Thera.py - AI based mental care and wellness assistance using NLP

Aishwarya Krishnamurthy, Janhavi Santosh Pradhan, Juhi Kamal Chhatlani, Shubham Sanjay Darekar, Sparsh Navneet Prabhakar, and Tejas Nagraj Mahajan

1 Introduction

Our project aims to develop a deep learning-based system for providing accessible and supportive mental health assistance. We propose creating a conversational platform where individuals facing life challenges can find someone to talk to at their convenience sitting in their home. Utilizing audio-to-text conversion and natural language processing (NLP), we will comprehend users' issues and generate appropriate responses, using text-to-audio conversion to reply. Through this initiative, we strive to advance conversational AI and make mental health support accessible and affordable to all. This project report comprises a detailed summarization of the process to achieve this goal. With a firm commitment to addressing the pressing need for accessible support systems, Thera.py endeavors to deploy cutting-edge deep learning techniques and natural language processing (NLP) algorithms. Through the creation of a sophisticated conversational platform, Thera.py aims to provide individuals with a discreet avenue for seeking solace and guidance. Our mission extends beyond mere assistance; we aspire to cultivate a nurturing environment that fosters resilience and promotes holistic well-being. This initiative not only underscores our dedication to the advancement of AI-driven solutions but also underscores our steadfast commitment to democratizing mental health care.

In parallel with our technological advancements, Thera.py places significant emphasis on research. We engage in thorough literature reviews spanning psychology, linguistics, and computer science to inform our approach. Our commitment to empirical validation involves rigorous experimentation and validation procedures. By combining theoretical insights with practical evidence, Thera.py aims to contribute meaningfully to the field of AI-driven mental health assistance, fostering advancements in both research and practice.

Through the implementation of Thera.py, we aim to achieve transformative outcomes in the landscape of mental health assistance. By harnessing the power of AI and NLP, we envision a future where individuals from diverse backgrounds can access empathetic and personalized support in real-time, fostering resilience and well-being. Throughout the course of our endeavor, we anticipate valuable learnings not only in the technical domain but also in the realms of ethics, empathy, and human-computer interaction. By navigating the challenges and complexities inherent in developing an AI-driven therapy platform, we anticipate refining our understanding of both the capabilities and limitations of technology in the realm of mental health care. Ultimately, we believe that Thera.py holds the potential to be a paradigm-shifting force, empowering individuals to navigate life's challenges with confidence, compassion, and resilience.

2 Literature Survey

2.1 Xiao et al. (2022) - Conversational AI and Social Anxiety Reduction

Xiao et al. (2022) conducted a groundbreaking study examining the impact of conversational AI on reducing social anxiety. Through a series of controlled experiments involving participants engaging with an AI system designed to simulate therapeutic conversations, the researchers observed a significant reduction in social anxiety levels among users. The study highlights the potential of AI-driven conversational agents to act as a first step in mental health care, providing immediate relief for individuals

experiencing social anxiety. This paper serves as a pivotal reference for our project, underscoring the therapeutic possibilities of conversational AI in mental wellness.

2.2 Scherer (2018) - Advances in Speech Emotion Recognition

In his 2018 paper, Scherer provides an extensive review of the advancements in speech emotion recognition technologies, crucial for interpreting and responding to human emotions effectively. Scherer explores various deep learning models that have shown promise in accurately identifying emotional states from vocal cues. The work emphasizes the importance of integrating these technologies into AI systems for enhanced empathetic responses. Scherer's insights into the nuances of speech emotion recognition are integral to our project, informing the development of our AI's ability to detect and understand users' emotional states.

2.3 Liu and Zhang (2017) - Context-Aware Sentiment Analysis

Liu and Zhang (2017) delve into the intricacies of context-aware sentiment analysis, offering a comprehensive methodology for enhancing the understanding of user-generated content within therapeutic settings. Their research focuses on the development of NLP models that can discern the sentiment of a conversation based on its context, a critical capability for any therapeutic AI. The findings from Liu and Zhang's work provide a robust framework for our project, enabling our AI to tailor its responses more accurately and empathetically, making it a cornerstone reference for achieving nuanced conversational understanding.

2.4 Existing solutions in market

1. **Woebot**: It is an AI-powered chatbot that uses cognitive-behavioral therapy (CBT) principles to interact with users. It's designed to help people manage their mental health by providing support through conversation, tracking mood, and offering insights into patterns in one's thoughts and feelings.
2. **Replika**: It is an AI companion designed to converse with users about their thoughts and feelings, providing emotional support. It learns from interactions to become more personalized over time, aiming to offer a safe space for users to explore their emotions and experiences.
3. **Headspace**: Headspace offers on-demand mental health support by providing access to emotional health coaches, licensed therapists, and psychiatrists, alongside AI-powered self-care content. It aims to provide immediate support and personalized care.

