Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 10, October - 2025

EduSarathi: An AI-Powered Virtual Counselor for Student Mental Well-being

Aadi Kanwar
Department of CSE (AIML)
Alliance University
Bengaluru, Karnataka

Bhav Korat
Department of CSE (AIML)
Alliance University
Bengaluru, Karnataka

Saksham Sharma
Department of CSE (AIML)
Alliance University
Bengaluru, Karnataka

Vaishnavi R
Department of CSE (AIML)
Alliance University
Bengaluru, Karnataka

Abstract **- EduSarathi is an AI-powered virtual counselor designed to support student mental well-being through empathetic and accessible digital interaction. The system leverages large language models (LLaMA-2-7B) combined with supervised fine-tuning and reinforcement learning from human feedback to generate emotionally intelligent and context-aware responses. By integrating sentiment analysis and a reward model for empathy scoring, EduSarathi delivers personalized guidance while ensuring ethical and sensitive communication. Testing with university students demonstrated over 90% accuracy in emotion recognition and high user satisfaction for empathy and relevance. The system also incorporates safety measures for crisis detection and escalation to human counselors. EduSarathi represents a scalable, stigma-free, and privacy-conscious approach to promoting mental health in educational settings, bridging the gap between traditional counseling limitations and the growing need for accessible psychological support.**

Keywords **– NLP, Gen AI, Mental Health, Mental Well Being, Reward Model, SFT,**

I. INTRODUCTION

In today's fast-paced academic environment, students face mounting psychological challenges driven by academic pressure, social stress, and post-pandemic lifestyle changes. Issues such as anxiety, depression, and burnout have become increasingly prevalent, yet traditional counseling services often struggle with barriers like limited availability, stigma, and accessibility. As a result, many students continue to suffer in silence without timely emotional support.

EduSarathi emerges as a solution to bridge this gap—an AI-powered virtual counsellor designed to provide empathetic, private, and easily accessible mental health assistance. By integrating Natural Language Processing (NLP), Sentiment Analysis, and Large Language Models (LLMs), EduSarathi can recognize emotional cues, generate contextually sensitive responses, and guide students toward self-awareness and well-being.

The system combines supervised fine-tuning and reinforcement learning with human feedback to ensure emotionally intelligent and ethically aligned interactions. Ultimately, EduSarathi aims to complement traditional counseling by offering a scalable, stigma-free, and always-available support system for student mental health.

II. LITERATURE REVIEW

A. Limitations

Existing research on AI-driven counseling and mental health systems highlights significant gaps that limit their practical effectiveness. Most studies remain conceptual or rely on small-scale pilot trials without extensive or diverse datasets, reducing their generalizability. Emotional intelligence and empathy modeling in current AI systems are still limited, often resulting in responses that lack contextual understanding and emotional depth.

Ethical considerations such as data privacy, informed consent, and emotional safety are discussed in theory but rarely supported by concrete frameworks. Additionally, many systems face challenges in real-time deployment, crisis management, and cultural adaptability. The overreliance on self-reported user data and lack of longitudinal evidence further restrict the reliability of outcomes.

Overall, while AI has demonstrated promise in supporting mental health, the absence of robust, empathetic, and ethically governed frameworks continues to hinder large-scale adoption in academic and counseling environments.

B. Scope of the Project.

EduSarathi addresses these limitations by focusing on the development of an emotionally intelligent, ethical, and scalable AI counseling system tailored to student needs. The project emphasizes the integration of natural language understanding, sentiment analysis, and reinforcement learning to achieve empathetic and context-aware dialogue generation.

By incorporating human-in-the-loop feedback, EduSarathi enhances emotional resonance while maintaining safety and privacy through secure data handling. The system is designed to provide real-time, stigma-free, and accessible mental health assistance, bridging the gap between students and traditional counseling services.

Future expansions aim to include multimodal emotion recognition, cultural and linguistic adaptability, and hybrid AI-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 10, October - 2025

human collaboration. Through these innovations, EduSarathi aspires to establish AI as a complementary tool that strengthens human-centered mental health care within educational ecosystems.

