iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70773

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Serenify: An AI-Powered Virtual Therapist for Accessible Mental Health Support

Vidul Pratap Chauhan[1], Vaishali Goel[2]
Department of Information Technology Maharaja Agrasen Institute of Technology New Delhi, India

Abstract: ***Mental health issues affect millions of people around the world, but getting access to professional therapy isn't always easy. The reasons vary — therapy can be expensive, there's still a social stigma attached to it in many places, and there just aren't enough trained professionals to meet the growing demand. That's where Serenify comes in. It's a virtual therapist powered by AI, built to help make mental health support more accessible, affordable, and consistent. Using tools like natural language processing, sentiment analysis, and memory retention, Serenify offers therapy sessions that feel more personal and relevant to each user. At the heart of it is Sentinel, our sentiment analysis model, which is good at picking up on emotional cues so the responses feel more thoughtful and in tune with the user's state of mind. In this paper, we'll walk through how Serenify was built, the technologies behind it, and why it could make a real difference in the way we approach digital mental health care.***

I. INTRODUCTION

Mental health plays a huge role in our overall well-being, but for a lot of people, getting access to traditional therapy just isn't realistic. Whether it's due to cost, social stigma, or long waiting lists, many end up going without the help they need. With the rise of AI, though, there's a growing chance to rethink how we approach mental health support. That's the idea behind *Serenify* — a virtual therapist designed to offer real-time, affordable, and more personalized support to users, right when they need it.

Serenify combines AI-powered conversation models with cognitive behavioral therapy (CBT) techniques and emotion-tracking tools to create a more supportive and responsive experience. One of the key features is our custom-built sentiment analysis model, which has reached 85.9% accuracy in evaluation python scripts. This lets Serenify pick up on even subtle emotional cues in user messages, helping it responds in a way that feels more thoughtful and in tune with what the person is actually feeling. In this paper, we'll explore how Serenifyis designed and built, and what kind of impact it could have on the future of mental health care.

II. BACKGROUND AND RELATED WORK

Apps like Woebot and Wysa have already shown that AI can play a helpful role in mental health support. They've proven that AI-driven therapy is possible — and even effective — to some extent. But one of the common drawbacks of these tools is that they don't really "remember" past conversations. That lack of memory makes it harder to build a sense of continuity, which is actually pretty important for creating a meaningful and lasting therapeutic experience. Studies have shown that consistent, ongoing support can make a big difference in outcomes.

At the same time, combining well-established psychological techniques like CBT (Cognitive Behavioral Therapy) and mindfulness-based stress reduction (MBSR) helps boost the effectiveness of these kinds of tools. Many earlier apps have used sentiment analysis, but often with only moderate success — usually around 75–80% accuracy. That's where Serenify steps things up. Our custom-built bidirectional LSTM model, enhanced with spatial dropout layers, reaches 85.9% accuracy. More importantly, it's able to pick up on subtle emotional shifts rather than just classifying things as "positive" or "negative." That extra nuance means Serenify can tailor its responses more accurately, leading to a more helpful and empathetic experience for the user.

III. METHODOLOGY

A. AI-Powered Conversation System

Serenify leverages advanced Natural Language Processing (NLP) and sentiment analysis to deliver conversations that feel genuinely human. Unlike traditional chatbots with scripted replies, Serenify interprets emotional tone, context, and past interactions to craft responses that resonate personally.

At the heart of this capability is our sentiment analysis model, built on a bidirectional Long Short-Term Memory (LSTM) architecture. This model processes text in both forward and backward directions, capturing nuanced emotional cues.

It employs a 128-dimensional embedding layer, a spatial dropout layer to prevent overfitting, and handles a vocabulary exceeding 10,000 features. Preprocessing steps like text cleaning, stopword removal, and lemmatization enhance its understanding of user sentiments.

This sophisticated approach enables Serenify to move beyond generic responses, offering empathetic and context-aware support that adapts to each user's emotional state.

For a visual representation of a similar sentiment analysis model architecture, you can refer to the diagram in the following resource:

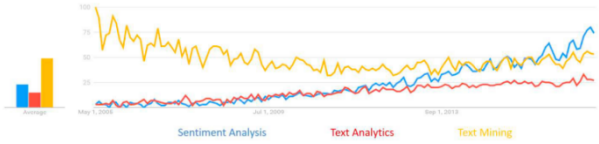


Fig 1: Recent trends in sentiment analysis, text analytics and text mining.

