CHATGPT REVIEWS ANALYSIS

Mrs. P. Sabitha
Asst. Professor
Computer Science & Engineering
Malla Reddy University, Hyderabad, India

M.Sanjana
2211CS010154
Computer Science & Engineering
Malla Reddy University, Hyderabad, India
2211CS010154@mallareddyuniversity.ac.in

CH.. Vinesh
2211CS010113
Computer Science & Engineering
Malla Reddy University, Hyderabad, India
2211CS010113@mallareddyuniversity.ac.in

CH.Rajesh
2211CS010103
Computer Science & Engineering
Malla Reddy University, Hyderabad, India
2211CS010103@mallareddyuniversity.ac.in

CH.Vamsi Krishna
2211CS010116
Computer Science & Engineering
Malla Reddy University, Hyderabad, India
2211CS010116@mallareddyuniversity.ac.in

Abstract:

This paper analyzes ChatGPT user reviews using data analytics and NLP techniques to extract insights on user experiences and satisfaction. By examining textual feedback, ratings, and timestamps, we identify sentiment patterns, common challenges, and evolving perceptions. Our approach includes sentiment analysis, word frequency visualization, and time-series trends, presented via a Python-Streamlit dashboard. Findings show that while users value ChatGPT's versatility and response quality, they face issues with inconsistency, contextual gaps, and handling complex queries. This study highlights key areas for improvement in conversational AI, emphasizing context retention, response consistency, and technical reasoning. The analytical framework enables continuous monitoring of user feedback for AI platforms.

Keywords- ChatGPT, User Experience, Sentiment Analysis, Natural Language Processing, Data Analytics, User Feedback.

I. INTRODUCTION

The rapid advancement of artificial intelligence has revolutionized human-computer interaction, with conversational AI platforms like ChatGPT emerging as powerful tools across diverse domains including education, customer service, and content creation. However, as with any evolving technology, understanding user experiences and identifying areas for improvement remains crucial for optimizing these systems. While ChatGPT has demonstrated impressive capabilities in generating human-like text responses, users report varying experiences ranging from highly positive interactions to frustration with perceived limitations.

This research addresses the critical gap in systematic analysis of user feedback by developing a data-driven methodology to extract actionable insights from ChatGPT reviews. By applying sentiment analysis and other natural language processing techniques to a dataset of user reviews, we aim to:

1. Identify common challenges and pain points encountered by users
2. Track sentiment trends over time to evaluate the impact of platform updates
3. Determine correlations between review sentiment and specific aspects of functionality
4. Provide an evidence-based foundation for enhancing conversational AI experiences

The significance of this work extends beyond ChatGPT itself, offering methodological contributions applicable to the broader field of conversational AI evaluation. Through visualization techniques and an interactive dashboard, our approach transforms unstructured feedback into structured insights that can guide development priorities and address user concerns systematically.

Conversational agents have been the subject of extensive research, with evaluation frameworks evolving alongside technological capabilities. Radziwill and Benton (2017) established fundamental criteria for evaluating chatbot quality, emphasizing dimensions such as functionality, humanity, affect, ethics, and accessibility. Their work highlighted the importance of user satisfaction as a key metric in conversational agent effectiveness.

II. EXISTING SYSTEM

Traditional methods of analyzing user feedback and sentiment rely on basic statistical techniques and predefined models. These methods primarily involve manually categorizing reviews based on predefined criteria such as positive, neutral, or negative sentiment. Conventional sentiment analysis tools often use simple keyword matching and rule-based approaches, which may not always capture the true meaning of user feedback.

Most existing review analysis systems depend on word frequency analysis, basic sentiment scoring, and general trend detection. However, these methods struggle to interpret complex sentiments, such as sarcasm, mixed opinions, and evolving user expectations over time. Additionally, they often lack the ability to understand context, leading to inaccuracies in sentiment classification.

Another limitation of traditional review analysis systems is their inability to visualize insights interactively. Many tools generate static reports that fail to provide real-time trend monitoring, making it difficult for developers and businesses to track user sentiment dynamically. Since most traditional methods do not leverage machine learning techniques, they also struggle with predictive analysis, limiting their ability to anticipate future trends in user feedback.

