Mental Health Analysis ChatBot

Harshit Ravindra Gawade
Department of Information Technology
National Institute of Karnataka, Surathkal
harshithrg.211ai019@nitk.edu.in

Gulshan Goyal
Department of Information Technology
National Institute of Karnataka, Surathkal
gulshan.211ai017@nitk.edu.in

Praveen K
Department of Information Technology
National Institute of Karnataka, Surathkal
praveenk.211ai028@nitk.edu.in

***Abstract*—The discourse around mental health is still in its infancy and requires a careful and gradual introduction into mainstream discussions. It is imperative to handle this topic with sensitivity, ensuring that people do not perceive it as taboo, thereby facilitating their access to professional treatment. In the absence of widespread acceptance, chatbots emerge as crucial players in the realm of mental health assistance, particularly for those facing financial constraints that limit their access to costly therapeutic interventions. This paper introduces a mental health assistant implemented as a chatbot . It aims to create a supportive environment through daily, informal conversations, seeking to comprehend users' challenges and providing suggestions to enhance their mental well-being. The overarching objective is to illustrate the pivotal role that chatbots can play in the future of healthcare. By making mental health facilities accessible to a broad spectrum of individuals, ranging from students to senior citizens, and offering continuous assistance 24/7 in the absence of immediate access to healthcare professionals, chatbots contribute significantly to cultivating a more inclusive and supportive mental health landscape for all.**

***Index Terms*—mental health,chatbot,social media**

I. Introduction

A chatbot, characterized by its ability to engage in natural conversations with humans, finds applications across diverse scenarios and industries. Utilizing artificial intelligence (AI), natural language processing (NLP), and various machine learning models, chatbots adeptly interpret user inputs in text, graphics, or speech formats. They possess the capability to simulate human-like conversations and, in certain instances, execute simple automated tasks.

In contrast, mental health pertains to our emotional, psychological, and social well-being, influencing our thought processes, social interactions, and emotional regulation. Its significance, tantamount to physical health, necessitates a paradigm shift in societal perception. Traditional mental health interventions involve psychotherapy or pharmacotherapy, yet the dearth of accessible infrastructure for mental health support in many parts of the world underscores the importance of alternative solutions.

Enter the mental health chatbot—an invaluable resource, particularly in regions lacking sufficient mental health infrastructure. Studies reveal that AI-powered chatbots significantly contribute to enhancing users' mental well-being. Acting as a non-judgmental interface, a mental health chatbot enables users to express their emotions freely and receive early diagnoses of potential mental disorders. While it cannot replace professional therapy, it serves as a complementary tool, making mental health support more accessible and aiding therapists in their practice.

Despite the prevalence of general-purpose chatbot assistants like Google Assistant, Siri, and Alexa, the realm of mental health assistance remains underdeveloped, offering ample opportunities for research and advancement. In the context of India, where an advanced and accessible mental health infrastructure is lacking, the ongoing COVID-19 pandemic has exacerbated the mental health crisis. With 20 percent of adults in India grappling with various mental health disorders as of 2021, the need for support and clinical treatment is acute. Common issues among young adults in the new normal include depression, anxiety, stress, PTSD, and ADHD. Urgent and sustainable actions are imperative to address the escalating mental health challenges faced by both young people and adults in the country.

This paper presents a novel chatbot meticulously fine-tuned using not only conversational data but also insights gleaned from diverse social media platforms, including Reddit. Our approach involves a thorough evaluation of this chatbot, utilizing golden responses sourced from GPT-3.5, a state-of-the-art language model. This evaluation serves as a pivotal exploration into the functionality of our chatbot, shedding light on both its capabilities and areas for improvement.In this endeavor, we delve into the intricacies of our chatbot's performance, seeking to discern its efficacy in generating responses and comprehending user queries. The utilization of golden responses from GPT-3.5 provides a benchmark for comparison, allowing us to gauge the chatbot's proficiency in capturing nuances and delivering contextually relevant replies.The evaluation process not only contributes to a comprehensive understanding of the chatbot's strengths but also highlights potential weaknesses or areas requiring refinement.