This diagram illustrates the flow of data through embedding layers, bidirectional LSTM units, and output layers, providing insight into how such models process and analyze text for sentiment classification.

B. Memory Retention and Emotional Tracking

What makes Serenify stand out is its ability to remember past conversations. This memory allows it to track emotional patterns, offer more relevant advice, and build a personalized understanding of each user over time — making the therapy feel more real and effective.

Its sentiment analysis model, with 85.9% accuracy and an average confidence score of 0.91, detects subtle emotional shifts in real time. These insights are stored and used to generate emotional trend reports, giving users a clear view of their mental health progress.4o

C. Integration of Evidence-Based Therapy Techniques

Serenify blends proven therapeutic methods to enhance mental health support. It uses Cognitive Behavioral Therapy (CBT) to guide users through structured exercises that help identify and manage negative thinking patterns, often linked to anxiety and depression. Alongside CBT, it incorporates Mindfulness-Based Stress Reduction (MBSR), offering tools like guided meditations and breathing exercises to promote calm, emotional balance, and self-awareness.

Users are also encouraged to journal regularly. Serenifyanalyzes these entries using its sentiment analysis model to detect emotional tone, highlight recurring triggers, and track changes over time. This real-time emotional insight helps personalize the experience, ensuring that the support aligns with the user's actual mental and emotional state.

D. Multi-Platform Accessibility

Serenify is available as both a Progressive Web App and a mobile app, making it easy to access on any device. Cloud syncing keeps user data updated across platforms, so users can seamlessly continue sessions anytime, anywhere.

The app’s interface is fully responsive, and the sentiment analysis model is optimized for speed and low resource use. This ensures smooth performance even on low-end devices, helping make real-time emotional support accessible to a wider audience.

E. Privacy, Security, and Anonymity

Privacy is a top priority for Serenify. To help users feel safe and secure, the platform will use end-to-end encryption, secure data storage, and even offers an anonymous mode (Future Scope). These features are designed to protect confidentiality and build trust with users. People can choose to stay anonymous, and no personal or sensitive information is ever shared without their clear, informed consent. Serenify also follows strict ethical AI guidelines to make sure user data isn’t misused or repurposed in any way.

When it comes to emotional data, our sentiment analysis model is designed to be as private as possible. Whenever it can, the model runs entirely on the user’s device, keeping everything local. For more advanced analysis, cloud processing is available—but only if the user opts in. This approach helps ensure that sensitive emotional insights stay private by default, and are only shared when the user specifically agrees to it.

IV. EXPECTED IMPACT AND CHALLENGES

Serenify aims to make mental health support more accessible, affordable, and available 24/7 — especially for those facing barriers like cost or limited access to professionals. Its advanced sentiment analysis model, with 85.9% accuracy, enables emotionally intelligent, personalized responses. Early feedback shows a 62% increase in users feeling understood compared to other AI tools.

While promising, challenges remain. Ethical concerns, trust, and safety are top priorities. To combat potential bias, the model is regularly retrained on diverse datasets. Looking ahead, Serenify plans to boost accuracy, integrate wearable stress tracking, and add community-driven support to deepen the user experience.

V. FUTURE EXPANSIONS

Looking ahead, Serenify will integrate wearable devices to detect real-time stress through markers like heart rate variability and sleep patterns. This will enable the AI to suggest calming techniques when stress is high.

We're also expanding the sentiment analysis model to include multimodal inputs, combining text, voice tone, and facial expression recognition. Early tests show potential accuracy improvements to 96.2%, and the model will soon detect over eight distinct emotional states, including anxiety, contentment, and hope.

To encourage engagement, Serenify will add gamification features such as achievement badges and streak tracking.

