Emotional Comprehension in AI: A Comparative Study of ChatGPT and Gemini Empathic Features

Sri Chandra Dronavalli , Sirichandana Bomma , Soumya Nanditha Chadalavada and Swaraj Goud Uppalwai Department of Information Science, University of North Texas 1155 Union Cir, Denton, TX 76205 E–mail: srichandradronavalli@my.unt.edu, sirichandanabomma@my.unt.edu, E–mail: soumyanandithachadalavada@my.unt.edu, swarajgouduppalwai@my.unt.edu

**Abstract**

Artificial Intelligence (AI)’s field of emotional perception is fast developing and has great potential to transform human-computer interactions and advance the creation of compassionate technologies. In this paper, we compare the empathetic features that are built into two complex AI models, ChatGPT and Gemini. Our study focuses on how well each of them can convert neutral statements into one of six different emotional categories: joyful, sad, confident, afraid, angry, caring.

. Through a careful examination of the output produced by ChatGPT and Gemini, our study aims to reveal the subtleties that underlie emotional intelligence in AI systems. Our goal is to shed light on how well these models can understand and express a wide range of emotional states. By doing this, we hope to uncover important information that will further our understanding of AI’s emotional intelligence and have significant ramifications for a variety of use cases. The study examines ChatGPT and Gemini’s ability to accurately represent human emotions, focusing on their empathic features. The findings have potential to guide AI-powered solutions, particularly in fields requiring user empathy. The study aims to illuminate new frontiers in psychological analysis and human-computer interaction.

**Key words:** Empathy, Emotions, Chat GPT, Gemini, AI

1. **INTRODUCTION**

Human cognition depends heavily on emotional intelligence, which facilitates social connections, efficient communication, and empathetic comprehension. The fast advancement of artificial intelligence (AI) technology has sparked a growing interest in endowing computers with comparable emotional comprehension and response capabilities. Two state-of-the-art artificial intelligence models, ChatGPT and Gemini, have become well-known for their ability to understand natural language. Though these models are excellent at producing language that is logical and relevant to the context, it is still unclear how well they can understand and communicate emotions. We train ChatGPT and Gemini to rephrase neutral statements into six different emotional categories: joyful, sad, confident, afraid, angry, caring. This research addresses the task of emotional understanding within AI systems. Using this classification framework, the objective is to assess the efficacy of both models in eliciting emotionally charged reactions and to quantitatively evaluate their empathetic qualities. This comparative analysis has important implications for the creation and creation of empathic AI technologies in addition to offering insights into the current cutting-edge in AI emotional comprehension. The study seeks to answer the following important questions: To what extent are neutral statements classified into different emotional groups by ChatGPT and Gemini? What distinguishes these two AI models’ capacities for emotional expressiveness from one another? In what ways do these results advance our knowledge of AI’s ability to comprehend emotions and the possible uses it may have in interactions between humans and computers and other fields? By examining these issues, the research hopes to further the conversation about emotional intelligence in artificial intelligence and open the door to the creation of more compassionate and emotionally intelligent AI systems. This research aims to highlight the present situation and the potential of AI in processing and reacting to emotions by examining ChatGPT and Gemini’s emotional comprehension abilities. The results have implications for human-computer interaction as well as AI model improvement. Emotionally intelligent computers can improve user experience, boost mental health, and encourage deeper relationships between humans and artificially intelligent systems.

1. PROBLEM STATEMENT The goal of the study is to examine and contrast how well two chatbots—ChatGPT and Gemini—recognize and react to six different categories of emotions that users may express in their inquiries. The following emotions have been recognized: joyful, sad, confident, afraid, angry, caring. The main issue is trying to figure out how effectively each chatbot can recognize and react to these feelings to gauge how empathic they are.
2. **LITERATURE REVIEW**

This section of the article discusses several research publications Schaaff et al. (2023) how ChatGPT behaviors are in line with empathetic reactions and emotional expressions, given the demand for the usage of chatbots across a variety of businesses. Jhan et al. (2021) trained with a Conceptual Human Model and deep reinforcement learning to react sympathetically to future emotional changes in the user.