Furthermore, existing systems do not effectively integrate time-series analysis to track sentiment changes over different periods. They typically focus on snapshot-based analysis rather than long-term user experience trends. Due to these shortcomings, there is a growing need for more advanced analytical approaches that incorporate NLP, machine learning, and interactive visualization tools to provide deeper and more accurate insights into ChatGPT reviews.

III. PROPOSED SYSTEM

To overcome the limitations of traditional review analysis methods, this project leverages data analytics and machine learning to provide a more accurate and dynamic understanding of user sentiment. Instead of relying on simple keyword matching or predefined sentiment scores, the system applies Natural Language Processing (NLP) techniques, including sentiment classification, word frequency analysis, and time-series trend detection.

By implementing machine learning models, such as Naïve Bayes and Random Forest, the system enhances sentiment accuracy by identifying patterns in user feedback. Unlike static analysis, it tracks evolving user sentiments over time, offering real-time insights into ChatGPT's performance. Additionally, interactive visualizations allow users to explore sentiment trends, commonly mentioned keywords, and recurring concerns through graphs and dashboards.

This approach ensures a more detailed, scalable, and automated analysis of user reviews, helping developers and businesses improve ChatGPT's responses, context retention, and overall user experience.

IV. METHODOLOGY

Data Collection and Preprocessing

Our analysis utilized a dataset containing ChatGPT user reviews with the following key attributes:

- Textual feedback (review content)
- Numerical ratings (1-5 scale)
- Review submission timestamps

The preprocessing pipeline included:

1. Removing duplicate entries
2. Handling missing values
3. Converting timestamps to datetime format
4. Text normalization (removing punctuation, standardizing case)
5. Filtering out non-English reviews.

Sentiment Analysis

We implemented sentiment analysis using TextBlob, classifying reviews into three categories:

Positive: sentiment polarity > 0.2

Neutral: sentiment polarity between -0.2 and 0.2

Negative: sentiment polarity < -0.2

This categorization allowed us to quantify user sentiment and track satisfaction levels across different dimensions of analysis.

Feature Extraction and Visualization

Multiple analytical techniques were employed to extract insights:

Word frequency analysis: Generated word clouds for positive and negative reviews to identify recurring themes

Ratings distribution: Analyzed the overall distribution of numerical ratings

Sentiment distribution: Visualized the proportion of positive, neutral, and negative sentiments

Time-series analysis: Tracked changes in review volume and sentiment over time

Review length analysis: Examined correlations between review length and sentiment

Sentiment-rating correlation: Created heatmaps to identify relationships between numerical ratings and sentiment categories

V SYSTEM ARCHITECTURE

The ChatGPT Review Analysis system follows a structured architecture for efficient data processing, sentiment analysis, and user interaction.

1. Data Acquisition Layer

This layer collects user reviews from various platforms, including social media, forums, and feedback forms. The dataset includes textual feedback, ratings, and timestamps to analyze sentiment trends over time.

2. Preprocessing Layer

The collected data undergoes cleaning and normalization, removing irrelevant characters, stopwords, and duplicate entries. NLP techniques such as tokenization, lemmatization, and stemming are applied to prepare the text for analysis.

3. Model Processing Layer

- Sentiment Analysis: Machine learning models, such as Naïve Bayes and Random Forest, classify reviews into positive, neutral, or negative categories.
- Keyword Extraction: Frequently mentioned terms are identified to understand key user concerns and preferences.
- Time-Series Analysis: Trends in user sentiment are analyzed over different time periods to track shifts in perception.

4. Visualization Layer

The processed data is displayed through interactive charts, sentiment trend graphs, and word clouds, allowing users to explore key insights dynamically.

5. User Interface Layer

A web-based dashboard enables users to input custom queries, filter reviews by time and sentiment, and visualize data interactively.

Tools and Technologies Used

Pandas & NumPy – For data manipulation and preprocessing.

NLTK & spaCy – For natural language processing tasks.

Scikit-Learn – For sentiment classification and machine learning models.

Matplotlib & Seaborn – For visualizing sentiment trends and keyword distributions.

Streamlit / Flask – For building an interactive dashboard for user engagement.