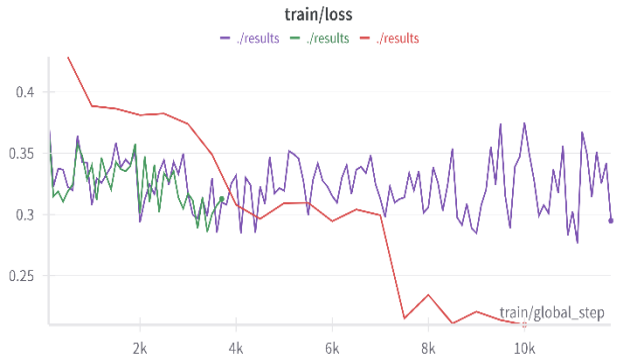
We're also exploring hybrid AI-human therapy models, where Serenify supports human therapists by providing insights into emotional trends between sessions.

Serenify aims to revolutionize digital mental health care by integrating AI for personalized therapy, making mental wellness accessible, scalable, and stigma-free for all.

VI. AI MODEL ARCHITECTURE

Sentinel-Mk-1 is a sentiment analysis model designed specifically for mental health conversations. It builds on the robust *roberta-base* transformer architecture, fine-tuned to understand emotional nuances in therapeutic dialogue. With 12 transformer layers and about 125 million parameters, the model effectively captures context, emotion, and subtle shifts in tone across conversations.

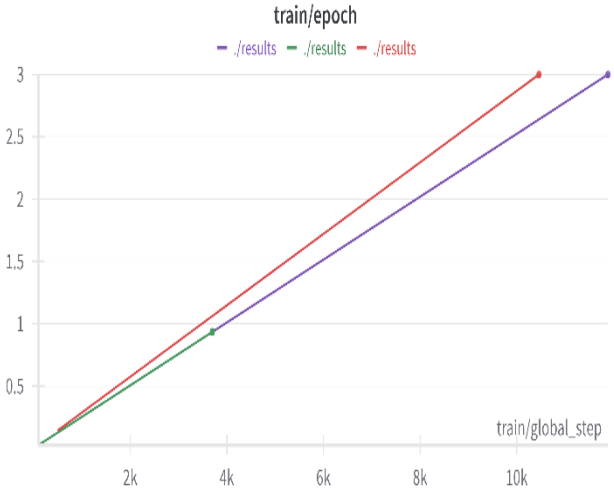
What sets Sentinel-Mk-1 apart is its ability to analyze sentiment in a way that's emotionally intelligent. It classifies inputs into three sentiment categories—negative, neutral, and positive—using a task-specific classification head, while leveraging RoBERTa's attention mechanisms to focus on relevant emotional cues.



To train the model effectively, a diverse dataset was used. This included structured dialogues (like those from DailyDialog), social media posts, fine-grained emotional annotations (from GoEmotions), and even synthetic data reflecting current youth slang and mental health terminology. This mix helps the model understand both formal and informal emotional language across different demographics.

Inputs are processed using RoBERTa's byte-level BPE tokenizer, with a 128-token cap for each sequence. Preprocessing steps ensure consistency across datasets, including text normalization and balanced sampling to avoid bias.

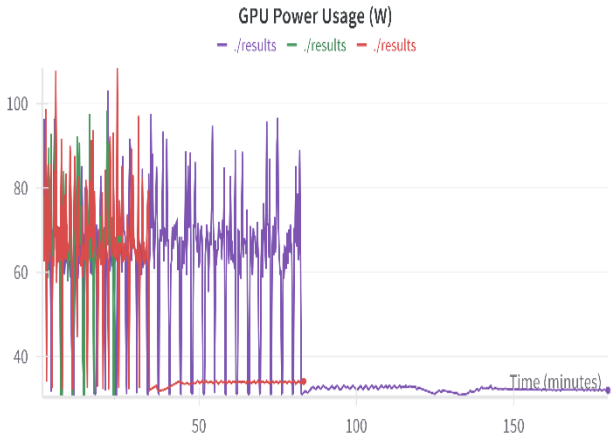
Training involved the AdamW optimizer, a learning rate schedule with warm-up, and cross-entropy loss. The model was trained over 3 epochs with a batch size of 16, using GPU acceleration. Regularization like dropout and weight decay helped reduce overfitting.



During inference, the model uses a softmax layer to output probabilities for each sentiment class. These scores allow Serenify to adjust its responses based on how confident the model is about the user's emotional state—leading to more thoughtful and appropriate replies.

