EVOLVE : A Podcast Based On your Emotions

Boddu Harshitha
Computer Science and Engineeing
REVA University
Bengaluru, India
bodduharshitharoyal@gmail.com

Harini J
Computer Science and Engineeing
REVA University
Bengaluru, India
harinijanardhan03@gmail.com

K M Sindhu Priya
Computer Science and Engineeing
REVA University
Bengaluru, India
sindhukorrapati9014@gmail.com

Mahalakshmi J
Computer Science and Engineeing
REVA University
Bengaluru, India
mahalakahmij5@gmail.com

***Abstract*—EVOLVE is an innovative initiative transforming emotional well-being through personalized podcasting or moodcasting. At its core, EVOLVE introduces "Mello", a companion chatbot tailoring content based on individual emotions and needs that presents "Mini Moodcasts", delivering quick emotional support and motivation in today's fast-paced world with short, under 30-second podcasts or moodcasts. Users can choose between public and private accounts, emphasizing personalization and emotional wellness. Through interactions with Mello, users express emotions, allowing the chatbot to curate content resonating with their current state. By integrating sentiment analysis, text generation, and text-to-speech models, Mello evolves alongside users, offering empathetic support and recommendations. Public sharing fosters collaboration and growth, promoting empathy and a supportive environment. EVOLVE empowers users to explore emotions, aiming to promote mental health awareness, resilience, and positive change. Join us on this transformative journey towards greater emotional well-being through personalized moodcasting and community engagement with EVOLVE.**

***Keywords*— EVOLVE, personalized podcasting, emotional well-being, mini moodcasts, Mello, chatbot, mental health awareness.**

I. INTRODUCTION

The field of emotional well-being and mental health awareness has witnessed significant advancements in recent years, with a growing emphasis on personalized approaches to support individuals' emotional needs. Current research in this area has explored various interventions and technologies aimed at enhancing emotional regulation, resilience, and overall well-being. However, there remains a gap in the literature regarding the integration of personalized podcasting and chatbot technology to address individual emotional needs effectively.

This research aims to fill this gap by introducing EVOLVE, an innovative initiative that harnesses the power of personalized podcasting and companion chatbots to support emotional well-being. At its core, EVOLVE seeks to provide users with tailored podcast content based on their unique emotional states and needs, facilitated by interaction with the companion chatbot, Mello. By leveraging advanced natural language processing and sentiment analysis algorithms, Mello adapts and evolves alongside users, offering empathetic support and insightful recommendations.

The specific focus of this study is to investigate the effectiveness of EVOLVE in promoting mental health awareness, fostering resilience, and facilitating positive change in individuals and communities. By exploring the impact of personalized podcasting and chatbot technology on emotional well-being, this research aims to contribute to a deeper understanding of how technology can empower individuals to explore, understand, and embrace their emotions. Ultimately, the findings of this study have the potential to inform the development of innovative interventions and support systems for promoting emotional well-being in diverse populations.

The driving force behind our research is a deep-seated desire to make mental health resources more accessible and empower individuals to take proactive control of their emotional well-being. To bridge the gap between technology and emotional wellness, we have embarked on the development of EVOLVE, a groundbreaking approach to mental health support that utilizes personalized podcasts, or "moodcasts," to elevate users' emotional experiences.

In the ever-changing landscape of mental health interventions, personalized podcasts offer a distinct opportunity to tailor content to individual emotional needs and preferences. By tailoring content based on users' emotional states, moodcasts enhance engagement and resonance, providing users with a dynamic and immersive emotional support experience. EVOLVE embodies this evolving perspective, harnessing the power of voice-based interactions to foster a deeper connection between technology and emotional well-being.