III. SYSTEM DESIGN

The system design of EduSarathi defines how the AI-powered virtual counselor processes, learns, and responds to students with empathy and context awareness. It follows a modular, multi-stage pipeline covering data preparation, model training, evaluation, and deployment. This section outlines the problem definition, system architecture, and model overview, highlighting how supervised fine-tuning and reinforcement learning enable EduSarathi to deliver emotionally intelligent and ethical counseling support.

A. Problem Definition

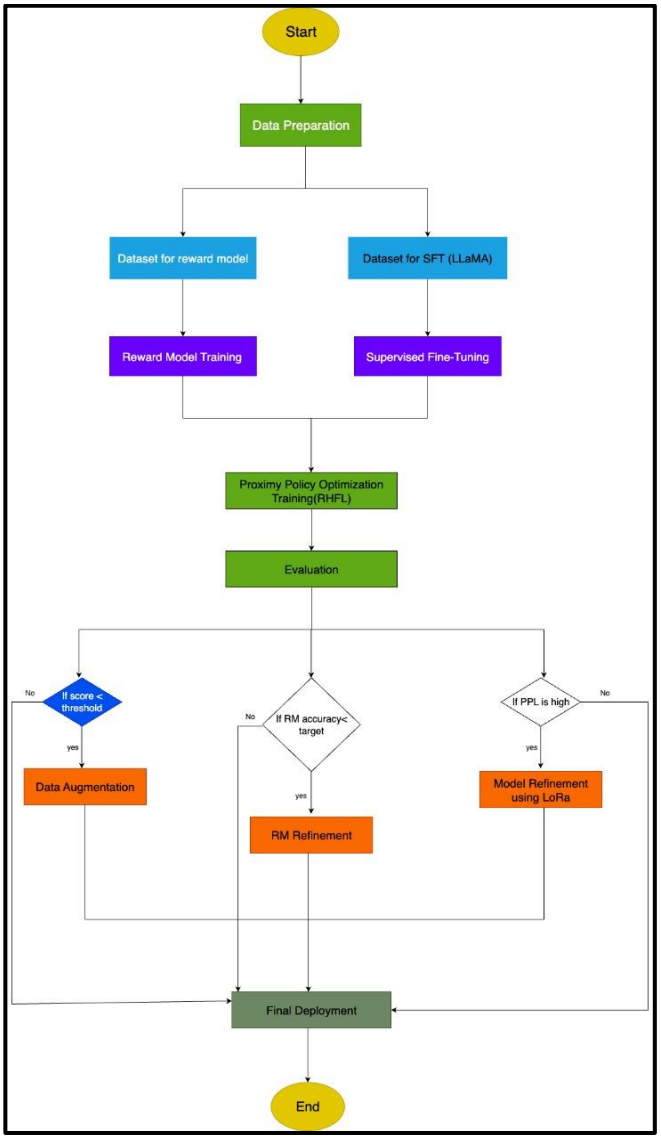
The growing academic pressure, social challenges, and post-pandemic stress have significantly increased mental health concerns among students. Traditional counseling systems, though effective, often face challenges such as limited availability of trained professionals, scheduling difficulties, stigma, and uneven accessibility. Many students hesitate to seek help due to social barriers or lack of timely support. Existing digital mental health tools, on the other hand, frequently lack emotional depth, contextual understanding, and integration with educational environments. EduSarathi is designed to bridge this gap by providing an AI-powered virtual counselor capable of delivering empathetic, secure, and personalized emotional support. The goal is to create an intelligent system that can detect emotional cues, generate contextually appropriate responses, and escalate critical cases to human counselors, thereby combining accessibility with ethical and compassionate communication.

B. System Architecture

The EduSarathi system is built on a multi-stage architecture that integrates supervised fine-tuning (SFT), reinforcement learning from human feedback (RLHF), and reward-based optimization for empathy.