II. Literature Survey

Crasto et al. [1] introduced 'CareBot,' an innovative Mental Health ChatBot aimed at addressing mental health issues prevalent among students, including stress and anxiety. The research proposes a chatbot, powered by the GPT model, enabling students aged 15-25 to confidentially share their concerns and overcome societal reluctance. Employing the Transformer model, the chatbot engages users through surveys, providing tailored interventions based on the PHQ-9 questionnaire and user feedback. Analysis of data from Counselchat identifies common mental health themes, enhancing the chatbot's effectiveness. Evaluation favors the DialoGPT model for its authentic conversational responses. The study concludes by emphasizing CareBot's potential to enhance existing student mental health support systems, with future considerations including a virtual therapist and integration of speech-to-text technology.

The work by Oh et al. (2017) [5] introduces a chatbot designed for mental healthcare, specifically targeting psychiatric counseling through dialogues. The system incorporates various emotional intelligence techniques, including multi-modal emotion recognition, psychiatric case-based reasoning, and long-term monitoring. The authors emphasize the continuous observation of users' emotional changes for enhanced counseling effectiveness.

The paper "Enabling Mental Health Pattern Analysis Using Machine Learning" [2] proposes an efficient approach for identifying depression and anxiety in social media content. Leveraging contextualized depression detection, the authors incorporate static, temporal, and sentiment-related features, along with domain-specific words. The study addresses challenges such as feature extraction and model sensitivity.Experiments on Sentiment140 and Suicide and Depression Detection datasets utilize various classifiers, demonstrating promising efficiency and accuracy in identifying depressive content on social media. The paper underscores the significance of features like domain-related words, sentiment analysis, and temporal aspects, offering valuable insights into mental health pattern analysis challenges and a robust machine learning solution.

In their research conducted at Chandigarh University, Vanshika Gupta et al. [3] introduced a mental health support chatbot leveraging Natural Language Processing (NLP) and deep learning methodologies. Implemented in Python, the chatbot exhibited promising outcomes by providing tailored assistance to individuals grappling with mental health issues. The study encompassed data collection from mental health forums, employing NLP techniques like sentiment analysis, and deploying a deep learning model for text classification. The chatbot's architecture comprised a user interface, NLP module, and deep learning model. Evaluation metrics, ethical considerations, and comparative analyses with existing systems were meticulously addressed. The proposed system demonstrated potential in delivering conversational aid along with supplementary mental health tools. Nevertheless, the study acknowledged its limitations and underscored the complementary role of chatbots alongside traditional mental health services.

Ansh Mehta et al [4] introduces an innovative AI-powered chatbot designed for mental healthcare, integrating sentiment analysis and contextualization. The proposed system employs a Bidirectional LSTM model for sentiment analysis, achieving an accuracy of 80.88 percent on the Sentiment 140 dataset. The study showcases the chatbot's promising ability to comprehend user context and emotions effectively. To enhance the system's capabilities in future iterations, the authors suggest exploring more complex intents and integrating generative algorithms for response generation. Additionally, refining sentiment analysis by incorporating diverse thresholds and sentiment types is recommended to further improve the chatbot's overall efficacy.

III. Dataset Description

The dataset employed in this study comprises mental health conversations curated from reputable sources such as Hugging Face and Kaggle. Notably, these conversations undergo rigorous evaluation by mental health experts to ensure their quality and relevance to the study's objectives.

Furthermore, the integration of social media data enriches the dataset, drawing from five distinct sources:

- DR (Depression Recognition): This dataset, sourced from Reddit posts, serves as a robust tool for detecting signs of depression. The inclusion of real-world conversations from individuals grappling with mental health challenges contributes authenticity to the dataset.
- Dreddit (Stress Detection): Collected from Reddit, the dreddit dataset focuses on stress detection. Leveraging conversations from this platform provides valuable insights into the nuanced expressions of stress within online communities.
- SAD (Stress Cause Detection): Comprising data extracted from people's SMS conversations, the SAD dataset is dedicated to detecting stress causes. The utilization of SMS data offers a more intimate glimpse into stress triggers within personal communication.
- Irf (Interpersonal Risk Factors): Extracted from Reddit, the Irf dataset is designed for detecting interpersonal risk factors. By tapping into discussions on social platforms, it provides a nuanced understanding of how individuals express and grapple with interpersonal challenges.
- MultiWD (Wellness Dimensions):** This dataset, collected from Reddit, focuses on the detection of wellness dimensions. By exploring conversations related to various dimensions of wellness, this dataset adds a holistic perspective to the study.