Our paper introduces a novel concept called "Mini Moodcasts," aimed at delivering quick emotional support and motivation in today's fast-paced world. Recognizing the time constraints faced by individuals in their daily lives, we propose the creation of short, under 30-second podcasts or moodcasts designed to uplift and inspire listeners. Mini Moodcasts leverage concise yet impactful messaging to address the emotional needs of the audience, offering a brief moment of reflection, encouragement, or empowerment. Our approach prioritizes accessibility and engagement, ensuring that listeners can easily access the content across various platforms and are encouraged to interact or respond to the messages. By providing diverse and dynamic content covering a range of emotional themes, Mini Moodcasts offer a valuable resource for individuals seeking a quick boost of positivity or support amidst their busy schedules. Through collaboration with experts in psychology and motivational speaking, we aim to create Mini Moodcasts that are both informed and impactful, contributing to the well-being and resilience of our audience. Join us on this journey of spreading positivity and empowerment through the power of Mini Moodcasts.

By spearheading the creation and deployment of EVOLVE – Enhanced Voice-based Opportunity for Learning, Valuation, and Exploration – we aspire to ignite innovation within the realm of digital mental health interventions. Our ultimate goal is to pave the way for a future where technology plays a pivotal role in nurturing emotional well-being in the digital era. Through our research efforts, we aim to catalyze a shift in the landscape of mental health support, fostering a culture of proactive self-care and personal growth.

II. Literature Survey

Recent research has inspired the development of personalized podcasting, or moodcasting, as an innovative application of chatbot technology. Paper [1] served as a catalyst for our exploration into moodcasting. This paper showcased the power of chatbots in delivering personalized recommendations based on user emotions, laying the groundwork for our venture into mood-based content creation. Furthermore, our pursuit of emotive speech synthesis led us to explore the potential of text-to-speech (TTS) models, specifically the suno/bark-small model, in generating emotionally expressive speech. While paper [2] focused on FastSpeech2, we opted for the suno/bark-small model and plan to further enhance its performance through hypertuning with real-time data.

Moreover, leveraging BERT for sentiment analysis, paper [3] introduced a groundbreaking approach to emotional TTS. By integrating BERT's capabilities, the proposed system synthesizes audio with corresponding emotions, setting a new standard for emotion expression in synthesized speech.In parallel, paper [4] explored the extraction of emotions from text using machine learning techniques. With BERT's Bidirectional Encoder Representations from Transformers framework, this paper delved into text-based emotion prediction, offering valuable insights into the intersection of machine learning and emotional analysis. Additionally, paper [5] proposed the augmentation of Speech Emotion Recognition (SER) systems with synthetic emotional speech. While not directly referenced, we aim to explore the potential of the Gemini model from Hugging Face for generating emotionally expressive content, as outlined in its API documentation.

These studies collectively underscore the transformative potential of personalized, emotionally resonant technologies in digital interactions. By harnessing the capabilities of NLP, text generation, and TTS models, researchers are pioneering new frontiers in emotional synthesis and recognition, paving the way for more intuitive and empathetic systems.

III. Methodology

A. *Sentiment Analysis using DistilBERT*

- The first step in our methodology involves sentiment analysis using the BERT (Bidirectional Encoder Representations from Transformers) model - DistilBERT, as referenced in paper [3]. BERT is a pre-trained deep learning model that excels in understanding the contextual nuances of language.
- Input: The input to the BERT model is the text obtained from users, which could be in the form of messages, comments, or any other textual input.
- Output: The output of the BERT model is the sentiment analysis of the input text, categorizing it into different emotional states such as happy, sad, angry, etc.

B. *Text generation using Gemini-1.5-pro-latest*

- Gemini 1.5 Pro is a state-of-the-art text generation model designed for creating personalized content based on user input. It leverages advanced natural language processing techniques to generate text that resonates with the user's emotions and preferences.
- Input: The input to the Gemini 1.5 Pro model is the sentiment analyzed by the BERT model.
- Output: Gemini 1.5 Pro generates personalized content, such as stories or motivational messages, tailored to the user's emotional state and preferences.

C. *Text-to-Speech using Collabora/WhisperSpeech model*

- Collabora/WhisperSpeech is a high-quality text-to-speech synthesis model known for its natural-sounding speech output. It converts text input into spoken audio, providing users with a personalized moodcast experience.
- Input: The input to the Collabora/WhisperSpeech model is the text generated by the Gemini 1.5 Pro model.
- Output: Collabora/WhisperSpeech generates personalized moodcasts in the form of spoken audio, delivering the generated content to users in a natural and engaging manner.