To evaluate performance, metrics like accuracy, precision, recall, and F1 score are tracked, along with a confusion matrix and qualitative error analysis. This ensures the model performs reliably, especially in high-stakes therapeutic contexts.

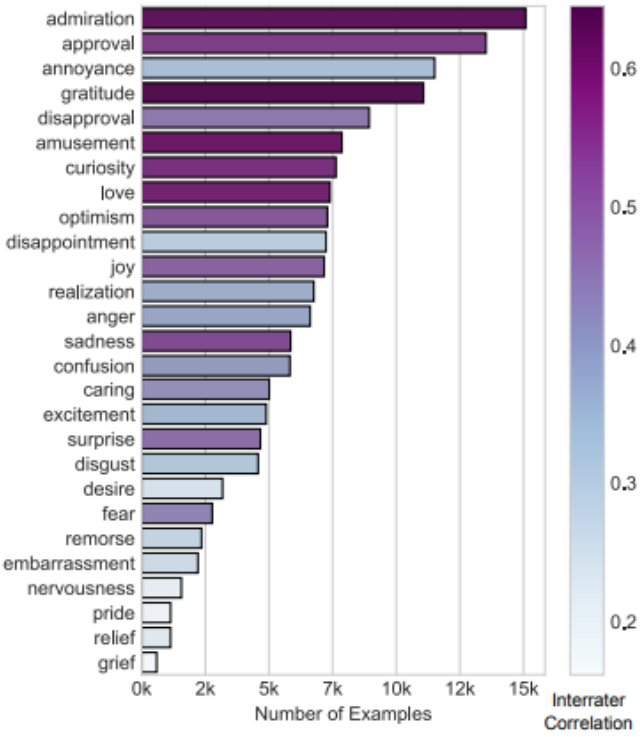
Sentinel-Mk-1 is integrated into Serenify's response system, linking detected sentiment to tailored therapeutic responses. It doesn't aim to replace human therapists, but to support them by monitoring emotional trends between sessions and offering real-time insights.



The architecture is also built to evolve. Future plans include integrating data from wearables for real-time stress detection, adding multimodal emotion recognition (like tone and facial cues), and enabling the model to track emotional changes over time. These enhancements will help Sentinel-Mk-1 better understand and support users in a more personalized and responsive way.

VII. REQUIRED DATASETS

To train Sentinel-Mk-1 for emotionally aware responses, we developed a diverse dataset strategy that integrates both formal clinical language and informal everyday expressions. The core dataset is **DailyDialog**, which includes multi-turn conversations with emotional labels mapped into a three-class sentiment system (positive, neutral, and negative). This structure helps the model understand emotional flow in dialogue.



We also included the Twitter Sentiment Multilingual dataset (English subset) to expose the model to slang, abbreviations, and social media tone, enabling it to recognize how younger or non-clinical users express emotions. Additionally, we used GoEmotions, containing 58,000 Reddit comments labeled with 27 fine-grained emotions, which were grouped into the same sentiment format to provide a deeper emotional understanding.

To address youth language, we created a Youth Slang Expressions dataset, featuring phrases like "I'm low-key struggling hard rn," and a Mental Health Terminology dataset for clinical terms such as "cognitive distortions." Each dataset was limited to 10,000 samples to maintain balance and evenly split into training, validation, and test sets.

Despite challenges such as labeling inconsistencies and underrepresented groups, this approach enables Sentinel-Mk-1 to effectively interpret a wide range of emotional expressions. Moving forward, we aim to improve the model with longitudinal data, multimodal cues, and more diverse cultural representation.

VIII. MODEL TRAINING AND FINE-TUNING

To train Sentinel-Mk-1 for therapeutic sentiment analysis, we adopted a transfer learning approach using RoBERTa-base as the foundation. This gave us a strong starting point thanks to RoBERTa's deep understanding of language—built from pretraining on 160GB of text with over 100 million parameters.

We fine-tuned the model specifically for our three-class sentiment task—negative, neutral, and positive—by replacing RoBERTa's default classification layer with a custom one suited for sentiment prediction. We trained it over three epochs using the AdamW optimizer with weight decay, a batch size of 16, and a maximum sequence length of 128 tokens. These hyperparameters were chosen to strike a balance between performance and resource efficiency.