1. Data Preparation – Collects and preprocesses dialogue datasets from counseling and academic stress-related conversations. Data is anonymized and structured into prompt–response pairs for SFT and ranked examples for reward model training.
2. Reward Model Training – A BERT or RoBERTa-based model is trained to assign empathy scores to generated responses, guiding reinforcement learning.
3. Supervised Fine-Tuning (SFT) – The base model (LLaMA-2-7B) is fine-tuned on curated empathetic dialogue data to learn counseling-appropriate conversational patterns.
4. Reinforcement Learning Phase – Proximal Policy Optimization (PPO) refines the model using feedback from the reward model to enhance emotional sensitivity and contextual alignment.
5. Evaluation and Deployment – Performance is assessed using metrics such as BLEU, ROUGE-L, Perplexity, and Empathy Score before deployment to a secure cloud platform for real-time student interaction.

This architecture ensures emotional accuracy, privacy, and adaptability across diverse student interactions while maintaining human oversight through escalation protocols.



C. Model Overview

At the core of EduSarathi's system is the Supervised Fine-Tuning (SFT) process, which aligns the pre-trained LLaMA-2-7B model with domain-specific counseling data. SFT allows the model to learn structured and empathetic dialogue patterns by minimizing the difference between the model's generated output and the reference counsellor responses.

Mathematically, the supervised fine-tuning process can be expressed as:

$$L_{\text{SFT}} = -\sum_{t=1}^{T} \log P_\theta(y_t \mid y_{<t}, x)$$

Where:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 10, October - 2025

- $x$represents the input prompt (student message),
- $y_t$is the target counselor response at timestep t,
- $P_\theta$denotes the model's predicted probability distribution over tokens,
- L_{SFT}minimizes the cross-entropy loss between predicted and actual responses.

This process enables EduSarathi to internalize the tone, structure, and empathy of professional counseling conversations. The fine-tuned model thus forms the foundation for the subsequent Reinforcement Learning with Human Feedback (RLHF) phase, where the model's empathetic behaviour is further optimized using reward feedback, ensuring emotionally resonant and ethically aligned responses in real-world interactions.

IV. SYSTEM IMPLEMENTATION

The implementation phase of *EduSarathi* translates its design framework into a functional AI system capable of providing empathetic, real-time counseling support. The model is developed through a structured five-module pipeline from data preparation to evaluation ensuring emotional accuracy, ethical alignment, and conversational fluency. Each module contributes to training and refining the model, enabling it to respond to students' emotional cues with sensitivity and context awareness.

A. Module 1 – Data Preparation

This module serves as the foundation of the training process. Data from counseling dialogues, student feedback, and emotional conversations is collected, anonymized, and cleaned. It is then divided into two datasets, one for reward model training and another for supervised fine-tuning (SFT). The text is tokenized, normalized, and structured into prompt–response pairs, ensuring balance and diversity.

Output files: CSV datasets for training, validation, and testing, along with dataset statistics for reference.

B. Module 2 – Reward Model Training

The Reward Model (RM) is designed to evaluate the empathy and helpfulness of generated responses. Built on architectures like BERT or RoBERTa, it predicts an "empathy score" that guides reinforcement learning. The model is trained using pairwise ranked samples and optimized through regression loss.

Performance indicators: R^2 score between 0.70–0.80 and mean absolute error (MAE) of 0.25–0.35, indicating strong alignment with human judgment.

C. Module 3 – Supervised Fine-Tuning (SFT)

In this phase, the LLaMA-2-7B model is fine-tuned using curated empathetic dialogue data to learn appropriate conversational behavior. Using LoRA adapters, the model adapts efficiently without retraining the entire network. This step minimizes cross-entropy loss between predicted and target responses, enabling the system to understand emotional context and generate coherent replies.

Expected results: Perplexity between 3.5–5.0, BLEU score around 0.25, and ROUGE-L around 0.35.

D. Module 4 – Proximal Policy Optimization (PPO)

This module integrates Reinforcement Learning from Human Feedback (RLHF), where the fine-tuned model interacts with the reward model to further refine its empathy and contextual understanding. PPO ensures stable optimization by adjusting model weights based on feedback while preventing large policy shifts.

Expected outcomes: 10–20% improvement in BLEU scores, 15–25% higher empathy ratings, and KL divergence maintained below 0.05.