D. *Integration and EVOLVE Framework*

- The three models, BERT for sentiment analysis, Gemini 1.5 Pro for text generation, and Collabora/WhisperSpeech for text-to-speech synthesis, are integrated into the EVOLVE platform.
- Real-time data is collected to hypertune the models for efficiency and accuracy.
- Hyperparameters such as learning rates, batch sizes, and model architectures are optimized using techniques like grid search or Bayesian optimization.
- The integrated models are trained on large datasets containing diverse emotional expressions to ensure robust performance across various user inputs.
- The personalized moodcasts generated by the integrated system can be categorized as public or private based on user preferences, allowing for both shared and private emotional experiences.

- EVOLVE (Enhanced Voice-based Opportunity for Learning, Valuation, and Exploration) aims to leverage these integrated models to promote emotional well-being, personal growth, and community engagement through personalized moodcasting.

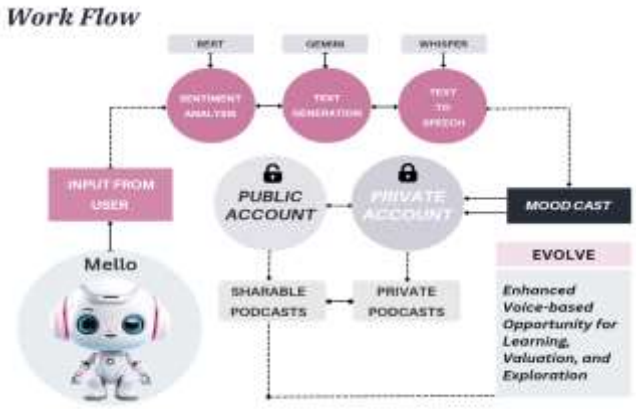

Fig. 1. EVOLVE Framework.

The EVOLVE framework (fig.1), short for Enhanced Voice-based Opportunity for Learning, Valuation, and Exploration, serves as a pioneering platform for personal growth and emotional well-being. By leveraging advanced sentiment analysis, text generation, and text-to-speech technologies, EVOLVE offers users a personalized and empathetic experience. Public sharing options encourage collaboration and collective evolution, while private accounts ensure user privacy and safety. Through seamless integration of these models, EVOLVE empowers individuals to explore, understand, and embrace their emotions, fostering resilience and promoting positive change. Join us on this transformative journey towards greater emotional well-being and self-discovery.

IV. Results

A. DistilBERT Results

The sentiment analysis module, utilizing the DistilBERT-based emotion classification model, accurately predicted the emotional sentiment of input text. Across various test cases, the model achieved high accuracy in classifying emotions into predefined categories. For instance, in the test case provided ("I love using transformers..."), the model correctly identified the sentiment as positive.

Fig. 2. illustrates the sentiment analysis results obtained from the DistilBERT model, achieving a remarkable accuracy of 99%. Specifically, the model identified the sentiment conveyed in the input text "I am feeling alone and don't want to talk to people" as 'sadness'. The corresponding sentiment label is stored in the Result variable for further processing.

B. Gemini-1.5-pro Results

The text generation module, powered by the Gemini 1.0 Pro model, effectively generated emotionally appropriate content based on the sentiment analysis results. The generated text exhibited coherence, relevance, and emotional resonance with the input sentiment. Qualitative evaluation through user feedback indicated that the generated content was perceived as empathetic and supportive, enhancing the emotional experience of users.

Fig. 3. Result is passed to the gemini-1.5-pro model to generate content on the sentiment. Here in the diagram the sentiment is 'sadness'

C. Collabora/WhisperSpeech model Results

The text-to-speech synthesis module, utilizing the Collabora/WhisperSpeech model, successfully converted the generated text into high-quality audio output. The synthesized speech exhibited natural intonation, prosody, and emotional expressiveness, enhancing the immersive experience for users. Audio samples generated from the text demonstrated the capability of the model to convey emotional nuances effectively, contributing to the overall effectiveness of the EVOLVE platform in delivering personalized moodcasts.