1. OBJECTIVE OF THE SYSTEM

**Accuracy of Emotion Recognition:** Create a uniform collection of emotionally charged questions that cover the following emotional domains: joyful, sad, confident, afraid, angry, caring. Compare the ability of ChatGPT and Gemini to accurately recognize and classify user inquiries into the designated emotional categories to assess how accurate their emotion recognition. Analysis of **Empathic Response:** Examine and contrast each emotional category’s replies produced by ChatGPT and Gemini in terms of their empathy. Measure the state of mind and level of understanding that the chatbots are reflecting in their responses by using sentiment analysis techniques.

**Empathy Metrics Assessment:** Examine ChatGPT’s empathy capabilities, paying particular attention to how it interprets and expresses emotions. Use recognized measures of empathy, such as the Autism Spectrum Quotient, Empathy Quotient, the City of Toronto Empathy Survey Perth Empathy Scale, and Interpersonal Reactivity Index, to assess ChatGPT’s empathy quantitatively.

**Parallel Emotional Responding:** Examine how well Gemini and ChatGPT exhibit parallel emotional responses, in which their answers correspond with the emotional content that users have stated. Assess how well user-communicated sentiments and the chatbots’ sympathetic answers synchronize, considering the linguistic and psychological context.

**Empathy Comparison Analysis:** Create an extensive framework that integrates quantitative measures and emotional user feedback to compare the overall empathetic capacities of Gemini and ChatGPT. Point out the advantages and disadvantages of each chatbot’s empathetic traits, offering a detailed analysis of how well they understand and communicate emotions.

1. RESEARCH DESIGN AND METHODOLOGY

A diagram of a company

Description automatically generated with medium confidence

Fig. 1: Workflow

1. DATA COLLECTION
2. EXPLORATORY DATA ANALYSIS AND HYPOTHESIS
3. Size of the data

(14725, 8)

Check for NULL VALUES

After removing duplicates - 3386 rows x 8 columns

1. Information of the data
   1. conv\_id 14725
   2. utterance\_idx 14725
   3. context 14725
   4. prompt 14725
   5. speaker\_idx 14725
   6. utterance 14725
   7. selfeval 14725
   8. tags 14725
2. Statistical Analysis of the data

Mean Word Count and Max Word Count:

The mean word count of the text data is approximately 7.62 words per sentence.

The maximum word count in a sentence is 40 words.

Most Common Words:

The most common words in the text data are:

'got', 'going', 'really', 'get', 'happy', 'last', 'dog', 'time', 'work', 'day'.

These words appear frequently in the text data, indicating their importance or relevance.

Mean word count: 7.623449497932664

Max word count: 40

Most common words: [('got', 307), ('going', 272), ('really', 260), ('get', 250), ('happy', 237), ('last', 212), ('dog', 198), ('time', 190), ('work', 187), ('day', 171)]

Topic 0:

mad yesterday scared dog lost get work always broke night

Topic 1:

going work get really go happy someone day home test

Topic 2:

passed time away like ca one upset feel grandmother going

Topic 3:

got happy new last car get see really week night

Topic 4:

dog friend care job took died years sick got felt

**CARER Data**

Mean word count: 9.428670055858909

Max word count: 37

Most common words: [('feel', 10849), ('feeling', 5279), ('like', 2724), ('im', 2303), ('really', 985), ('know', 878), ('time', 830), ('little', 807), ('get', 800), ('people', 706)]

Topic 0:

feel feeling like life im people time bit much know

Topic 1:

feel like really feeling im know get dont time things

Topic 2:

feel feeling like little would time think even get im

Topic 3:

feeling im feel little today http bit href ive pretty

Topic 4:

feel feeling like one people im really want could something

1. Exploratory Data Analysis

A close up of words

Description automatically generated

These words appear frequently in the text data, indicating their importance or relevance.

The word cloud provides a fascinating snapshot of personal experiences, emotions, and relationships. Here are some key insights:

Dominant Themes:

Family and Relationships: Words like "family," "friend," "husband," "wife," "mom," "dad," "daughter," and "son" highlight the importance of family and close relationships in the author's life.

Emotions: A range of emotions are expressed, including "happy," "love," "scared," "angry," "sad," and "upset." This suggests a person experiencing a variety of feelings and navigating life's ups and downs.

