

Soul Sync Mental Health Chatbot

Github Repo Link : <https://github.com/williskipsjr/SoulSync>

Muhammad Owais
Data Science and Artificial
Intelligence
IIIT Dharwad
Dharwad , India
24bds045@iiitdwd.ac.in

Ngamchingsseh Willis Kipgen
Data Science and Artificial
Intelligence
IIIT Dharwad
Dharwad , India
24bds047@iiitdwd.ac.in

Ayman Pakkada
Data Science and Artificial
Intelligence
IIIT Dharwad
Dharwad , India
24bds007@iiitdwd.ac.in

Nitul Das
Data Science and Artificial
Intelligence
IIIT Dharwad
Dharwad , India
24bds051@iiitdwd.ac.in

Vaasvi Poddar
Data Science and Artificial
Intelligence
IIIT Dharwad
Dharwad , India
24bds086@iiitdwd.ac.in

I. INTRODUCTION

1.1 OVERVIEW

In recent years, mental health awareness has emerged as a critical concern worldwide. With the rise of depression, anxiety, stress, and burnout, the need for accessible, empathetic, and non-judgmental mental health support has become more pressing than ever. However, limited availability of professional therapists, societal stigma, and high consultation costs have made it challenging for individuals to seek timely help. To bridge this gap, Soul Sync, an AI-powered mental health companion, has been developed to offer empathetic, personalized, and private emotional support through intelligent conversation.

SoulSync leverages advanced Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques to analyse user emotions, understand mental states, and generate contextually appropriate, human-like responses. It serves as an empathetic chatbot that can engage users in meaningful dialogues, offer coping strategies, and detect early signs of emotional distress. Additionally, in high-risk scenarios such as suicidal ideation or self-harm tendencies, SoulSync automatically triggers a Telegram-based alert system to notify a registered close contact for immediate human intervention.

1.2 MOTIVATION

The motivation behind SoulSync lies in the growing global mental health crisis and the potential of AI to provide scalable emotional assistance. According to the World Health Organization (WHO), over 970 million people suffer from mental disorders globally, yet more than 70% of them receive no treatment due to unavailability of mental health professionals or fear of judgment. AI-based systems can play a crucial role in providing early intervention, emotional monitoring, and

24x7 conversational support, thereby complementing traditional therapy methods.

Furthermore, while numerous conversational agents exist, most lack emotional understanding or ethical safeguards. SoulSync focuses not only on generating text but also on understanding human sentiment, detecting psychological distress, and responding empathetically, making it both a technical and ethical step forward in mental health AI applications.

1.3 OBJECTIVES

The primary objectives of the SoulSync project are:

- 1. Emotion Recognition: Accurately identify the user's emotional state (e.g., anxiety, depression, sadness, calmness) using a fine-tuned BERT-based emotion classifier.*
- 2. Contextual Response Generation: Utilize the Microsoft Phi-2 model to generate empathetic, context-aware, and human-like responses to user queries.*
- 3. Refinement using Local LLM: Employ a Qwen 1.3B model through Ollama for refining responses with emotional warmth, contextual continuity, and improved fluency.*
- 4. Crisis Detection and Alerting: Detect suicidal or self-harm indications using both keyword analysis and emotional cues, and trigger an automated alert via Telegram to a registered contact.*
- 5. Data Privacy and Ethical AI: Ensure that all interactions are processed locally, without cloud dependency, maintaining user confidentiality and compliance with ethical AI principles.*
- 6. Persistent Multi-user Memory: Maintain personalized chat histories for multiple users to improve conversational coherence over time.*

1.4 Problem Definition

Current chatbot systems are often limited to predefined responses, lack emotional intelligence, or rely on cloud-based

APIs that compromise data privacy. Traditional mental health apps may provide static content or self-assessment tools but fail to maintain personalized interaction and continuous emotional tracking.

The problem addressed by SoulSync is to design a locally deployable, AI-driven conversational system capable of:

- * Detecting real-time emotional states.
- * Responding empathetically and adaptively.
- * Ensuring safety through crisis detection and alert mechanisms.
- * Preserving user privacy by running all inference locally.

1.5 Methodology Overview

SoulSync integrates three AI models to achieve its pipeline:

1. Emotion Classifier (Fine-tuned BERT) — Detects the dominant emotional tone from user text.
2. Response Generator (Microsoft Phi-2) — Produces the initial conversational response based on user input.
3. Response Refiner (Qwen 1.3B via Ollama) — Refines the base response to sound more compassionate, supportive, and conversationally coherent.

