Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

EmotionSync: Emotion-Aware Recommendation and Response Generation Systems

Gitesh Patil
Dept. of Computer Engineering
Pune Institute of Computer Technology
Pune, India

Sakshi Mahajan
Dept. of Computer Engineering
Pune Institute of Computer Technology
Pune, India

Shloka Shetty
Dept. of Computer Engineering
Pune Institute of Computer Technology
Pune, India

Samruddhi Nevse
Dept. of Computer Engineering
Pune Institute of Computer Technology
Pune, India

Dr. Arati Deshpande
Dept. of Computer Engineering
Pune Institute of Computer Technology
Pune, India

***Abstract*—Modern conversational AI systems increasingly require emotional intelligence to enhance user engagement and support emotionally sensitive applications such as mental health counseling. This survey reviews recent advancements in emotion-aware dialogue, recommendation, and empathetic response generation systems. We categorize methods across classical ML, deep learning, and multimodal approaches, present comparison insights, highlight research gaps, and outline the proposed EmotionSync architecture for real-time emotion-driven conversation and recommendation.**

***Index Terms*—Affective Computing, Emotion-Aware AI, Empathetic Dialogue Systems, Sentiment Analysis, Emotional Recommendation Systems, Transformer Models, Multimodal Learning, Natural Language Processing (NLP)**

I. Introduction

In the modern era of Artificial Intelligence, conversational systems have progressed significantly from basic rule-based or text-only chatbots to advanced virtual assistants.These chatbots are capable of understanding context, intent, and even human emotions. While these systems demonstrate strong performance in information retrieval and task automation, most remain emotionally neutral, which limits their effectiveness in domains that require sensitivity, trust, and sustained user engagement ,particularly in mental health support and well-being applications. Emotion-aware AI seeks to solve this problem by incorporating principles from affective computing, sentiment analysis, and deep learning to recognize emotional cues from user inputs. By enabling machines to understand and interpret emotions, these systems can now generate more empathetic, supportive, and human-like interactions, thereby improving user satisfaction and emotional connection.

This survey paper examines a range of latest approaches to emotion detection, empathetic response generation, and emotion-informed recommendation systems, offering a comparative analysis of state-of-the-art methods proposed in recent research. It identifies key challenges such as fragmented system designs, limited personalization, poor generalization to unseen emotions, and deployment complexity. Building upon these insights, the paper introduces EmotionSync, a proposed multimodal architecture that unifies emotion detection, context-aware recommendation delivery, and empathetic response generation within a single framework. By using multimodal emotional cues and continuous feedback, EmotionSync aims to improve user interaction quality, deliver adaptive personalization, and provide meaningful support for mental well-being through emotionally intelligent conversation.

II. Literature Review

A. Emotion Detection Models

1. Zhang et al. (2020) proposed a BERT-based emotion classifier integrated with a transformer decoder for dialogue generation. It effectively captured contextual emotional transitions, but required large annotated datasets and significant computation for training.

2. Poria et al. (2018) introduced a multimodal emotion recognition model using text, audio, and video cues for human-computer interaction. Although highly accurate, the multimodal fusion made deployment complex and resource heavy.

3. Hazarika et al. (2020) presented a Conversational Memory Network (CMN) that tracked emotional context across conversation. While emotion continuity improved, memory requirements were high and it had limited scalability for real-time use.

B. Empathetic Response Generation Models

4. Rashkin et al. (2019) released the EmpatheticDialogues dataset and trained models to produce empathetic responses. Their method improved emotional tone, but lacked personalization and real-time emotion adaptation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

5. Zhou et al. (2020) introduced the Emotional Chatting Machine (ECM), where emotion embeddings guided Seq2Seq responses. It generated emotionally aligned replies, but sometimes overfit emotion labels, reducing generalization.

6. Lin et al. (2021) combined sentiment analysis with reinforcement learning to dynamically control emotional tone in conversations. The model generated emotionally adaptive responses but was computationally expensive.

7. Majumder et al. (2020) proposed a GRU-based dialogue system that maintained emotional consistency. While effective in emotion continuity, it struggled with long multi-turn dialogues and complex context retention.

C. Emotion-Aware Recommendation Systems

8. Li et al. (2021) presented a sentiment-enhanced recommendation hybrid that combined collaborative filtering with sentiment signals. It improved user satisfaction but performed poorly in cold-start situations.

9. Sun et al. (2021) developed a transformer-based multimodal recommendation model leveraging emotional cues in conversation text. Recommendations were contextually accurate, but performance dropped for new emotional states.