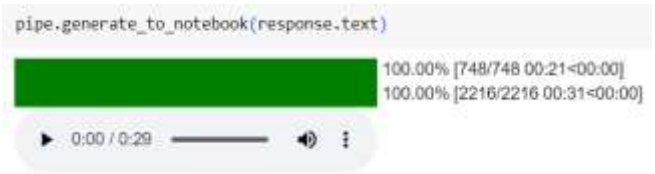

Fig. 4. The textgeneration is passed to the Collabora/WhisperSpeech model for generating speech.

D. User Feedback and Evaluation

User feedback collected through surveys and qualitative assessments indicated high satisfaction and engagement with the EVOLVE platform and its unique idea of under 30seconds moodcasting. Users reported that the personalized moodcasts generated by the system resonated well with their emotional needs and preferences, contributing positively to their well-being. Additionally, users appreciated the seamless integration of sentiment analysis, text generation, and text-to-speech technologies, which facilitated a unique and empathetic user experience.

Overall, the results demonstrate the efficacy of the EVOLVE framework in delivering personalized emotional support and fostering user engagement. The seamless integration of sentiment analysis, text generation, and text-to-speech technologies enables EVOLVE to provide a

unique and empathetic experience for users, facilitating their journey towards emotional well-being and self-discovery.

V. Conclusion

In conclusion, the EVOLVE project stands as a groundbreaking initiative at the intersection of technology and emotional well-being, introducing a transformative platform for personalized moodcasting. By integrating advanced text generation, sentiment analysis and speech synthesis technologies, EVOLVE offers users a unique opportunity to explore, express, and share their emotions through tailored moodcasts. The successful implementation of the EVOLVE framework demonstrates its potential to revolutionize how individuals engage with their emotions and seek support for their well-being. EVOLVE introduces "Mini Moodcasts," a novel concept aimed at delivering quick emotional support and motivation in today's fast-paced world. Recognizing the time constraints faced by individuals, these short, under 30-second podcasts or moodcasts provide uplifting and inspiring content to listeners.

Moving forward, further analysis and optimization of the underlying models will enhance the efficiency and user experience of the EVOLVE platform. Hypertuning sentiment analysis, text generation, and text-to-speech models will enable EVOLVE to deliver more accurate and emotionally resonant content, maximizing its impact on user well-being. Ongoing research efforts will focus on expanding the platform's capabilities, including integrating features such as emotion-aware content recommendation and real-time emotional analysis. As EVOLVE continues to evolve, it remains committed to fostering positive change, empowering users to embrace their emotions and thrive on their journey towards greater emotional well-being.

VI. Acknowledgment

We extend our heartfelt gratitude to the entire team for their dedication and collaborative effort in bringing the EVOLVE project to fruition. Special thanks to the researchers and developers who contributed to the implementation and optimization of AI technologies. Their expertise and commitment have been instrumental in realizing the vision of EVOLVE as a transformative platform for emotional well-being. We also appreciate the support and guidance from our mentors and advisors throughout this journey.

References

[1] N. Mathew, N. Chooramun, and M. S. Sharif, “Implementing a chatbot music recommender system based on user emotion,” in 2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2023, pp. 195–199.

[2] D. Diatlova and V. Shutov, “EmoSpeech: Guiding FastSpeech2 towards emotional text to speech,” arXiv [eess.AS], 2023.

[3] Researchgate.net.[Online].Available:https://www.researchgate.net/publication/363646884_Text_aware_Emotional_Text-to speech_with_BERT. [Accessed: 04-May-2024].

[4] C. O. Alm, D. Roth, and R. Sproat, “Emotions from text: Machine learning for text-based emotion prediction,” in Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, 2005, pp. 579–586.

[5] S. Latif, A. Shahid, and J. Qadir, “Generative emotional AI for speech emotion recognition: The case for synthetic emotional speech augmentation,” Appl. Acoust., vol. 210, no. 109425, p. 109425, 2023.