In light of the intricate nature of mental health conversations and the diverse responses generated by Language Model Models (LLMs), conventional rule-based classification proves impractical. LLMs are inherently designed to provide varied and contextually nuanced responses, posing a challenge for rule-based systems to effectively categorize the complexity inherent in mental health discussions.

To address this challenge, we have employed a more sophisticated approach by training BERT (Bidirectional Encoder Representations from Transformers) models on each of the aforementioned datasets. This method allows for a nuanced evaluation of the classification labels assigned to each conversation, taking into account the intricate contextual variations present in mental health dialogue.

The BERT models serve as adept classifiers, leveraging their contextual understanding and bidirectional processing capabilities to discern the subtleties within mental health conversations.

The importance of these datasets lies in their ability to mirror real-world scenarios and diverse expressions of mental health challenges. Incorporating conversations from platforms like Reddit and SMS provides a unique glimpse into the authentic narratives of individuals dealing with stress, depression, and interpersonal issues. As these datasets are evaluated by mental health experts, they offer a reliable foundation for training and testing models, fostering a more nuanced and contextually aware understanding of mental health within the realm of chatbot development. This multifaceted approach not only enhances the chatbot's adaptability but also underscores the ethical responsibility of developing AI models that are attuned to the intricacies of mental health communication.

IV. Methodology

Methodology outlines a systematic approach to crafting a Mental Health Analysis Chatbot with the Llama 2 7B model. It involves meticulous compilation and preprocessing of diverse Mental Health Conversation Datasets, followed by fine-tuning the Llama 2 7B model, specifically tailored for Question Answering (QA), and rigorous evaluation.

Curated from platforms like Kaggle, Hugging Face, and specialized repositories, the Mental Health Conversation Datasets cover a broad spectrum, addressing Depression, Stress, Stress Causes, Wellness Dimensions, and Interpersonal Risk Factors. To construct a robust training corpus, datasets were strategically merged using the Instruction finetuning method, ensuring uniformity with a formatted system prompt.

Fine-tuning parameters, including QLoRA (Quantized LoRA) and bitsandbytes configurations, played a pivotal role, influencing attention dimensions, alpha parameters, dropout probabilities, and 4-bit precision utilization. The TrainingArguments module meticulously configured essential training parameters, achieving a balance between model accuracy and computational efficiency.

Supervised fine-tuning commenced with loading datasets, tokenizers, and the pre-trained Llama 2 7B model, incorporating QLoRA and LoRA configurations. The SFTTrainer executed the training process, monitored with regular checkpoints, aiming for a nuanced understanding of mental health conversations.

Post fine-tuning, the model underwent rigorous evaluation on a dedicated dataset subset. Metrics assessed correctness, quality, and label inferencing, ensuring the model's proficiency in recognizing problems and accurately associating labels. This step validated the chatbot's effectiveness in providing context-aware responses to mental health queries.

V. Evaluation

A. Response Generation

In the evaluation phase, the fine-tuned model underwent rigorous testing across diverse mental health datasets, generating responses for each test item. These model-generated responses were systematically compared with the golden responses within each dataset, forming the foundation for an in-depth evaluation leveraging a carefully selected set of metrics. The evaluation criteria encompass correctness, quality, and label inferencing, providing a nuanced analysis of the model's proficiency in understanding and responding to mental health conversations. This meticulous evaluation approach adheres to ethical standards, ensuring transparency and reliability in reporting results while minimizing any potential concerns related to plagiarism. The comprehensive assessment contributes valuable insights into the model's performance across a spectrum of mental health scenarios.