D. Unified Empathetic Dialogue + Recommendation Architectures

10. EmotionSync (Proposed Study) integrates emotion detection, empathetic response generation, and personalized recommendation. Unlike previous works focusing on individual components, EmotionSync combines multimodal emotion analysis, context retention, and adaptive personalization for mental-health support applications.

III. Research Gaps

There is a notable lack of unified conversational AI systems that combine emotion detection, empathetic response generation, and personalized recommendation mechanisms within a single framework. Most existing solutions treat these components as independent modules, resulting in interactions that fail to adapt completely to the user's emotional and contextual state. Furthermore, many systems do not incorporate real-time emotional feedback or continuous learning, which significantly limits their ability to personalize interactions over time and respond effectively to changes in user behavior or mental state. This shortcoming reduces long-term engagement and reduces the potential impact of such systems, particularly in sensitive applications like mental health support.

In addition, current emotion-aware models often struggle when exposed to unseen, ambiguous, or rare emotional states, largely due to over-reliance on predefined emotion labels and limited generalization capabilities. The lack of robust multimodal and deployment-friendly architectures further constrains real-world adoption, as many approaches either depend on a single modality (such as text alone) or are too computationally complex for scalable deployment across platforms. Beyond technical limitations, serious concerns remain regarding bias, fairness, and ethical responsibility in emotionally sensitive conversations. In mental health contexts especially, biased emotion interpretation, unsafe response generation, or inadequate safeguards can lead to harmful outcomes, underscoring the need for transparent, accountable, and ethically grounded emotion-aware AI systems.

IV. Comparative Review

Study	Emotion Input	Response Method	Strength	Limitation
Rashkin et al. (2019)	Text	Seq2Seq	Empathy learning	Limited personalization
ECM (Zhou et al., 2020)	Text	Emotion Embedding	Strong emotional tone	Overfits emotion labels
Poria et al. (2018)	Text/Audio/Video	Fusion	High accuracy	Complex deployment
Sun et al. (2021)	Text	Transformer	Context awareness	Poor unseen emotion handling
Hsu et al. (2021)	Multimodal	LLM-based	Rich emotion capture	High cost

Fig. 1: Comparative analysis of emotion-aware dialogue systems

V. Proposed Approach:EmotionSync

EmotionSync is a hybrid AI system combining emotion detection, response generation, and recommendation subsystems. The architecture uses a CNN-LSTM model for emotion classification based on text and audio features. Once the dominant emotion is detected, it triggers the response generator (using fine-tuned GPT or T5 models) and the recommendation engine (using vector embeddings and user context).

Key Components:

- Emotion Detection: CNN-LSTM based hybrid model with pre-trained embeddings (GloVe or BERT).
- Response Generation: Transformer-based model fine-tuned for empathetic dialogue (GPT- 2 or DialoGPT).
- Recommendation Engine: Contextual and emotion-aware retrieval model using cosine similarity on vectorized user histories.
- Integration Layer: REST API communication between emotion detection, NLP response generator, and recommendation subsystems. EmotionSync's advantage lies in combining emotional understanding with personalized engagement which is crucial for mental health applications.

VI. Proposed Architecture: EmotionSync

The proposed EmotionSync AI architecture is designed as a unified system comprising of four tightly integrated modules. It begins with an input layer that accepts a detected emotion vector along with dialogue context and user profile information. This input is processed by a response generation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

module that uses open-source large language models such as GPT-Neo or LLaMA, enhanced through emotion-aware prompting to generate empathetic and contextually appropriate replies.A recommendation system module combines rule-based emotional triggers with machine-learning–based ranking approaches, such as LightFM or Surprise, to deliver personalized suggestions. These outputs are brought together in a fusion layer with a feedback loop that integrates generated responses and recommendations while continuously tracking user emotions to enable ongoing system improvement. Key innovations of this architecture are emotion-conditioned prompting for enhanced empathy, continuous personalization driven by emotional feedback, and real-time deployment with a 3D avatar to support engaging human–AI interaction.

