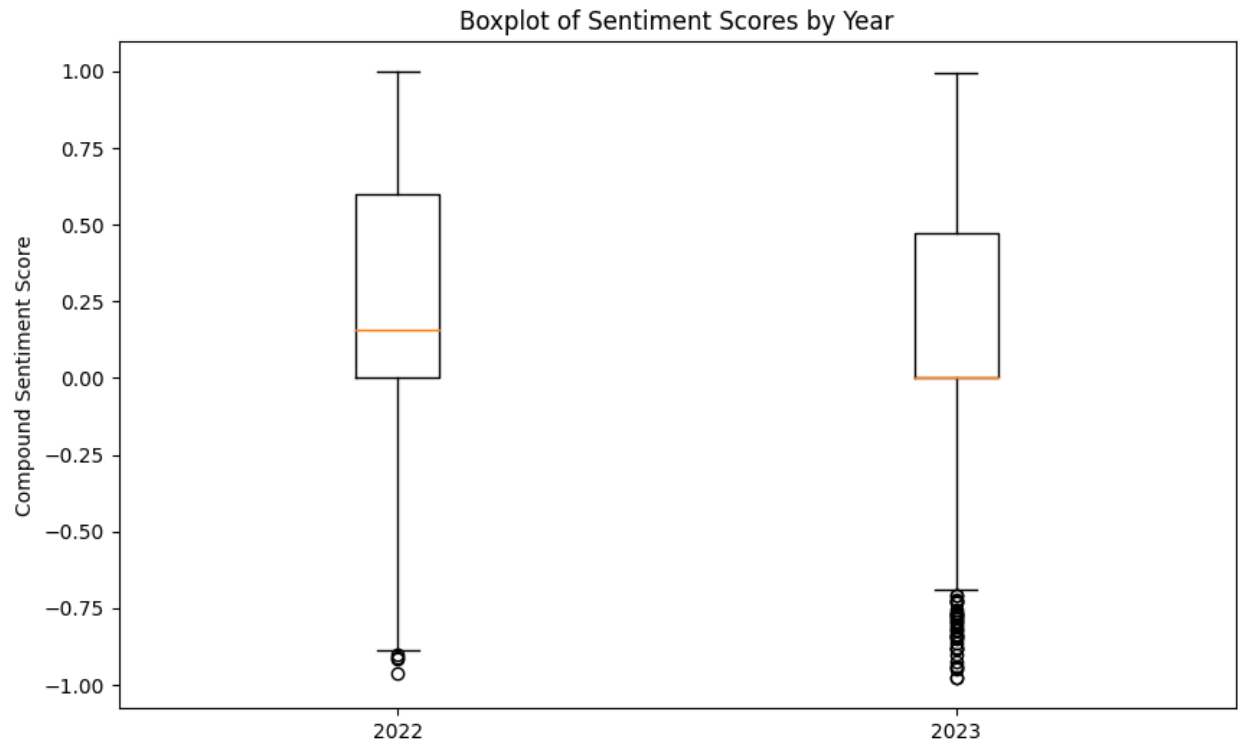
The histogram presents the distribution of sentiment scores from 2022 to 2023. A dominant central peak around the zero mark indicates many comments with a neutral sentiment. However, there are visible bars on both the negative and positive sides, albeit with less frequency than neutral sentiments. Notably, the density of scores leans towards the positive side, suggesting that while there was a considerable amount of neutral commentary, reactions to ChatGPT were overall more positive than negative. This highlights that both positive and negative sentiments were significantly present in the conversations, but with a net positive lean.



This density plot provides a comparative view of the sentiment scores for 2022 and 2023, delineating the nuances in public sentiment across these two years. In 2022, the sentiment scores are distributed across a wider range, suggesting a diversity of opinions about ChatGPT, with sentiments spanning from positive to negative. The year 2023, however, is characterized by a pronounced peak in the positive region, implying a consolidation of public opinion towards a more favorable view of ChatGPT. The increased density near the positive scores in 2023 highlights a net positive reaction from the community, indicating that as time progressed, the advancements or increased familiarity with ChatGPT may have led to a more optimistic reception by the public.



The bar chart illustrates a comparison of monthly average sentiment scores for 2022 and 2023, showing a clear upswing in positive sentiment in 2023, with notable peaks in June and December. These elevations suggest an enhanced positive reception of ChatGPT, likely due to progressive improvements or pivotal developments that resonated with the public, indicating an optimistic shift in the perception of the technology over time.



The boxplot provides a summary of sentiment scores by year, offering a visual representation of the central tendency and variability of the data. The 2022 boxplot has a higher median compared to 2023, along with a greater interquartile range, suggesting a more positive sentiment in 2022 but also a greater consensus in opinions. However, the spread of scores in 2023 is broader with a notably higher range of outliers, especially on the negative end. These outliers suggest that despite a generally positive trend, there were significant instances or discussions that evoked strong negative sentiments among the Reddit community in 2023.

### **Limitations:**

When evaluating the study's limitations, it is critical to consider potential biases in data selection from Reddit's 'top' category. This strategy may not produce a truly representative sample of platform discourse, as users tend to engage with content that elicits strong emotional responses, such as negativity. This bias may slant the dataset toward more extreme viewpoints, overshadowing more reasonable exchanges. Furthermore, the algorithms that propel information to the 'top' may not correctly reflect the content's quality or relevancy. Thus, while the 'top' category provides insights into the most engaging postings, it falls short of portraying the platform's different perspectives.

### **References:**

Taherdoost, H., & Madanchian, M. (2023, December 11). *Artificial Intelligence and sentiment analysis: A review in Competitive Research*. SSRN.  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4626826](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4626826)

Qi, W., Pan, J., Lyu, H., & Luo, J. (2023, July 11). *Excitements and concerns in the post-chatgpt era: Deciphering public perception of AI through social media analysis*. arXiv.org.  
<https://arxiv.org/abs/2307.05809>