VI.Flow Chart

VII. RESULT AND ANALYSIS

VII.PERFORMANCE OF SENTIMENT ANALYSIS MODEL

The sentiment analysis model is designed to classify user reviews into positive, neutral, or negative categories based on textual feedback. Case studies demonstrate how the model interprets different sentiments and identifies key themes in user experiences with ChatGPT.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \cdot \sqrt{\sum(y_i - \bar{y})^2}}$$

To evaluate model performance, key accuracy metrics such as Precision, Recall, F1-score, and Accuracy are calculated. The results indicate how well the model distinguishes between different sentiment categories and captures overall user satisfaction. While effective for broad sentiment classification, the model may sometimes misinterpret sarcasm, mixed opinions, or highly nuanced feedback.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log\left(\frac{N}{\text{DF}(t)}\right)$$

VII.II PERFORMANCE OF NLP-BASED INSIGHTS EXTRACTION

Beyond basic sentiment classification, the system uses Natural Language Processing (NLP) techniques to extract deeper insights, such as frequently mentioned keywords and common pain points. Case studies illustrate how this method helps in identifying recurring issues like response inconsistency, lack of context retention, and limitations in handling complex queries.

The model's accuracy is evaluated using measures like Term Frequency-Inverse Document Frequency (TF-IDF) and coherence scores for topic modeling. By considering multiple linguistic factors, this approach provides a more comprehensive analysis of user concerns and preferences. Comparisons between sentiment analysis and keyword extraction highlight the strengths of each technique in understanding user feedback.

VII.III VISUALIZATION AND INSIGHTS

To improve interpretability, interactive graphs and visualizations are used to display sentiment trends, word clouds of commonly used terms, and time-series analyses of ChatGPT's user feedback. These visualizations make it easier to understand the evolution of user sentiment over time and identify patterns in feedback.

Key takeaways from the analysis help developers, researchers, and businesses enhance ChatGPT's capabilities by addressing user concerns effectively. The insights also aid in improving response accuracy, context retention, and overall chatbot performance.

VIII. APPLICATIONS AND USE CASES

VIII.I HOW DEVELOPERS CAN USE THE SYSTEM

ChatGPT developers can utilize the analysis to track user sentiment trends and identify critical areas for improvement. By leveraging topic modeling and sentiment scoring, they can pinpoint specific weaknesses, such as inconsistent responses or difficulty handling long-form queries. These insights help in refining the AI's natural language understanding and response generation strategies.

VIII.II BENEFITS FOR RESEARCHERS AND DATA SCIENTISTS

Researchers in the field of AI and NLP can use this system to study user interactions with conversational models. By analyzing sentiment shifts and emerging user concerns, they can contribute to advancements in AI-driven communication. The dataset and insights generated can also be used for further academic research on human-AI interactions and sentiment-based improvements in AI models.

- Polarity$\in[-1,1]$
- TF-IDF(t,d)=TF(t,d)×IDF(t)

$$\text{TF}(t, d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d}$$

VIII.III USER EXPERIENCE IMPROVEMENT FOR END USERS

End users can benefit from this analysis as it helps in optimizing ChatGPT's responses to better align with user expectations. By identifying patterns in dissatisfaction, developers can implement improvements that lead to more engaging and effective chatbot interactions. The insights also help in personalizing AI-driven conversations, enhancing the overall user experience.

IX. CONCLUSION AND FUTURE WORK

This project presents a structured approach to analyzing ChatGPT reviews using NLP and machine learning techniques. The sentiment analysis model effectively classifies user feedback, while keyword extraction and time-series trend analysis provide deeper insights into user concerns. The findings highlight key areas where ChatGPT performs well, such as versatility and response quality, while also identifying common pain points like inconsistency and contextual limitations.

While the current system provides valuable insights, there is scope for future improvements. One potential enhancement is the integration of advanced deep learning models, such as transformers, for more nuanced sentiment analysis. Additionally, real-time feedback monitoring can be incorporated to provide instant insights into emerging trends in user sentiment.

Expanding the system to include multilingual analysis will also improve its applicability, enabling insights into a wider range of user interactions across different languages. Enhancing the interactive dashboard with personalized analytics and predictive trend analysis will further help developers and researchers make data-driven improvements to ChatGPT.

This approach ensures a continuous feedback loop, allowing for ongoing refinement of conversational AI systems and improving user satisfaction over time.