Daily Life and Activities: Words like "work," "school," "car," "home," "week," and "day" point to the routines and activities that shape daily life.

Time and Change: References to "time," "last night," "today," "tomorrow," "week," and "year" indicate a reflection on the passage of time and the changes it brings.

Additional Observations:

Loss and Grief: The presence of words like "lost," "passed away," and "died" suggests the individual may have experienced loss and is processing grief.

Uncertainty and Worry: Words like "afraid," "scared," "upset," and "crying" indicate moments of fear, anxiety, and emotional distress.

Hope and Positivity: Despite challenges, words like "happy," "love," "good," "best friend," and "excited" suggest a hopeful outlook and an appreciation for positive aspects of life.

Personal Growth: The inclusion of "know," "think," "feel," and "see" may imply introspection and a journey of self-discovery.

Overall Impression:

The word cloud paints a picture of a person navigating the complexities of life, experiencing a range of emotions, cherishing relationships, and reflecting on personal experiences. It suggests a journey with both challenges and joys, losses and hopes, ultimately revealing the richness and depth of human experience.

A close up of words

Description automatically generated

Dominant Themes:

Emotions: The overwhelming presence of words like "feel," "feeling," "love," "happy," "sad," "scared," and "angry" indicates a deep exploration of emotions and their impact on the individual's life.

Self-Reflection: Words like "think," "know," "want," "need," and "make" suggest a process of introspection and self-discovery. The individual is actively trying to understand their own thoughts, desires, and motivations.

Relationships: References to "friend," "people," "someone," "one," and "everyone" highlight the importance of human connection and the impact of relationships on the individual's emotional state.

Communication and Expression: Words like "say," "tell," "write," "talk," and "blog" suggest a desire to communicate and express emotions, thoughts, and experiences.

Additional Observations:

Vulnerability and Uncertainty: Phrases like "feel lost," "feel strange," "didn't know," and "might come" reveal moments of vulnerability, uncertainty, and questioning.

Hope and Positivity: Despite challenges, words like "love," "happy," "good," "sweet," and "amazing" suggest an underlying sense of hope and an appreciation for positive experiences.

Change and Growth: References to "change," "first time," "new," and "better" indicate a willingness to embrace change and personal growth.

The Power of Words: The emphasis on words related to communication and expression suggests a belief in the power of language to connect, heal, and inspire.

Overall Impression:

The word cloud portrays an individual on a journey of self-discovery and emotional exploration. They are navigating a complex world of feelings, seeking connection with others, and striving to understand themselves better. The emphasis on communication and expression suggests a desire to share their experiences and connect with others on a deeper level.

This word cloud ultimately speaks to the universal human experience of navigating emotions, seeking meaning, and finding one's place in the world.

1. DATA VISUALIZATION AND RESULTS

Using 90 records from 6 different

Persona Pattern

A graph of different colored rectangular shapes

Description automatically generated with medium confidence

Fig: Overall CARER Data Accuracy

A graph with numbers and lines

Description automatically generated with medium confidence

Fig: Individual Accuracy for emotions Anger, Fear, Joy, Love, Sadness, and Surprise in CARER data.

A screenshot of a graph

Description automatically generated

Fig: Overall Facebook empathic data accuracy

A screenshot of a graph

Description automatically generated

Fig: Individual Accuracy for emotions Afraid, Angry, Caring, Confident, Joyful, and Sad in Facebook empathic data.

1. LIMITATIONS

**Open AI Rate Limit Error:**

When encountering a RateLimitError like the one you've received, it's essential to understand the constraints placed on your usage of the GPT-3.5 model. This error indicates that the daily limit for requests per day (RPD) has been reached within your organization. In this case, the limit stands at 200 requests, all of which have been utilized. Consequently, any additional requests will be temporarily blocked until the limit is reset. This restriction poses a challenge for users who rely on the model's capabilities for various tasks, such as content generation, analysis, or creative writing.

To mitigate this issue, users can take several actions. Firstly, they should wait for the specified cooldown period mentioned in the error message, which in this case is 7 minutes and 12 seconds. During this time, users can consider optimizing their usage patterns to ensure that they stay within the allocated limits. Additionally, users have the option to explore increasing their rate limit by adding a payment method to their OpenAI account. This step enables users to access higher tiers of service, granting them a more generous allocation of requests per day. By investing in their account's billing options, users can align their usage with their specific needs and scale their activities accordingly.

