

# e-Therapist: Conversational Agent for Mental Health

Sneha Oram\*, Pushpak Bhattacharyya

## Abstract

India faces a significant mental health crisis with a projected 20% of the population expected to suffer from mental illnesses by the end of the year. However, there are only 0.3 psychiatrists, 0.07 psychologists, and 0.07 social workers per 100,000 people approximately. A conversational agent can hence step in to bridge the gap in the high ratio between professionals and mental health support seekers. We propose **e-therapist**, a cordial conversational agent, that can act as a first point of contact for the support seekers. It can work with 4 mental health conditions anxiety, depression, OCD (obsessive-compulsive disorder), and PTSD (Post Traumatic Stress Disorder). The results are obtained by fine-tuning a few pre-trained models. Comparative analysis shows that the BLEUScore, Perplexity, and BERTScore F1 on the generated response are 0.44, 190.73, and 0.4207, which are competitive with the state-of-the-art performance. The results point out that the Mistral model performs better in the conversational systems related to mental health. The code is available in the GitHub <sup>1</sup>.

## 1 Introduction

Mental health conditions account for 30% of the non-fatal disease burden, globally, by the World Health Organization (WHO). WHO also estimates depression and anxiety disorders cost the global economy \$1 trillion per year in lost productivity. These data motivate a need for robust diagnostic and mitigation systems for mental health conditions, to alleviate individual suffering and reduce subsequent societal and economic burdens associated with mental health issues.

Often people turn up for conversational agents where they can express their emotions and experiences (generally stigmatized). There are a variety of peer-to-peer platforms that offer text-based interactions, however, they lack best therapy practices.

It's been decades since the idea of a conversational agent that can handle general queries. Nowadays there are chatbots on websites of e-commerce industries, aviation industries, and several other organizations and industries. However, a conversational agent for mental health has a different set of characterization, principles, and goals.

Category	Description listed in ICD-10
MDD	mood disorder, hopelessness, worthlessness, lack of energy, reduced activity
OCD	repeated unwanted thoughts (obsession), urge to continuously repeat something (compulsion)
Anxiety	nervous disorder, worry, uncontrollable racing thoughts, difficulty in concentrating, sleeping
PTSD	recurrent distressing memories of the traumatic event, negative thoughts about oneself, difficulty maintaining close relationships

Figure 1: Description of different mental disorders (Saha et al., 2022b)

A mental health condition is built on top of sentiment and emotion dynamics. The mental health support seekers or the posters in the chatbots showcase a range of valence, arousal, and dominance in their emotions. It is important in the part of the model to be empathetic in its response while dealing with such situations. Empathy is believed to be the primary emotional skill that is to be incorporated in the conversational agents. (Saha et al., 2022a) has worked on incorporating motivation and empathy in the model's response in a virtual assistant. Previous research has tried incorporating motivational, empathetic, and politeness in the responses, generated by Conversational AI. (Mishra et al., 2023) introduced a persona-aware conversational agent that generates polite responses with appropriate interpersonal behavior (understanding, helpful, empathetic, etc).

Most recently Large Language Models (LLMs) (Brown et al., 2020); (Ouyang et al., 2022), especially ChatGPT 2 and GPT-4 (OpenAI, 2023), have exhibited strong general language processing ability to mental health-related text. Four kinds of mental health conditions are of concern in this work

<sup>1</sup>[https://github.com/Snehauser/e-Therapist\\_R-D](https://github.com/Snehauser/e-Therapist_R-D)

MDD (Major Depressive Disorder), PTSD, anxiety, or OCD, and details are mentioned in brief. Quantitative research shows that instilling optimistic behavior fueled by hope and motivation improves symptoms in terms of positive psychological transformation and a favorable alliance in mental health support (Jahanara, 2017).

Previous research has tried to separately evaluate user utterances and CA (Conversational Agent) utterances. To form correct encoding and draw conclusions, the utterance encoding goes through the attention layer (or HAN-Heirarchical Attention Network). (Saha et al., 2022a) has taken the users' utterances for classification tasks using the Double Attention System. As the sophisticated LLMs in recent times are exhibiting excellent linguistic and semantic cues, this study has not used attention. This study in this paper is focused on extensive data analysis, and comparative analysis between LLMs on the custom real-world dataset. Hence, we summarize our contributions as -

1. Performed extensive data analysis on the dataset with sentiment polarity and word cloud in the users' utterances.
2. Performed comparative analysis between LLMs namely Llama2-7b and Mistral-7b on the custom dataset.