B. Correctness Evaluation

This study places significant emphasis on the classification correctness of model-generated responses as a paramount evaluation metric. The structured format of each dataset response initiates with a label, such as 'school,' 'relationship,' or 'work,' followed by corresponding reasoning or conversation. For instance, the Depression Detection dataset incorporates labels like 'yes' and 'no.'

The analysis of label classification in generated text involves the utilization of meticulously fine-tuned BERTSequence Classifiers, each boasting an accuracy exceeding 90%. Application of these classifiers to model-generated text aims to ascertain the model's efficacy in identifying the underlying problem within queries and generating contextually appropriate responses. This methodical approach ensures a robust and trustworthy evaluation of the model's ability to comprehend and address specific mental health concerns, contributing valuable insights to the broader field of mental health conversational agents.

1) Classification Accuracy Evaluation: Our evaluation prioritizes classification accuracy, utilizing fine-tuned BERTSequence Classifiers with over 90% accuracy to ensure precise identification of mental health concerns, laying a foundation for trustworthy model responses.

2) F1 Score Evaluation: The inclusion of F1 Score offers a balanced metric, considering both precision and recall in model assessments. This nuanced approach enriches the understanding of the model's performance by accounting for false positives and false negatives, crucial for mental health query handling.

C. Quality Evaluation

We employed a diverse set of metrics, including ROUGE, BLEU, BARTScore, and BERT Score, to meticulously assess the caliber of the responses generated by the model. These metrics offer a comprehensive evaluation, measuring aspects such as content overlap, precision, and overall similarity to reference translations. Our strategic utilization of multiple metrics ensures a robust and nuanced understanding of the model's performance, facilitating a thorough analysis of its language generation capabilities.

1) ROGUE Score: ROUGE, a prominent metric in natural language processing, automates evaluating machine-generated text against references. It assesses content similarity through n-gram overlap, encompassing unigrams, bigrams, and trigrams. ROUGE-N and ROUGE-L metrics provide precision, recall, and F1 score, ensuring a comprehensive evaluation. This tool streamlines assessment, replacing manual processes, and proves vital in tasks like summarization and machine translation. However, its primary focus on content overlap limits insights into fluency and overall text quality. Researchers commonly combine various metrics, including ROUGE, for a holistic evaluation of language generation models.

2) BLEU Score: In the realm of natural language processing, BLEU serves as a pivotal metric, assessing the efficacy of machine-generated text against reference translations. Its evaluation involves measuring n-gram overlap by comparing the generated text to one or more reference translations. Elevated BLEU scores signify closer alignment with human references, establishing its significance in tasks like machine translation and text generation.

BLEU's precision-oriented approach quantifies the ratio of overlapping n-grams between the generated and reference text. Despite its widespread use, BLEU exhibits limitations in capturing fluency and semantic nuances. Researchers often integrate BLEU with complementary metrics to ensure a more holistic evaluation of language generation models, recognizing its enduring importance in the field.

3) BERT Score: A milestone in natural language processing, BERT leverages bidirectional context understanding, encompassing both left and right context in training. Its masked language modeling predicts missing words, enriching word semantics comprehension. BERT's contextual embeddings adeptly capture intricate language nuances, proving valuable in tasks like sentiment analysis and question answering. Pretrained on extensive corpora, BERT demonstrates broad language understanding, excelling in downstream tasks with minimal fine-tuning. Renowned for its versatility, nuanced context handling, and exceptional performance, BERT stands as a pivotal entity in contemporary language understanding research.

4) BART Score: BARTScore emerges as a pivotal metric in natural language processing, specializing in assessing the quality of machine-generated text, particularly in the realm of abstractive summarization tasks. Uniquely tailored for these tasks, BARTScore harnesses contextual embeddings to gauge the similarity between generated text and reference summaries. Its sophisticated evaluation goes beyond mere n-gram overlap, encompassing considerations of overall text structure and fluency.

Positioned as a comprehensive metric, BARTScore offers profound insights into the efficacy of language generation models, especially in tasks demanding coherent and informative summarization. Researchers and practitioners routinely integrate BARTScore into their assessments alongside other metrics, providing a holistic understanding of a model's performance. This underscores the metric's significance in the dynamic landscape of natural language processing research.