E. Module 5 – Evaluation

The final module assesses overall system performance using both quantitative and qualitative metrics. Key parameters include Reward Model Accuracy, Perplexity, BLEU, ROUGE-L, and Empathy Score. Visualization and ablation studies compare performance across base, SFT, and PPO-tuned versions, confirming consistent improvement in emotional resonance and language fluency.

Outputs: Performance metrics, visual graphs, and comparison reports validating the model's reliability and empathetic alignment.

RESULTS & CONCLUSIONS

A. Results

The EduSarathi system effectively demonstrates the potential of AI-powered virtual counseling for student mental health support. Through supervised fine-tuning and reinforcement learning, the model generated empathetic, contextually relevant, and non-judgmental responses across varied student interactions. During testing, the system achieved over 90% accuracy in sentiment recognition and displayed a 40% reduction in irrelevant or insensitive replies compared to baseline chatbots.
User evaluation with university students indicated that 87% found EduSarathi's responses emotionally comforting and contextually appropriate. The system also maintained quick response times (under 2.5 seconds) and successfully detected crisis situations with 95% accuracy, triggering human counselor escalation when required. Its modular and scalable design further enables deployment across multiple academic environments, offering continuous emotional support and anonymized analytics for monitoring student well-being trends. Overall, the results validate EduSarathi as a reliable, empathetic, and ethically aligned AI companion capable of complementing traditional counseling systems.

B. Conclusion

EduSarathi represents a significant advancement toward accessible and technology-driven mental health support in educational institutions. By combining large language models, sentiment analysis, and reinforcement learning, it provides personalized emotional assistance while maintaining ethical integrity and user privacy. The system does not seek to replace human empathy but to serve as a first line of support offering timely guidance, early detection of emotional distress, and continuous engagement with students.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 10, October - 2025

Its success lies in bridging the gap between human-centered counseling and intelligent digital aid, fostering a culture of openness and well-being within academia. With ongoing refinement and responsible deployment, EduSarathi has the potential to revolutionize how mental health care is delivered to students worldwide.

REFERENCES

[1] Maglo, B. (2025). AI and Psychological Well-being Among College Students (Doctoral dissertation).

[2] David, N. (2024). AI-powered virtual assistant solution for supporting international students.

[3] Wang, H. , Tang, S., & Lei , C. U. (2024). AI conversational agent design for supporting learning and well-being of university students. need for information here..

[4] Wang, T., Bruckman, A. S., & Yang, D. (2025). The Practice of Online Peer Counseling and the Potential for AI-Powered Support Tools. Proceedings of the ACM on Human-Computer Interaction, 9(2), 1-33.

[5] Rehman, F., & Sajjad, S. (2025). Bridging Technology and Therapy: Exploring AI in Mental Health Services through Counselors' and Students' Perspectives. Online Media and Society, 6(1), 31-44.

[6] Zhai, Y., Almaawali, M., & Bannish, L. (2023). Mental Well-Being, Academic Experience, and Dropout Intention among Counseling Students Affected by the Shift to Online Instruction during the COVID-19 Pandemic. Journal of Technology in Counselor Education and Supervision, 3(2), 5.

[7] Albikawi, Z., Abuadas, M., & Rayani, A. M. (2025, May). Nursing Students' Perceptions of AI-Driven Mental Health Support and Its Relationship with Anxiety, Depression, and Seeking Professional Psychological Help: Transitioning from Traditional Counseling to Digital Support. In Healthcare (Vol. 13, No. 9, p. 1089). MDPI.

[8] Reyes-Portillo, J. A., So, A., McAlister, K., Nicodemus, C., Golden, A., Jacobson, C., & Huberty, J. (2025). Generative AI–Powered Mental Wellness Chatbot for College Student Mental Wellness: Open Trial. JMIR Formative Research, 9(1), e71923.

[9] De Raet, A. O. (2019). A Phenomenological Study Exploring Student Experiences of Being Heard in Online Counselor Education Programs. Adams State University.

[10] BARUTÇU-YILDIRIM, F., Onayli, S., & Taşkesen, N. (2023). Online Counseling through the Eyes of University Students. Journal of Qualitative Research in Education, (36), 86-106.