## 2 Related Work

(Hua et al., 2024) has discussed how one of the major research topics in the field of LLM in Mental Health is dedicated to intelligent chatbots. The chatbots can provide a range of mental health services like analysis, prediction, and support. (Ma et al., 2023) developed a conversation app for offering on-demand, and non-judgmental support that can boost user confidence and aid self-discovery. Because of the stigma, the LLMs employing mental health support have to be private in handling users' information, ensuring responsible and effective application. (Lai et al., 2023) leveraged LLMs for question-answering tasks in psychological consultation settings. The authors show Psy-LLM framework that can serve as a front-end tool for healthcare professionals.

Some researchers have also worked on developing dialogue agents to provide therapeutic support "therapybots" (Fitzpatrick et al., 2017). Empathy is one of the major cues that the chatbot should impart in a conversation with a user. (Morris et al.,

*Sample Conversation-1*  
**User1** > I am really struggling to not take more sleeping pills and just pass out. I want to not wake up for a week. I feel worthless of myself not doing anything. I want to take the easy way out and just slip away...  
**VA** > That's not you talking User1, it's the depression. Give yourself some TLC....do something you love like.....You are here for a reason that you're not allowed to know about. Hugs, and keep venting...  
**User1** > I doubled my dose to sleep. But I'm still awake and typing fine.  
.  
**VA** > It's very easy to go from 2 to 3 to 4. The recommended dosage is there for a reason and unless you are a medical professional, you should follow what they say.  
.  
**VA** > You need to take responsibility for your thoughts, feelings, and behavior. It is not our fault we have this disease, we are not to blame. But we have to be responsible. Own it. Take care of yourself.

Figure 2: Sample conversation from MDD thread (Saha et al., 2022b)

2018) has worked in incorporating empathy in psychotherapeutic text-based conversations. There are a few studies that have explored politeness in improving the sense of empathy and compassion in a conversation ((Kim et al., 2018);(Lucas et al., 2014)). The competence role in responses to help-seeking posts on mental health was investigated in (Lahnala et al., 2021). (Althoff et al., 2016) investigated large-scale counseling dialogue and presented an SMS text-based counseling service.

The majority of the previous research emphasizes the stigma and prejudice in society, which points to the responsible application of LLMs in mental health. NLP-based mental is still an ambiguous and unexplored subject, and there still exists a huge gap between NLP applications and mental health services.

## 3 Dataset

A conversational dataset, Motivate has been used for training and inferencing. The Motivate dataset contains 7067 dyadic conversations in 4 domains of MDD (Major Depressive Disorder), PTSD, anxiety, or OCD. This dataset has leveraged a peer-to-peer platform Psychcentral<sup>2</sup>, to collect dyadic conversational data. The dataset statistics and the sample distribution in illnesses are mentioned in Figure 3. Figure 2 shows the sample conversation in the dataset. It shows that the CA has to give away simple, and relevant texts imparting comfort and assistance to the posters. This is how the dataset is curated with CA outputs.

The majority of the datasets dealing with mental health come with limitations. Some of the databases use social media content of anonymized users expressing their thoughts and feelings through posts with no specific dyadic or

<sup>2</sup>[www.psychcentral.org](http://www.psychcentral.org)

multi-party discussions to draw on. Standard conversational datasets are slowly emerging (example) with featured pairs of seeker posts and supporter responses, without dialogic structure. This is taken as a motivation by (Saha et al., 2022c), who curated it, giving rise to the MotiVate dataset. The base dataset is provided by Psychcentral, which is a peer-to-peer text-based forum. In this forum, anonymous individuals can talk about their mental health problems and get help and advice from others who have had similar emotions, difficulties, and grievances. There are subforums that deals with different mental health domains like anxiety and panic attacks, schizophrenia, and so on. For data collection, 10k multi-party interactions are considered from four distinct subforums: OCD, Anxiety, PTSD, and MDD are gathered.

In Figure 3 the details of the dataset are mentioned briefly. Each dialogue represents alternate user and virtual assistant utterances collected and curated from a peer-to-peer text conversational platform.

Criteria	Statistics				
	Total	MDD	OCD	Anxiety	PTSD
# of dialogues	7067	4046	1000	1000	1021
# of utterances	25947	16257	2461	2784	4445
# of utterance per dialogue (avg.)	3.67	4.01	2.46	2.79	4.36
# of utterance per dialogue (max.)	129	129	14	16	25
Maximum user utterance length (# of words)	3319	3319	1337	1028	2112
Maximum VA utterance length (# of words)	2869	2851	2869	1024	2116
# of unique users	2139	1060	349	323	407
# of unique words	56336	35666	14108	12135	16427

Figure 3: Dataset statistics (Saha et al., 2022b)

**Data preprocessing:** The dataset has text in a dialogue format with alternate user and VA (Virtual Assistant) utterances. To bring the dataset into the standard conversational data format given by the openassistant/guanaco dataset. The labels of the utterances that is the user ID and VA are changed to Human and Assistant. All the datasets in 4 domains of mental health are combined to form 7067 rows of dialogues to perform analysis further.