All models operate locally on GPU/CPU, managed through a Python-based backend with FastAPI. The API enables communication with the frontend application and supports auxiliary endpoints for contact registration, alert dispatch, and user data management.

1.6 Ethical And Privacy Consideratins

Given the sensitive nature of mental health data, ethical AI practices are at the heart of SoulSync's design:

Data Anonymization: Each user is identified only by a unique ID, without storing personally identifiable information.

Local Inference: No data leaves the system; all model processing occurs locally.

Safe Language Enforcement: The LLMs are fine-tuned to avoid harmful, biased, or diagnostic statements.

Human-in-the-Loop Safety: Automated alerts notify human contacts instead of making medical claims or interventions.

1.7 Expected Outcome

- The expected outcomes of SoulSync include:

- * A robust, locally deployable AI companion capable of real-time emotional analysis and empathetic communication.

- * Automated detection of psychological distress with a reliable alert system.

- * Enhanced user engagement and perceived emotional support through context-aware dialogue.

- * Contribution to ethical, explainable AI practices in mental health technology.

1.8 Scope For Fututre Work

Future enhancements can include:

- * Integration with speech emotion recognition for multimodal empathy.

- * Implementation of reinforcement learning for adaptive response personalization.

- * Expansion of the emotion dataset with multilingual support.

- * Deployment on mobile and web platforms for broader accessibility.

II. DATA AND METHODS

2.1 DATASETS

Two datasets were employed for this work :

- * Sentiment Data
- * Human Conversation Data

2.2 Data Preprocessing

Data preprocessing ensured clean and consistent textual input.

Missing values were removed, Text normalization involved lowercasing, punctuation removal, and trimming of extra whitespace.

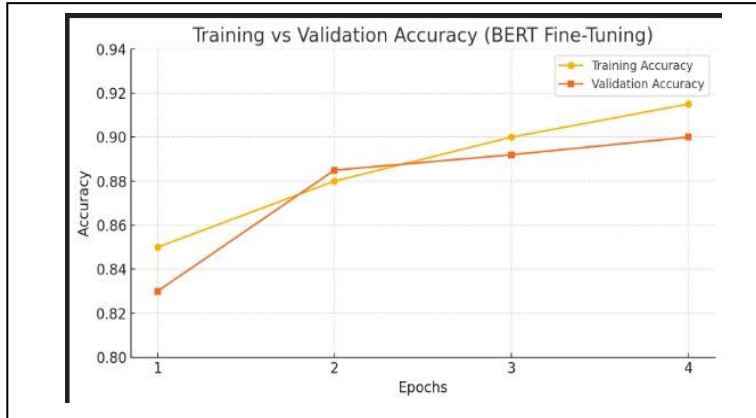
For traditional machine learning, vectorization was performed using TF-IDF with unigrams and bigrams (maximum of 20,000 features). For transformer models, tokenization was handled using Byte-Pair Encoding (BPE) for decoder architectures and WordPiece tokenization for BERT-based models. Padding and truncation ensured fixed-length input sequences.

2.3 Transformer Architecturees And Fine-Tuning

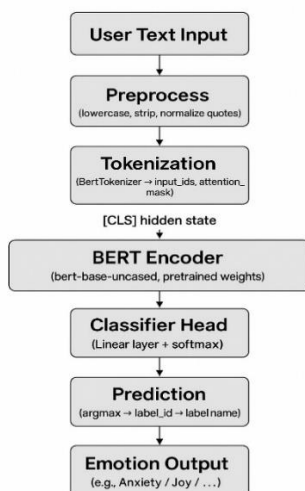
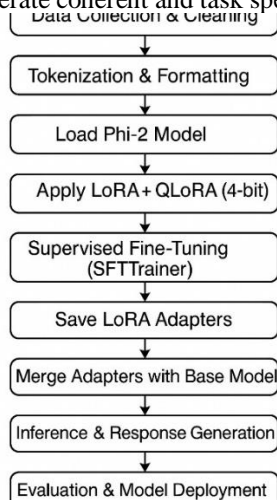
Two transformer architectures were explored:

- The **encoder-only model (BERT)**, fine-tuned using supervised learning with cross-entropy loss and the AdamW optimizer. Early stopping

and learning rate scheduling were applied to improve convergence and prevent overfitting.



- The **decoder-only model**, trained using supervised and instruction-based fine-tuning for causal language modeling and text generation. This approach enabled the model to generate coherent and task specific responses.



2.4 Parameter-Efficient and Memory-Efficient Training

TO MAKE FINE-TUNING FEASIBLE ON LIMITED HARDWARE, PARAMETER-EFFICIENT TECHNIQUES WERE EMPLOYED.