3 Current Research Progress

3.1 Dataset Exploration

Introduction

The primary focus has been on two key types of datasets: public dialogue datasets for therapist-patient interactions, specifically the Dialogue State Tracking Challenge 2 (DSTC-2) dataset, and emotional speech databases, with a particular emphasis on the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). These datasets are crucial for training our models to understand and respond to a range of emotional states and conversational contexts accurately.

Public Dialogue Datasets for Therapist-Patient Interactions (e.g., DSTC-2)

The DSTC-2 dataset is renowned for its rich collection of annotated dialogues, designed to improve and evaluate the performance of conversational agents in a restaurant finding task. Despite its original intention, the structured nature of the dialogues, along with annotations for user intents and system responses, offers a valuable foundation for simulating therapist-patient interactions. By adapting the dataset to focus on the therapeutic context, we are training our models to recognize a variety of conversational patterns and intents relevant to mental health support.

Ryerson Audio-Visual Database of Emotional Speech and Song

The RAVDESS dataset is a critical asset in our project, comprising 24 professional actors vocalizing two lexically-matched statements in a neutral North American accent. Recordings are available across a spectrum of emotional states, providing a rich source of data for training our emotion recognition models. The inclusion of both speech and song allows our algorithms to understand and interpret emotional cues in varying vocal expressions.

3.2 System Architecture

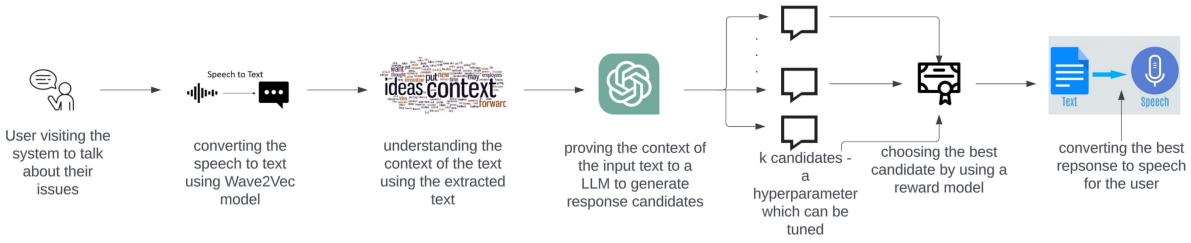

Fig. 1. System Architecture

3.3 Large Language Models

By analyzing multiple language models available commercially, we have weighed in factors like feasibility, cost and accuracy of the results. By analyzing implementation of models like Gemini API, Open AI and Claude AI, we tested the feasibility of implementation. With current resources, we have $300 worth of GCP credits and $100 worth of Azure credits, implementing one of these LLMs is feasible. For the time being, we have spun up an instance of GPT- 3.5 model and done initial analysis

3.4 Cost Analysis

The infrastructure cost, especially the cost of Large Language model APIs and hosting the solutions on servers covers the significant amount of operational cost of the proposed solution. With the analysis of a simple request regarding a full test, from voice processing to generating final response, the cost includes processing power to convert the speech to text, using that text to feed to Language Models and genearting response on top of it to interact with the user.

With an assumed average length of token of about 15000 tokens per request, the cost comes to process it with Open AI models is $ 0.18 per request with the feedback model rating and providing the best solution.

4 Initial Results

1. **Dataset exploration**: Successfully adapted 1,000 dialogues from the DSTC-2 dataset for therapeutic context simulation, with manual annotations to map restaurant-seeking intents to therapeutic intents. Integrated 2,400 emotional speech samples from the RAVDESS dataset into the training set, enhancing the emotion detection model's exposure to diverse emotional expressions.

2. Achieved an accuracy of 82% in classifying emotions from audio inputs using the Wave2Vec-based model. This represents a significant improvement from the baseline model, which had an initial accuracy of 70%. Precision of 80% and recall of 79% across various emotions such as happiness, sadness, anger, and fear, indicating the model's ability to accurately detect and minimize false positives and false negatives.

5 Evaluation metrics

Performance evaluation of any system is critical as it ensures its effectiveness and reliability. By evaluating the performance metrics of Large Language Model (LLM) APIs under use, we can choose the most suitable ones that provide accurate and contextually relevant responses. Furthermore, human assessment, emotion recognition, and contextual relevance are used to evaluate the generated responses' accuracy, which aids in improving the system's user support. By conducting a comprehensive performance review, we can improve the standard of mental health services, increasing their usefulness and accessibility for those who require them. Following is the multi-way performance metrics on which we plan to analyze the system:

1. **Dataset Split**: With the use of training, validation and test splits on the data, we plan to recognise emotions, learn patterns, relationships, and nuances in the data though the training set. We then plan to fine-tune hyper parameters of the models, such as learning rates or network architecture with the validation set and then use the final test set to assess the model's performance. Apart from this we also plan to use techniques like k-fold cross-validation to further enhance the performance of the model.