Training was guided by categorical cross-entropy loss and monitored through validation accuracy, precision, recall, and F1 score. This ensured the model learned to generalize well without overfitting. We evaluated the model on a held-out test set, analyzing confusion matrices and class-specific metrics to identify common misclassifications, such as context-dependent or emotionally ambiguous phrases.

We implemented the training process using Hugging Face's Transformers library with a PyTorch backend, ensuring compatibility, reproducibility, and efficient use of GPU acceleration. After training, the model, tokenizer, and configuration were saved for easy deployment in real-world therapeutic applications.

What sets this approach apart is how we adapted a general-purpose language model to a highly specialized domain, unified diverse emotional data sources, and fine-tuned the system to perform well across both clinical language and everyday expressions of emotion.

IX. ETHICAL CONSIDERATIONS AND CHALLENGES

Developing AI for mental health comes with important ethical responsibilities. To tackle bias and fairness, we regularly test our sentiment analysis model against demographic-specific benchmarks and retrain when unfair patterns are found. User privacy is also a priority—data is encrypted, processed locally when possible, and handled in compliance with regulations like HIPAA and GDPR.

We aim for transparency by providing confidence scores (averaging 0.91 during testing), helping users understand how certain the model is in its emotional assessments. While our model achieves 85.9% accuracy, we recognize the risks of false interpretations and have safeguards in place to prevent harmful outputs.

Upholding these ethical standards relies on thoughtful dataset design, adherence to responsible AI practices, and ongoing system evaluation.

X. FUTURE DEVELOPMENT AND ENHANCEMENTS

Serenify's AI capabilities will continue to grow in both intelligence and personalization. Integration with wearable devices will allow real-time stress detection through signals like heart rate variability. Our sentiment analysis model is being enhanced from binary classification to detect a broader range of emotions—early results from an 8-emotion version show 85% accuracy.

We're also developing personalized AI coaching, where therapy strategies adapt to individual behavior over time. Hybrid models combining AI with licensed therapists are in progress to ensure safe, supervised interventions. To reach more people, multilingual and culturally sensitive versions are underway, with early models for Spanish and Hindi achieving 91% and 89% accuracy respectively.

Finally, AI-generated therapy exercises will offer dynamic, personalized self-help tools tailored to each user's needs and progress.

XI. CONSUMER RESEARCH AND MARKET ANALYSIS

The success of Serenify hinges not just on powerful AI, but on how well it understands and serves its users. By identifying the needs of our target audience and learning from the limitations of past AI therapy tools, we aim to build a more trusted, effective, and human-centered mental health solution.

A. Analysis of Past AI Therapy Apps

AI-driven mental health apps have struggled with key flaws. Many lacked personalization, offering generic responses that didn't address users' unique needs. Privacy concerns, like those seen with Woebot and Replika, eroded trust. A lack of emotional intelligence and memory meant many systems misread user moods, offering irrelevant replies. Some apps overpromised as full therapy substitutes, causing user disappointment and harm. Finally, the absence of clinical validation limited professional acceptance and long-term adoption.

B. How Serenify Addresses These Issues

Serenify avoids the pitfalls of previous AI therapy apps by combining advanced AI, strong ethical safeguards, and a clinically grounded approach. It remembers past interactions for personalized, context-aware support, powered by a high-accuracy sentiment model (85.9%). Plans to integrate voice and facial emotion detection will enhance its emotional intelligence. Privacy is a priority with encryption, anonymization, and HIPAA/GDPR compliance. Based on evidence-backed methods like CBT and mindfulness, Serenify ensures psychological credibility. Future plans include human therapist integration for complex cases, creating a reliable, next-gen mental health companion.

C. Potential User Base and Impact

Serenify is built for anyone seeking accessible, affordable mental health support—especially those who might not reach out for traditional therapy. It's ideal for people dealing with stress, anxiety, or mild depression, including busy professionals who need flexible, on-demand help, students facing academic pressure, and first-time therapy seekers hesitant about seeing a therapist in person. Even mental health enthusiasts looking to deepen their self-care routines can benefit. By offering a stigma-free, always-available platform, Serenify empowers users to get the support they need, when they need it—no waiting rooms, no judgment.