VII. Discussion

The integration of emotion detection, recommendation, and response generation modules enables EmotionSync to function as a cohesive and emotionally intelligent system that can maintain empathy while adapting to evolving user needs. By continuously analyzing the user's emotional state alongside conversational context, the system is able to generate responses that are not only coherent but also emotionally appropriate and supportive. Simultaneously, the recommendation component leverages this emotional understanding to suggest actions, resources, or content that match the user's current state, creating a more personalized interaction experience. Unlike prior models that treat emotion recognition as an isolated preprocessing step or restrict emotional cues to surface-level response tuning, EmotionSync embeds emotional intelligence directly into the decision-making and conversational flow. This unified architecture allows emotional insights to influence both what the system says and what it recommends, resulting in smoother dialogue transitions, stronger contextual relevance, and more human-like interactions. As a result, EmotionSync bridges the gap between emotion-aware perception and intelligent action, enabling sustained empathy, adaptive personalization, and more meaningful long-term engagement with users.

VIII. Conclusion

Prior research in emotion-aware conversational AI has explored various strategies for incorporating emotional understanding into dialogue systems, yet each approach has many limitations. Early sequence-to-sequence models, such as the work by Rashkin et al. (2019), emphasized empathy learning from textual inputs but suffered from limited personalization. Emotion-embedding–based models like ECM (Zhou et al., 2020) improved emotional tone consistency but often overfit to predefined emotion labels, reducing flexibility in real-world interactions. Multimodal fusion approaches, including Poria et al. (2018), achieved higher emotion recognition accuracy by combining text, audio, and video signals, though their complexity hindered scalable deployment. Transformer-based models (Sun et al., 2021) enhanced contextual awareness but struggled with unseen emotional states, while recent LLM-based systems (Hsu et al., 2021) captured richer emotional cues at the cost of high computational and deployment overhead.

A key limitation across these existing models is their design, where emotion detection, response generation, and recommendation logic operate largely in isolation. Most systems rely on uni-modal emotion detection—typically text or speech alone—resulting in shallow emotional understanding. Response generation is often constrained to fixed emotion tags, producing repetitive or rigid emotional expressions, while recommendation mechanisms tend to be static and rule-based, offering limited adaptability over time. Furthermore, personalization is generally minimal, with little to no incorporation of long-term user profiles or emotional histories, thereby restricting the system's ability to evolve with user needs.

EmotionSync addresses these gaps through a unified, end-to-end architecture that tightly incorporates multi-modal emotion detection, emotion-aware response generation, and adaptive recommendation delivery. Unlike existing models, EmotionSync leverages both textual and speech cues for deeper emotional understanding, employs open-source LLMs with emotion-conditioned prompting to generate empathetic and context-sensitive replies, and combines rule-based emotional triggers with machine-learning based ranking techniques such as LightFM or Surprise for dynamic personalization. By maintaining user profiles and continuously incorporating emotional feedback, EmotionSync enables sustained personalization and adaptability, bridging the divide between emotional intelligence and decision-making. This design positions EmotionSync as a more scalable, empathetic, and user-centric alternative to pre-existing emotion-aware conversational AI systems.

References

[1] T. Zhang et al., "ECM: An Emotionally Intelligent Chatbot," in Proc. AAAI Conf. Artificial Intelligence, 2018.
[2] Y. Li et al., "DailyDialog: A manually labelled multi-turn dialogue dataset," in Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL), 2017.
[3] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models," in Proc. ACL, 2019.
[4] L. Wang et al., "Empathetic Dialogue Generation via Sensitive Emotion Recognition and Sensible Knowledge Selection," in *Findings of the Association for Computational Linguistics: EMNLP 2022*, Abu Dhabi, United Arab Emirates, Dec. 2022, pp. 4634–4645, Association for Computational Linguistics, doi:10.18653/v1/2022.findings-emnlp.340.
[5] S. Poria et al., "A review of affective computing: From unimodal analysis to multimodal fusion," Information Fusion, vol. 37, pp. 98-125, 2017.
[6] Y. Dong et al., "Controllable Emotion Generation with Emotion Vectors," *arXiv preprint* arXiv:2502.04075v1, 2025.
[7] N. Asghar et al., "Affective neural response generation," Machine Learning, vol. 107, no. 11, pp. 2145-2177, 2018.
[8] P. Colombo, C. Clavel, and G. Staiano, "Affect-driven dialog generation," in Proc. ACL, 2019.
[9] N. Lubis, S. Sakti, K. Yoshino, and S. Nakamura, "Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, New Orleans, Louisiana, USA, Feb. 2018, pp. 5293–5300.
[10] R. Shantala, G. Kyselov, and A. Kyselova, "Neural Dialogue System with Emotion Embeddings," in *Proceedings of the 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, IEEE, 2018, doi:10.1109/SAIC.2018.8516696.