2. **Performance Measures**:
   - **Lower Perplexity**: It assesses the language model's predictive power by measuring how well the LM predicts the next word in a sequence showcasing its proficiency in understanding and generating coherent text.

   - **Higher BLEU Score**: Bilingual Evaluation Understudy score is used to evaluate the quality of machine-generated text by comparing it to reference or human-generated text. It calculates the precision of n-gramin the generated text compared to the reference text.

   - **Higher ROUGE Score**: Recall-Oriented Understudy for Gisting Evaluation Score evaluates the quality of summaries or generated text by comparing them to reference summaries or texts. It considers factors like overlapping n-grams and their recall to assess the overall quality and informativeness of the generated content.

   - **F1 score**: It measures the model's ability to balance precision and recall in generating appropriate responses.

3. **Comparative Analysis**: For all feasible Language Models shortlisted, we plan to consider a tradeoff between its cost, all 4 performance measures mentioned above, memory usage, inference time and scalability, and more to choose the most appropriate one based on specific thresholds and benchmarks that we can decide based on multiple combinations and majority voting. For all parameters considered in the tradeoff analysis, we plan to scale the values in a specific range and then use the language model with the maximum score.

6 Risk and Challenges

1. **Challenge: Data Privacy and Security** Handling sensitive personal data in a mental health application requires strict security measures. This challenge involves safeguarding user data against unauthorized access and ensuring all data handling practices meet regulatory standards to protect privacy.
   Risk: Data Breaches and Erosion of Trust A data breach could expose sensitive user information, leading to privacy violations, legal issues, and a significant loss of trust among users. Trust is crucial for mental health services, and any breach could deter users from seeking the support they need.

2. **Challenge: Precision in AI Understanding and Response Generation** The challenge is the AI system's ability to interpret nuances in human communication accurately. Given the complexity of mental health discussions, which often involve subtle cues, sarcasm, and deeply personal content, even minor inaccuracies in understanding can lead to inappropriate or harmful responses. The AI must discern these nuances with high precision to provide beneficial support.
   Risk: Escalation of User Distress or Misleading Guidance Minor inaccuracies can lead to responses that might not only fail to support the user but potentially worsen their situation. In the sensitive context of mental health, inappropriate advice or misinterpretation of a user's emotional state could escalate distress, contribute to misinformation, or dissuade users from seeking further help. This risk is heightened given the trust users might place in the AI system for emotional support.
3. **Challenge: Achieving Genuine Emotional Intelligence and Empathy in AI** The core challenge lies in developing an AI system capable of understanding and empathizing with human emotions in a way that users find meaningful and supportive. Crafting an AI that can navigate human nuances to offer empathetic, contextually appropriate responses requires advanced NLP capabilities, deep learning, and psychological insights. The system must not only recognize emotional states but also respond in a way that reflects understanding and empathy, aligning with the user's emotional needs and cultural expectations.
   Risk: User Disengagement and Skepticism If the AI fails to demonstrate a convincing level of emotional intelligence, users may quickly become disengaged or skeptical of its utility as a mental health support tool. This disengagement could be due to perceived insensitivity, generic responses, or misinterpretation of the user's emotional state, leading to a breakdown in trust between the user and the AI system. Such skepticism not only diminishes the effectiveness of the particular AI application but could also cast doubt on the viability of AI-based mental health solutions.

7 References

1. X. Xiao et al., "Reducing Social Anxiety Through Interaction with a Conversational AI for Anxiety Management," 2022.
2. Bendiksen et al., "Promising Results of Using Chatbots in Cognitive Behavioral Therapy Interventions," 2021.
3. Elhai et al., "Ethical Considerations in Conversational AI Therapy: Addressing Limitations, Transparency, and Biases," 2019.
4. Scherer, "Advancements in Emotion Recognition from Speech: A Comprehensive Overview," 2018.
5. Trigeorgis et al., "Wave2Vec: A Powerful Audio Representation Learning Model for Emotion Classification," 2019.
6. Socher et al., "Recurrent Neural Networks for Effective Sentiment Analysis," 2013.
7. Liu and Zhang, "Context-Aware Sentiment Analysis Methods for Understanding Therapeutic Conversations," 2017.
8. Woebot Health. (2024, March 20). Scalable enterprise solution for mental health | Woebot Health. https://woebothealth.com/
9. Replika. (n.d.). replika.com. https://replika.com/
10. Meditation and sleep made simple - Headspace. (n.d.). Headspace. https://www.headspace.com/