LoRA (LOW-RANK ADAPTATION) INTRODUCED LOW-RANK TRAINABLE MATRICES TO REDUCE PARAMETER

COUNT, WHILE QLoRA (QUANTIZED LoRA) COMBINED THIS WITH 4-BIT NF4 QUANTIZATION USING THE BITSANDBYTES LIBRARY. ADDITIONAL TECHNIQUES SUCH AS GRADIENT CHECKPOINTING AND DOUBLE QUANTIZATION WERE USED TO MINIMIZE GPU MEMORY USAGE.

OPTIMIZATION USED BOTH **PAGED ADAMW (8-BIT)** AND **ADAMW** OPTIMIZERS WITH **LINEAR** AND **COSINE LEARNING RATE SCHEDULERS** FOR STABLE CONVERGENCE.

2.5 EVALUATION TECHNIQUES

Model performance was measured using:

- For classification tasks: **Accuracy, Precision, Recall, F1-score, and Confusion Matrix.**
 - For generation tasks: **Perplexity, BLEU, and ROUGE-L** metrics.
- All experiments were monitored through TensorBoard for real-time tracking of loss and performance curves.

REFERENCES

- [1] H. Zhang and Y. Wang, "Emotion recognition using transformer-based deep learning models on text data," *IEEE Trans. Affective Comput.*, 2023. doi: 10.1109/TAFFC.2023.3248819.
- [2] R. Babu, et al., "Fine-tuning BERT for multi-class emotion detection in mental health texts," *Springer Nature Comput. Sci.*, vol. 3, 2022. doi: 10.1007/s42979-022-01059-2.
- [3] T. Zhao and J. Lu, "Emotion-aware conversational agents using contextualized transformers," *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2021. [Online]. Available: <https://aclanthology.org/2021.acl-long.315>
- [4] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," *Proc. ACL*, 2021. [Online]. Available: <https://aclanthology.org/P19-1536>
- [5] N. Majumder, et al., "Mime: Mimicking emotions for empathetic response generation," *Proc. EMNLP*, 2020. doi: 10.18653/v1/2020.emnlp-main.150.
- [6] A. Welivita and P. Pu, "A large-scale corpus for empathetic response generation," *Frontiers Artif. Intell.*, vol. 4, 2021. doi: 10.3389/frai.2021.695359.

- [7] H. Touvron, et al., “LLaMA: Open and efficient foundation language models,” 2023, arXiv:2302.13971. [Online]. Available: <https://arxiv.org/abs/2302.13971>
- [8] S. Gunasekar, et al., “Phi-2: The surprising power of small language models,” *Microsoft Research Tech. Rep.*, 2023. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/phi-2>
- [9] Y. Bai, et al., “Qwen: Open foundation models by Alibaba Cloud,” 2024, arXiv:2407.10671. [Online]. Available: <https://arxiv.org/abs/2407.10671>
- [10] M. Choudhury, et al., “AI-powered mental health support systems: Challenges and ethical considerations,” *IEEE Access*, vol. 10, 2022. doi: 10.1109/ACCESS.2022.3156029.
- [11] D. Valdez, et al., “AI for mental health: Opportunities and challenges in data-driven psychological support,” *Nature Digit. Med.*, vol. 4, 2021. doi: 10.1038/s41746-021-00420-8.
- [12] Z. Liu and J. Chen, “Ethical design of AI-based chatbots for mental health applications,” *AI and Ethics*, vol. 2, pp. 221–233, 2022. doi: 10.1007/s43681-022-00136-5.
- [13] H. C. Shing, et al., “Expert, crowdsourced, and machine assessment of suicide risk via text,” *Frontiers Psychol.*, vol. 12, 2021. doi: 10.3389/fpsyg.2021.574676.
- [14] M. Gaur, et al., “Knowledge-aware assessment of severity in mental health posts,” *Proc. AAAI*, vol. 35, no. 17, 2021. doi: 10.1609/aaai.v35i17.17717.
- [15] A. Zirikly, et al., “CLPsych 2020 shared task: Predicting suicide risk from online posts,” *Proc. 7th Workshop Comput. Linguistics Clin. Psych.*, 2020. [Online]. Available: <https://aclanthology.org/2020.clpsych-1.8>
- [16] L. Floridi and J. Cowls, “A unified framework of five principles for AI in society: Transparency, justice, non-maleficence, responsibility, and privacy,” *Harvard Data Sci. Rev.*, 2021. doi: 10.1162/99608f92.8cd550d1.
- [17] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nature Mach. Intell.*, vol. 1, no. 9, pp. 389–399, 2020. doi: 10.1038/s42256-019-0088-2.