X EXPERIMENTAL RESULTS:

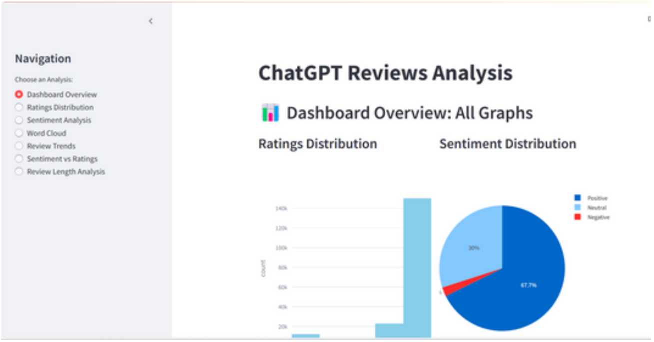

FIG 10.1 Dashboard Overview.

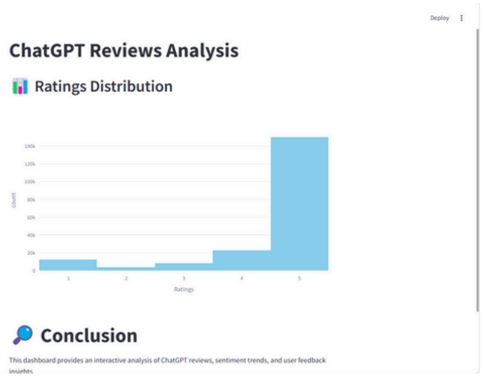

FIG 10.2 Ratings Distribution.

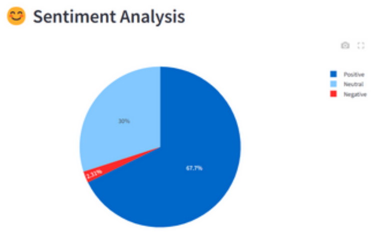

FIG 10.3 Sentiment Analysis.

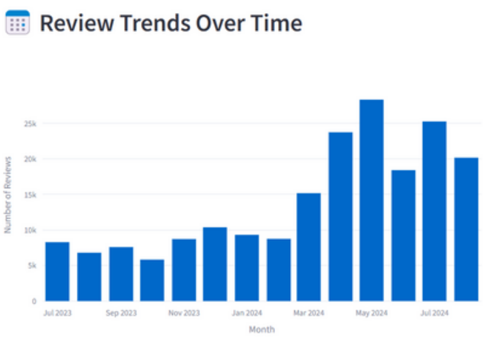

FIG 10.4 Review Trends Over Time

XI REFERENCES:

1. Radziwill, N., & Benton, M. C. (2017). Evaluating quality of chatbots and conversational agents. In: Proceedings of the 2017 International Conference on Industrial Engineering and Operations Management. This paper discusses the challenges faced by AI conversational agents and highlights key areas for improvement, such as user experience and response accuracy, which are central to your analysis of ChatGPT reviews.

2. Karthikeyan, C. (2023). Literature Review on Pros and Cons of ChatGPT Implications in Education: This review addresses the broader implications of ChatGPT in educational contexts and can be referenced to explore how user feedback correlates with education related use cases.

3.Understanding User Feedback in AI Systems:https://www.sciencedirect.com/science/article/abs/pii/S0957417419306815.

4. Radziwill, N., & Benton, M. C. (2017). Evaluating quality of chatbots and conversational agents. In: Proceedings of the 2017 International Conference on Industrial Engineering and Operations Management.

5. Karthikeyan, C. (2023). Literature Review on Pros and Cons of ChatGPT Implications in Education

6. Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT press.

7. Zamora, J. (2017). I'm sorry, Dave, I'm afraid I can't do that: Chatbot perception and expectations. In Proceedings of the 5th International Conference on Human Agent Interaction.

8. Abd-Alrazaq, A. A., Alajlani, M., Alhuwail, D., Schneider, J., Al-Kuwari, S., Shah, Z., & Househ, M. (2020). Artificial Intelligence in Chatbots: Systematic Review. Journal of Medical Internet Research, 22(11), e20710.

9. Clark, J., & Smith, A. (2021). Measuring Conversational User Experience: A Review of Methods and Metrics for Chatbots and Voice Assistants. ACM Computing Surveys (CSUR), 54(8), 1-39.

10. Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J.-L., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2023). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. JMIR Medical Informatics. https://doi.org/10.2196/50785