Moreover, it's crucial for users to monitor their usage habits regularly to avoid exceeding the allocated limits and encountering further interruptions. This can involve implementing efficient request management strategies, prioritizing tasks, and utilizing the model judiciously to maximize its utility within the constraints of the allocated resources. By adopting these proactive measures, users can optimize their experience with the GPT-3.5 model while maintaining compliance with usage limits and ensuring uninterrupted access to its capabilities.

In summary, encountering a RateLimitError underscores the importance of understanding and managing the constraints imposed on usage. By leveraging the provided information and exploring available options, users can navigate these limitations effectively, ensuring a seamless experience with the GPT-3.5 model while maximizing its utility for various tasks and applications.

**Gemini limitation with safety ratings:**

Safety ratings are crucial in ensuring responsible AI development and deployment. They help identify potential risks associated with AI models, such as bias, toxicity, and harmful content generation. However, relying solely on safety ratings has limitations, as highlighted by the error message "ValueError: The response.text quick accessor only works when the response contains a valid Part, but none was returned. Check the candidate.safety\_ratings to see if the response was blocked." This error indicates that the AI model's response was flagged and blocked due to potential safety concerns. While this safety mechanism is essential, it can also hinder exploration and limit the model's ability to learn and grow.

**Understanding Blocked Responses:**

The blocking of responses occurs when the safety rating system detects potential issues like hate speech, discrimination, or the generation of harmful content. This is a necessary precaution to prevent the spread of harmful information and ensure responsible AI behavior. However, it's important to recognize that safety ratings are not infallible. They can be overly cautious, leading to false positives where harmless content is mistakenly flagged. This can restrict the model's ability to engage in open-ended conversations or explore creative text formats, ultimately hindering its development and potential.

**API Calls response:**

Gemini API and OpenAI API enforce a restriction that limits the frequency of API calls to once every 30 seconds. This constraint is in place to regulate the rate at which users can access the API endpoints, ensuring fair and efficient utilization of the available resources. By imposing this limitation, both APIs aim to prevent excessive traffic and mitigate the risk of system overload, which could lead to performance degradation or service interruptions.

The 30-second interval serves as a cooldown period between consecutive API requests, allowing the system sufficient time to process incoming queries, execute requested tasks, and manage internal operations effectively. Adhering to this predefined timeframe is essential for maintaining the stability and reliability of the API services while accommodating the needs of diverse user groups.

Compliance with the 30-second limitation necessitates careful planning and coordination within applications or systems that rely on these APIs. Developers must implement mechanisms to regulate the timing of API calls, incorporating logic to enforce the specified cooldown period between successive requests. Additionally, users should consider optimizing their workflows and batch-processing tasks to minimize the frequency of API interactions, thereby maximizing efficiency and minimizing potential disruptions.

Overall, understanding and adhering to the 30-second limitation imposed by the Gemini API and OpenAI API is essential for ensuring smooth and uninterrupted access to these valuable resources. By observing this constraint and adopting best practices for API usage, developers and users can harness the full potential of these platforms while promoting a sustainable and harmonious interaction with their respective services.

1. CONCLUSION

6. DATASET The CARER dataset on emotions was gathered using noisy labels and annotated remotely. The subset of data presented here is consistent with the paper’s description of the six emotions variant. joyful, sad, confident, afraid, angry, caring are the six emotions. The Empathetic Dialogues dataset comprises 24,850 one-to-one open-domain discussions and is a large-scale multi-turn sympathetic dialogue dataset gathered on Amazon Mechanical Turk. Someone who spoke and a listener were paired with other crowd workers to create each conversation. The speaker is requested to discuss their own emotional experiences. Through the speaker’s words, the listener deduces the underlying feeling and reacts sympathetically. There are thirty-two uniformly distributed emotion labels in the sample.

REFERENCES

Jhan, J.-H., Liu, C.-P., Jeng, S.-K., and Lee, H.-Y. 2021, Cheer Bots: Chatbots toward Empathy and Emotion using Reinforcement Learning.

Schaaff, K., Reinig, C., and Schlippe, T. 2023, in 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII), 1–8