**Data Analysis:** As a part of exploratory data analysis, assessments like sentiment analysis and word cloud computations are performed. Figure 4 represents the dominance of negative polarity texts in the dataset compared to positive polarity. The word cloud for depression, anxiety, OCD, and PTSD sub-domains are given which gives an overview of the most frequently used words in the text.

Often a lexical source is used for the detection

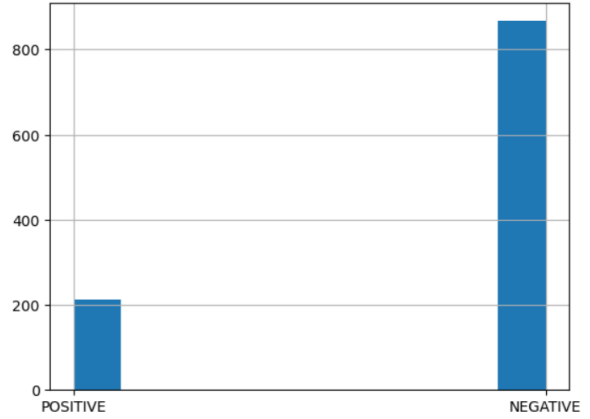


Figure 4: Sentiment polarity from Anxiety thread

of a mental health condition in a user. A word cloud gives an overview of a few words that can be helpful in the detection or classification task for NLP-based Mental Health. It can be observed that the word cloud of PTSD does not reveal the presence of a typical mental health condition. This can be attributed to the data quality for the particular sub-domain of PTSD. However, all other word clouds represent the name of the mental health condition directly. This reflects the fact that people already diagnosed with a particular mental health condition are seeking a conversational agent for further guidance and assistance. This is a typical situation in Western countries like the USA and Europe, as the data has its origin in the West. This context may not be prevalent in India.

## 4 Methodology

The idea adopted in this work is to take a pre-trained model and evaluate the generated response. The models used are Llama2-7b-chat-hf and Mistral-7b.

For the limited computing power, we have used LoRa (Low-Rank Adaptation) and a quantized fine-tuning approach. In Low-Rank adaptation the pre-trained weights are frozen, while the weight matrix in the inner layers is converted to low-rank matrices to trace the weight change. The quantized version brings down the memory format of the weights for a faster inferencing process.

## 5 Experiments

Two models, Llama2-7b-chat-hf and Mistral-7b are employed for fine-tuning with MotiVate dataset. The settings with - epoch - 1, learning rate -  $2e - 4$ , and padding-right are





Table 2: Comparison between generated responses by 2 models

Seeker Utterance	Generated Response by Llama2	Generated Response by Mistral
I know i have no control, this could have happened to anyone. and when the crisis came i did everything i could. but... i nearly failed, i almost did not make it. and everytime i put my little girl to bed i wonder if this will be the last time i see her. it is just the stupid ptsd, anxiety and depression are having a field day	Thank you for sharing your story with me. It takes a lot of courage to open up about your struggles, and I want you to know that you are not alone. Many people experience mental health issues, including PTSD, anxiety, and depression, and it's important to remember that you are not to blame for what happened. It's understandable that you are feeling scared and worried about your daughter's safety, and it's	I'm really sorry to hear that you're going through a difficult time right now. It takes a great deal of courage to reach out for support, and I want you to know that you're not alone. I can't imagine how painful it must be to experience intrusive thoughts and feelings, especially when it comes to the safety and well-being of your child. It's important to acknowledge that what you've been through is not your fault, and that having
In the fall of 2010, i lost a job and my depression became unmanageable. the level of distress i go in and out of now is beyond anything i have ever experienced in past years. i've always had social anxiety and depression are partnered up like a tag team holding me down.	Hello there, thank you for reaching out for support. Losing a job can be a significant life change that can impact your mental health, and it sounds like you're experiencing a lot of distress right now. It's important to remember that you're not alone in this, and there are resources available to help you manage your mental health.	I'm really sorry to hear that you've been going through a difficult time lately, losing a job and experiencing unmanageable depression and social anxiety. I can only imagine how challenging and overwhelming this must be for you. First and foremost, I want to assure you that you're not alone in this. Many people experience depression and social anxiety, and there are resources and support available to help you manage these feelings. It's important to take care of yourself during this time. This might include things like getting enough sleep.
I'm wondering how people cope with every day anxiety without panic and without causing extra stress. I want to just live without having to try to control and have a plan. I have a planning personality and do this naturally. But sometimes it just gets overwhelming.	Hello there It's completely normal to feel overwhelmed by everyday anxiety, especially when you have a planning personality. Here are some strategies that may help you cope with anxiety without causing extra stress: 1. Practice mindfulness: Mindfulness is the practice of being present in the moment, without judgment. It can help you focus on what's happening right now, rather than worrying about the future or past. Try incorporating mindfulness into your daily routine by taking a few deep.	I'm glad you've reached out for support with managing everyday anxiety. It's understandable that you're looking for ways to cope without adding extra stress or relying too heavily on planning. Here are some suggestions that might help: 1. Mindfulness practices: Mindfulness meditation, deep breathing exercises, and progressive muscle relaxation are all effective ways to reduce anxiety in the moment. These practices can help you focus on the present and bring your attention away from worrying thoughts. 2. Physical activity: Regular exercise can help reduce anxiety and improve your overall mood. Aim for at least 30