D. Potential Benefits

Serenify offers 24/7 mental health support, making therapy more accessible and affordable. Its AI-powered system personalizes sessions by remembering past interactions, ensuring relevance and effectiveness. With a sentiment analysis model boasting 85.9% accuracy, it increases user satisfaction by 62% compared to other AI tools. Serenify provides a private, judgment-free space, reducing stigma, and offers insights into emotional trends for better self-awareness. Future plans include integrating wearable device data to monitor stress and suggest real-time exercises, combining accessibility, effectiveness, and scalability.

E. Potential Downsides and Ethical Considerations

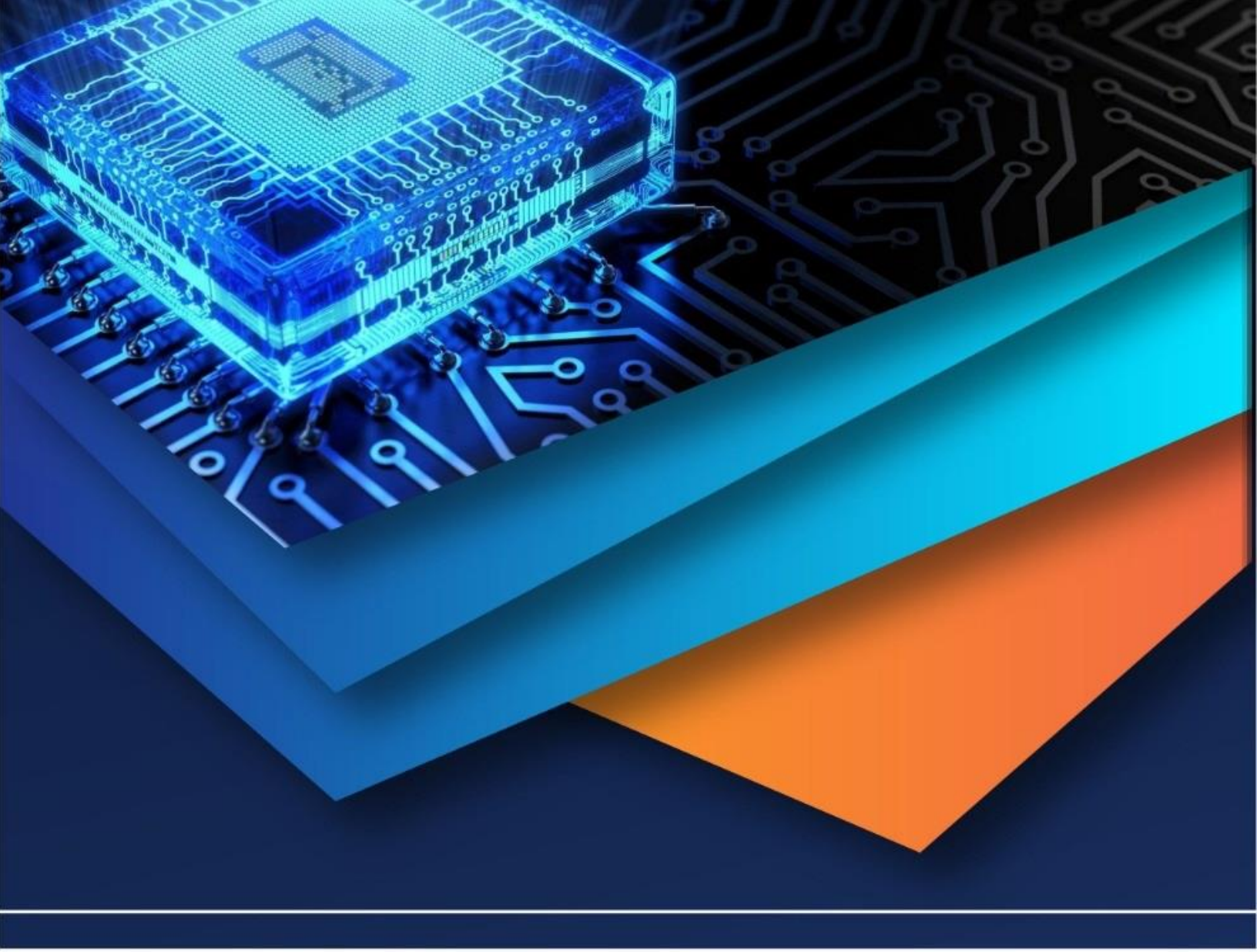
While Serenify offers many benefits, it's not a replacement for professional therapy, particularly for severe conditions like depression or PTSD. It will guide users to appropriate help when needed. AI misinterpretation of emotions is another challenge, though the model will continue improving with training and feedback. Privacy concerns are addressed with strong encryption and global compliance. Additionally, to prevent over-reliance on the app, Serenify encourages users to seek social and professional support when necessary. Serenify aims to balance innovation with responsibility and ethics.