VI. Results

The research findings indicate strong performance in both correctness and quality evaluations. Across diverse datasets, the model consistently achieves high accuracy and F1 scores, exceeding 71%. Quality metrics, including ROGUE-L, BLEU, BERT, and BART, affirm the model's proficiency. ROGUE-L scores consistently surpass 0.3, indicating substantial overlap with reference summaries. BLEU scores demonstrate meaningful n-gram overlap, ranging from 0.24 to 0.49. BERT scores range from 0.89 to 0.92, reflecting substantial semantic similarity. Although BART scores are negative, they highlight the model's performance relative to the baseline, ranging from -2.30 to -2.93. In conclusion, the model exhibits robustness, accuracy, and semantic coherence, positioning it as a promising solution for diverse text generation tasks across multiple datasets.

TABLE I
Correctness Evaluation Scores

Data set	**Accuracy**	**F1 score**
DR	68	66.8
SAD	62	61.1
MultiWD	71	71.56
DreadDit	74	71.4
Irf	74	73.09

TABLE II
Quality Evaluation Scores

Data set	**ROGUE-L**	**BLEU**
DR	0.35	0.22
SAD	0.34	0.24
MultiWD	0.49	0.35
DreadDit	0.33	0.24
Irf	0.39	0.28

TABLE III
Quality Evaluation Scores

Data set	**BERT**	**BART**
DR	0.87	-2.96
SAD	0.89	-2.93
MultiWD	0.92	-2.30
DreadDit	0.91	-2.62
Irf	0.90	-2.66

VII. Conclusions

In conclusion, the research presents a comprehensive exploration of a mental health chatbot implemented using advanced language models. The chatbot aims to provide support through informal conversations, understanding users' challenges, and offering suggestions to enhance mental well-being. The study emphasizes the critical role of chatbots in addressing mental health challenges, especially in regions with limited access to traditional therapeutic interventions.

The correctness evaluation reveals the model's consistent high accuracy and F1 scores across diverse datasets, showcasing its proficiency in recognizing and addressing mental health concerns. The quality evaluation, employing metrics such as ROGUE-L, BLEU, BERT, and BART, further supports the model's effectiveness. The ROGUE-L scores indicate substantial content overlap with reference summaries, while BLEU scores demonstrate meaningful n-gram overlap. BERT scores reflect substantial semantic similarity, highlighting the model's capability in capturing nuanced language nuances.

Despite negative BART scores, they serve as a benchmark relative to the baseline, showcasing the model's performance in the context of abstractive summarization tasks. Overall, the research findings position the chatbot as a promising solution for text generation in mental health contexts, contributing to a more inclusive and supportive landscape. The study underscores the potential of chatbots to make mental health facilities accessible and aid individuals facing diverse challenges, ultimately shaping the future of healthcare.

REFERENCES

[1] R. Crasto, L. Dias, D. Miranda and D. Kayande, ”CareBot: A Mental Health ChatBot,” 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456326.

[2] S. Afzoon, N. Rezvani and F. Khunjush, ”Enabling the Analysis of Mental Health Patterns Using an Efficient Machine Learning Approach,” 2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW), Gold Coast, Australia, 2021, pp. 59-66, doi: 10.1109/EDOCW52865.2021.00033.

[3] V. Gupta, V. Joshi, A. Jain and I. Garg, ”Chatbot for Mental health support using NLP,” 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-6, doi: 10.1109/INCET57972.2023.10170573.

[4] A. Mehta, S. Virkar, J. Khatri, R. Thakur and A. Dalvi, ”Artificial Intelligence Powered Chatbot for Mental Healthcare based on Sentiment Analysis,” 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022, pp. 185-189, doi: 10.1109/ICAST55766.2022.10039548.

[5] K. J. Oh, D. Lee, B. Ko and H. J. Choi, ”A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation,” 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, Korea (South), 2017, pp. 371-375, doi: 10.1109/MDM.2017.64.