Often the output by the Llama-7b model keeps repeating even though the parameter of max\_seq\_len is maintained low. Table 2 represents the response by 2 models given the same prompt.

## 7 Conclusions

The results show that with similar settings Mistral-7b model is performing better in generating responses. The models do carry the capability of understanding texts related to mental health in zero-shot settings. The LLMs used here, have come a long way with their response being inherently polite. On fine-tuning, the 2 models are picking empathetic cues essential for mental health conditions. The LLMs are a good fit for conversational systems. The study can be concluded by emphasizing the fact that fine-tuning the LLMs in the mental health domain, is generating satisfactory responses, given that sufficiently high-quality data are fed to the models.

## 8 Limitations

The study in this work is only limited to text data. In the future, we will be extending our work to incorporate multi-modal cues of images, audio, and video data. The problem of mental health is region-specific. The people in the western countries are experiencing mental health conditions for a different set of reasons and people in India are experiencing mental health conditions for a whole new set of reasons. A comparative study of mental health conditions in different regions will be explored in the future. Moreover, the research work is skewed toward the English language, and other languages will be taken into consideration moving forward.

## Ethics Considerations

Research work in mental health is a critical case, it is important to follow certain standardizations and guidelines. The dataset does not carry any information about the user's profile. Any kind of personal information is not disclosed during preprocessing, training, and inferencing. This research work does not make any therapy recommendations or diagnostic claims. All the copyrights of the data belong to Psychcentral. While working with computational models for mental health, we acknowledge that the responses trying to aid can have the opposite effect. Thus, risk mitigation steps are appropriate in this context. We emphasize the point that the system does not intend to make any clinical diagnosis or treatment of the disorder. The model is made to focus on the linguistic and semantic cues for learning a generation policy.

## References

- Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, 4:463–476.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health*, 4(2):e7785.

- Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li, Yi-han Sheu, Peilin Zhou, Lauren V Moran, Sophia Ananiadou, and Andrew Beam. 2024. Large language models in mental health care: a scoping review. *arXiv preprint arXiv:2401.02984*.
- Mustafa Jahanara. 2017. Optimism, hope and mental health: Optimism, hope, psychological well-being and psychological distress among students, university of pune, india. *International Journal of Psychological and Behavioral Sciences*, 11(8):452–455.
- Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a machine tend to teenagers’ emotional needs? a study with conversational agents. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*, pages 1–6.
- Allison Lahnala, Yuntian Zhao, Charles Welch, Jonathan K Kummerfeld, Lawrence An, Kenneth Resnicow, Rada Mihalcea, and Verónica Pérez-Rosas. 2021. Exploring self-identified counseling expertise in online support forums. *arXiv preprint arXiv:2106.12976*.
- Tin Lai, Yukun Shi, Zicong Du, Jiajie Wu, Ken Fu, Yichao Dou, and Ziqi Wang. 2023. Psy-llm: Scaling up global mental health psychological services with ai-based large language models. *arXiv preprint arXiv:2307.11991*.
- Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It’s only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100.
- Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2023. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *AMIA Annual Symposium Proceedings*, volume 2023, page 1105. American Medical Informatics Association.
- Kshitij Mishra, Priyanshu Priya, Manisha Burja, and Asif Ekbal. 2023. e-therapist: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13952–13967.
- Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. 2018. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of medical Internet research*, 20(6):e10148.
- R OpenAI. 2023. Gpt-4 technical report. arxiv 2303.08774. *View in Article*, 2(5).
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.
- Tulika Saha, Vaibhav Gakhreja, Anindya Sundar Das, Souhitya Chakraborty, and Sriparna Saha. 2022a. Towards motivational and empathetic response generation in online mental health support. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 2650–2656.
- Tulika Saha, Saichethan Reddy, Anindya Das, Sriparna Saha, and Pushpak Bhattacharyya. 2022b. A shoulder to cry on: Towards a motivational virtual assistant for assuaging mental agony. In *Proceedings of the 2022 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 2436–2449.
- Tulika Saha, Saichethan Miriyala Reddy, Sriparna Saha, and Pushpak Bhattacharyya. 2022c. Mental health disorder identification from motivational conversations. *IEEE Transactions on Computational Social Systems*.