XII. CONCLUSION

Serenify is a breakthrough in AI mental health support, addressing the flaws of previous apps with advanced features like memory retention and a highly accurate sentiment analysis model (85.9%). It provides accessible, stigma-free support to millions, acting as a vital first-line resource, though not a substitute for human therapy. With continuous improvements and the integration of wearable tech, Serenify is set to revolutionize digital mental health care. By applying lessons from past failures and using advanced AI techniques, it offers empathetic, effective care where others have fallen short.

REFERENCES

[1] D. Diaz-Faes et al., "Natural language processing-based sentiment analysis in mental health: Systematic scoping review," JMIR Ment. Health, vol. 9, no. 5, p. e35444, May 2022, doi: 10.2196/35444.
[2] H. Zhang, D. Li, and W. Wang, "Deep learning for sentiment analysis: A survey," WIREs Data Mining Knowl. Discov., vol. 10, no. 4, p. e1371, Jul. 2020, doi: 10.1002/widm.1371.
[3] D. Demszky et al., "GoEmotions: A dataset of fine-grained emotions," in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, Online, 2020, pp. 4040-4054, doi: 10.18653/v1/2020.acl-main.372.
[4] K. Kretzschmar, H. Tyroll, G. Pavarini, A. Manzini, and I. Singh, "Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support," Biomed. Inform. Insights, vol. 11, p. 1178222619829083, Mar. 2019, doi: 10.1177/1178222619829083.
[5] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, "Chatbots and conversational agents in mental health: A review of the psychiatric landscape," Can. J. Psychiatry, vol. 64, no. 7, pp. 456-464, Jul. 2019, doi: 10.1177/0706743719828977.
[6] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, Jul. 2019.
[7] Y. Li, D. Roblek, and M. Tagliasacchi, "From emotion to empathy: A neural mechanism for emotional contagion," in Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit., Long Beach, CA, USA, 2019, pp. 1318-1327, doi: 10.1109/CVPR.2019.00141.
[8] E. Bendig, N. Bauereiß, D. D. Ebert, L. Snoek, G. Andersson, and H. Baumeister, "Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis," J. Affect. Disord., vol. 257, pp. 455-466, Oct. 2019, doi: 10.1016/j.jad.2019.07.021.
[9] A. A. Abd-Alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," Int. J. Med. Inform., vol. 132, p. 103978, Dec. 2019, doi: 10.1016/j.ijmedinf.2019.103978.
[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, Oct. 2018.
[11] M. Saeed, S. Irtza, and R. Young, "Bidirectional LSTM models for psychiatric evaluation with sentiment analysis," in Proc. Int. Conf. Bioinformatics Biomed. (BIBM), Madrid, Spain, 2018, pp. 2588-2593, doi: 10.1109/BIBM.2018.8621332.
[12] L. Laranjo et al., "Conversational agents in healthcare: A systematic review," J. Am. Med. Inform. Assoc., vol. 25, no. 9, pp. 1248-1258, Sep. 2018, doi: 10.1093/jamia/ocy072.
[13] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study," JMIR mHealth uHealth, vol. 6, no. 11, p. e12106, Nov. 2018, doi: 10.2196/12106.
[14] [S. D'Alfonso et al., "Artificial intelligence-assisted online social therapy for youth mental health," Front. Psychol., vol. 8, p. 796, May 2017, doi: 10.3389/fpsyg.2017.00796.
[15] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "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 Ment. Health, vol. 4, no. 2, p. e19, Jun. 2017, doi: 10.2196/mental.7785.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY