Emotional State Engine: A Dynamic Frame work Modeling Emotions for Conversational AI conditioned with Emotion

Mukiibi Moses, Shema Nkindi Giscard, Lymo Johnson Samwel, Tirop Meshack

Department of Smart Computing, Kyungdong University Global

Republic of Korea , March 2026

ABSTRACT

Recent developments in conversational artificial intelligence have created systems that can generate fluent and contextually consistent conversation. Nevertheless, the majority of conversational actors do not possess inertial emotional awareness, which constrains their potential to have an empathetic conversation. Current emotion-sensitive dialogue systems are generally designed to assume that emotions are fixed labels that can be attributed to individual utterances, but not the dynamic characteristics of human emotional conditions. This paper describes the Emotional State Engine (ESE), a computational model of emotional dynamics in conversation with a continuous Valence-Arousal-Dominance (VAD) representation and a temporal state updating model. This system combines emotion recognition with a neural dialogue model to produce responses based on the changing emotional state of the user. Through experimental demonstrations, the proposed approach is proved to be emotionally continuous across dialogue turns and enhance appropriate contexts of generated responses. The proposed architecture helps bridge the gap between emotion recognition and dialogue generation, helping to make conversational agents smarter about emotions..

Keywords: *affective computing, conversational AI, emotional dynamics, dialogue systems, empathetic AI*

I. INTRODUCTION

Human communication is inherently emotional. Emotional context influences how individuals interpret language and formulate responses. Consequently, conversational artificial intelligence systems that fail to model emotional dynamics often produce responses that appear insensitive or contextually inappropriate.

Research in *Affective Computing* aims to enable machines to recognize and respond to human emotions. The field was pioneered by Rosalind Picard, who proposed that computing systems should be able to “recognize, interpret, and simulate human emotions” [1]. Since then, advances in machine learning have enabled emotion detection from text, speech, and facial expressions.

Meanwhile, neural language models like GPT-2 and DialoGPT have advanced conversational response generation in a major way. DialoGPT was trained using more than 147 million conversation exchanges and is capable of generating fluent open-domain dialogues responses [2].

Despite these developments, emotional indicators are still regarded by the majority of conversational AIs as fixed labels on individual messages. Nevertheless, psychological studies prove that emotions change through time and affect the following cognition process [3]. Thus, it is crucial to model emotional interactions between turns in dialogue to create more human-like conversational agents.

Recent research has attempted to incorporate emotion awareness into dialogue systems using datasets such as *EmpatheticDialogues*, which contains thousands of emotionally grounded conversations [4]. Although such approaches enable emotion-conditioned response generation, they still lack explicit mechanisms for modeling emotional state transitions.

This paper proposes the **Emotional State Engine (ESE)**, a framework designed to maintain a continuous emotional representation throughout conversation. The system integrates emotion detection with a dynamic state update mechanism and uses this evolving state to guide dialogue generation.
This work has contributed:

1. Valence-Arousal-Dominance (VAD) representation of a dynamic emotional state model.
2. An emotional momentum modeling temporal emotional state update process.
3. Combination of emotional condition with neural conversational generation.
4. Evidence of emotional persistence on conversational reactions.

II. TECHNOLOGICAL GAP

Although affective computing and conversational AI are progressing at a high rate , several limitations persist technologically.

A. Static Emotion Classification

Typically, most Emotion Recognition Systems will classify every message as an individual discrete category of emotions like Joy, Anger or Sadness. Based on lexical approaches (sometimes referred to as lexicon-based approaches) [5] or emotion classifiers using neural networks [6], Emotion Recognition Systems assign only one label per utterance (message). However, this approach fails to identify emotional continuity between interactions. Research has also indicated that emotion representation is complex and temporal in nature, therefore static classification alone is inadequate for representing real-life emotional behavior [7].

B. Weak Integration with Dialogue Models

The main objective of neural dialogue systems, including DialoGPT and other transformer-based architectures, is to generate fluent dialogue rather than have emotional reasoning throughout the course of a conversation (as presented in Ref. 2). Though these systems can create grammatically correct and flowing dialogue, they do not provide methods to sustain emotional state during the course of a conversation. In fact, many emotion-aware dialogue systems are based on low-level conditioning systems (i.e., the addition of an emotion label at the end of the input prompt).

C. Lack of Emotional Memory

For the most part, most conversational agents work as if every interaction is independent from one another, meaning they don't have emotional memory to track if the user's emotional state has improved or not through a conversation. In many cases, empathetic dialogue models (e.g., CAiRE and MoEL) improve the empathy of the responses produced by a conversational agent but still rely primarily on single-turn emotion predictions [8][9].

D. Research Gap

There is a gap between how we detect emotions and how we generate dialogue based on those emotions; in other words, there is a lack of systems that can create a model of an individual's emotional state across many interactions. The Emotional State Engine will help to solve this problem in the following ways:

- Maintaining a continuous and stable emotional state representation
- Updating the emotional state as the dialogue progresses.
- Conditioning generation of dialogues on evolving emotional context and phrases.

III. RELATED WORK

A. Affective Computing

Affective computing's goal is to provide machines with the ability to identify and react appropriately to human emotion. The key idea of Picard's work was that artificial computing systems could be able to identify and respond according to feelings found in physical expressions and emotions [1]. Researchers have since investigated numerous methodologies (physiological, emotional, verbal), in addition to systems for identifying the facial and emotional signals of an individual and adapting the system's corresponding behavior [10].

The Valence-Arousal-Dominance model proposed by Russell provides a continuous representation of emotional states, allowing emotions to be represented in a three-dimensional affective space [3].

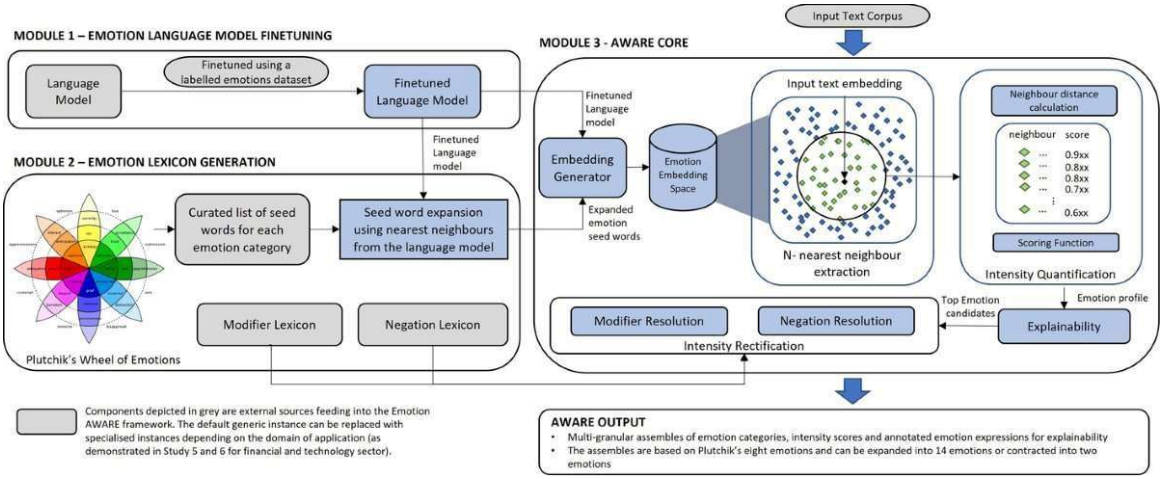

Fig. 1. *Emotion AWARE framework: a multi-module architecture combining a number of fine-tuned language models, emotion lexicon generation or Plutchik's Wheel of Emotions, and nearest-neighbour (KNN)embedding-based classification to explain emotion intensity quantification [developed from related literature].*

B. Emotion Recognition in Natural Language Processing

The topic of emotion detection within text has been extensively explored through the utilization of both lexicon-based techniques and machine learning models. To support large-scale, emotion classification efforts, word-emotion association lexicons (developed by Mohammad and Turney) have been created [5]. Recent advancements in deep learning have made use of transformer architectures to identify emotion signals in textual data [6]. Figure 2 presents an exemplary multimodal architecture for emotion recognition, which integrates speech emotion recognition (SER) and text emotion recognition (TER) embeddings using cross-attention mechanisms (CMT), to allow for improved detailed emotion classification results.

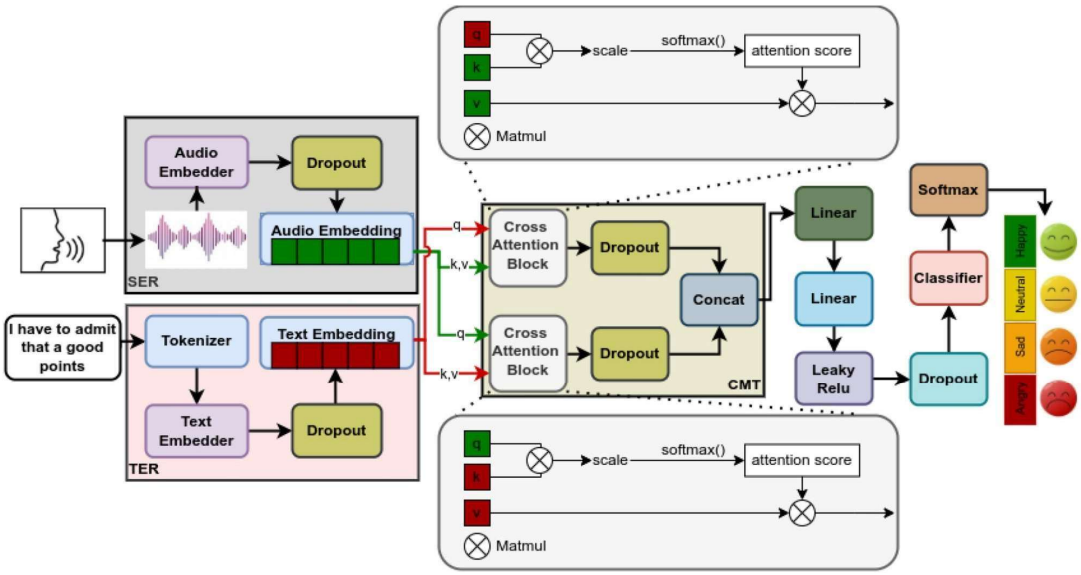

Fig. 2. *Cross-Modal Transformer (CMT) architecture for multimodal emotion recognition. Speech (SER) and text (TER) embeddings are fused via dual cross-attention blocks, followed by linear projection and softmax classification into emotion categories (Happy, Neutral, Sad, Angry).*

C. Neural Dialogue Systems

The neural models for conversation have developed dramatically through the use of encoders, and decoders in the past 10 years for generating dialogue [11], to more recently using hierarchical dialogue architecture models and large-scale neural networks to create conversational agents [12], and finally the introduction of DialoGPT demonstrates that large pretrained transformers can produce very high-quality responses in conversational formats [2]. The pipeline for chatbot dialogue management presented in Figure 3 is a standard example of dialogue management employed by the aforementioned recent research. The figure illustrates the process of intent analysis, context tracking, knowledge retrieval, and response generation used to manage conversations with chatbots.

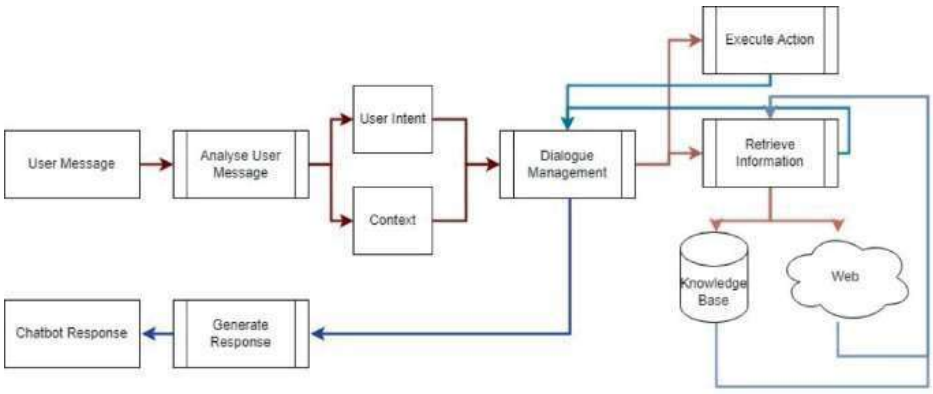

Fig. 3. *Standard chatbot dialogue management pipeline: user message analysis extracts intent and context, which drives dialogue management, knowledge base retrieval, and response generation.*

D. Empathetic Dialogue Models

Many studies have tried to put emotional awareness into dialogue systems, introducing the EmpatheticDialogues dataset so as to train models which are capable of generating responses that are emotionally supportive [4].

Other research proposed other empathetic neural dialogue models such as CAiRE and Mixture-of-Empathetic-Listeners (MoEL) [8][9]. These were able to improve response empathy but didnt explicitly model the dynamics of emotions with dialogue turns. Figure 4 shows an example deploying of a chatbot emotionally aware and integrated with another external platform (Slack), which combines emotion detection and machine translation as the client- side actions.

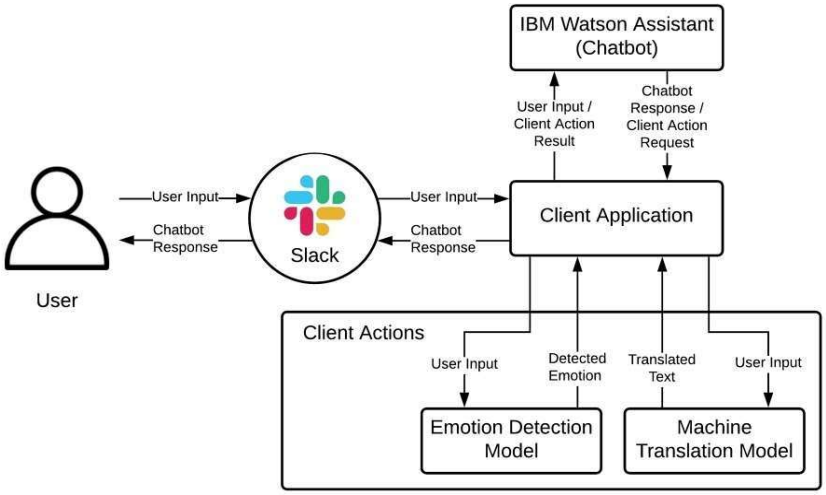

Fig. 4. *Architecture of an emotion-aware chatbot deployment integrating IBM Watson Assistant with Slack, including client-side emotion detection and machine translation modules [adapted from prior deployment study].*

IV. METHODOLOGY

A. System Architecture

The Emotional State Engine is a framework that has four primary parts (1) user input processing, (2) emotion detection, (3) the Emotional State Engine core, and (4) emotion-conditioned dialogue generation. This system receives user messages, recognizes emotional expressions, modulates internal emotional state, and produces responses which depend on the emotional state.

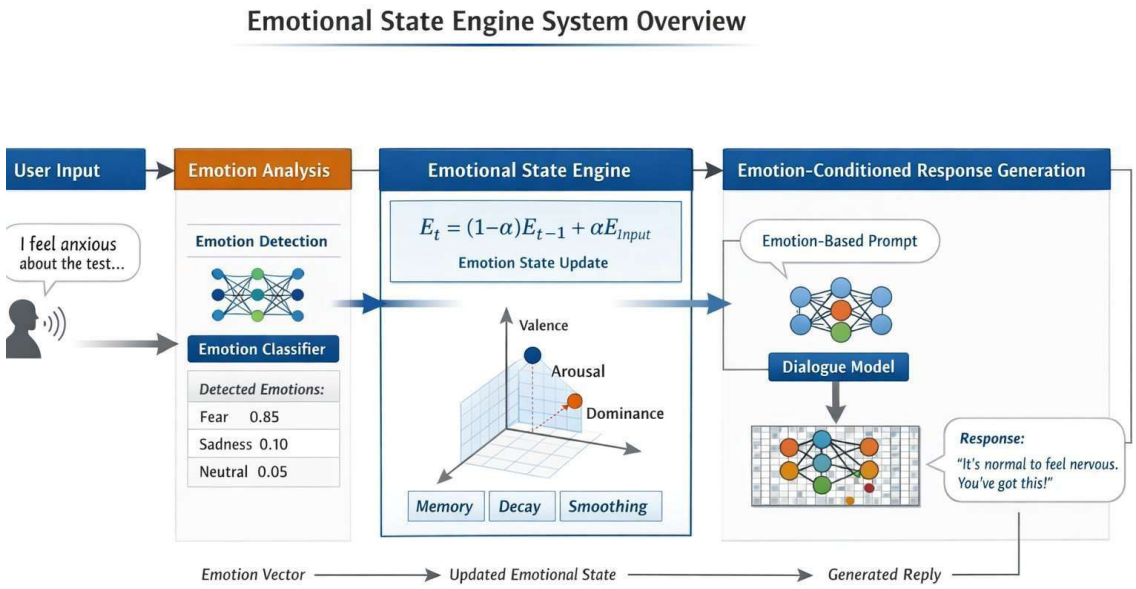

Fig. 5. *Engine System overview- Emotional State Engine. Emotion analysis module (detection + classification) is fed on with user input. The emotion state engine is updated with the resulting emotion vector through a weighted average equation which includes memory, decay and smoothing. The revised situation conditions the dialogue model in order to produce empathetic answers.*

B. Emotional Representation

The system displays emotional state using the **Valence-Arousal-Dominance (VAD)** model [3][14][15]. Each emotional state is shown as a continuous vector:

```
E = (v, a, d)
```

Dimension	Description
Valence (v)	Positive vs. negative affective quality of the emotion
Arousal (a)	Emotional intensity or activation level
Dominance (d)	Degree of control, confidence, or power

Table I: VAD Emotional Dimension Definitions

C. Emotional State Update Mechanism

In order to model emotional continuity, the Emotional State Engine updates the emotional state following the interaction with a weighted exponential smoothing formula:

```
E_t = (1 − α) × E_{t-1} + α × E_input
```

Variable	Meaning
E_t	Updated emotional state at turn t

E_{t-1}	Previous emotional state
E_input	Detected emotion vector from current message
α (alpha)	Emotional momentum coefficient ($0 < \alpha < 1$)

Table II: Emotional State Update Variables

This process facilitates changes in emotions without sudden shifts of emotional state as a result of temporary emotional cues. A low value of alpha results in slower adaptation of emotions (a higher memory), whereas a high value of alpha results in the system being more responsive to recent input.

D. Emotional State Engine Algorithm

The algorithm below illustrates how emotional states are able to be maintained and updated as the conversation takes place:

Step	Operation
1	Initialize emotional state: E_0 = (0, 0, 0)
2	Receive user message U_t
3	Apply emotion detection model to U_t
4	Extract emotion vector E_input from detection output
5	Update emotional state: $E_t = (1-\alpha)E_{t-1} + \alpha E_{input}$
6	Construct emotion-conditioned prompt from E_t
7	Send prompt to dialogue model (DialoGPT)
8	Generate response R_t
9	Output response to user
10	Repeat from Step 2 for next dialogue turn

Table III: Emotional State Engine Algorithm

DialoGPT provides response generation, which is a transformer model that has been pretrained on large-scale dialogue data [2]. The emotional state vector also affects the response generation and it is prompt conditioned such that the model would generate emotionally consistent responses.

E. Emotional Transition Matrix

The Emotional State Engine represents the evolution of emotional states between conversational turns and thus has an emotion transition representation:

Previous Emotion	Current Input Emotion	Resulting Emotional State
Neutral	Fear	Mild anxiety
Fear	Fear	Strong anxiety
Fear	Joy	Emotional recovery
Joy	Joy	Positive reinforcement
Joy	Anger	Emotional conflict
Anger	Calm	Emotional stabilization

Table IV: Emotional Transition Matrix

V. EXPERIMENTAL SETUP

An experimental conversational system based on the Emotional State Engine was created in Python and NLP libraries based on transformers. The system implemented DialoGPT as the dialogue model and the emotion classifier VAD extraction was pre-trained.

Simulated conversational scenarios including: were used to test the system.

- Anxiety situation (test stress and exam preparation)
- Happiness (personal success and good news) scenario.
- Frustration situation (technical problems and misunderstanding after misunderstanding)
- General exchange of information (neutral conversation)

To note transitions in emotional state and confirm emotional continuity, each scenario entailed several conversational turns. The momentum coefficient 0.3 was selected as a compromise between emotional memory and quick adaptation.

VI. EVALUATION METRICS

There were three evaluation criteria used:

1. Emotional Consistency: consistency between the emotion elicited by a user and the tone/content of the generated response.

2. Emotional Continuity: continuousness of the development of the emotional state of the dialogue in turns, assessed by the course of VAD values.

3. Response Latency: computing efficiency of end-to-end response generation pipeline.

VII. RESULTS

Experiments show that the Emotional State Engine supports consistent emotional representations during a conversation. The emotional update mechanism enables the system to monitor eventual change in emotional condition, and not respond to separate messages.

Figure 6 provides an illustration of how the valence dimension changed in four conversational turns in the anxiety scenario. The negative valence (-0.30) decreases slowly as the system supports empathetically the system reaching a positive valence +0.25 at turn 4.

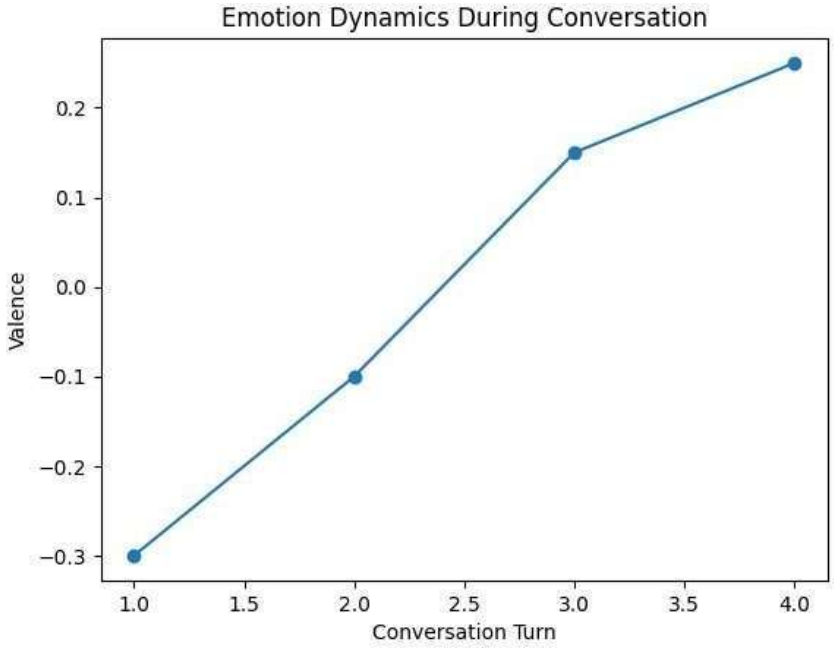

Fig. 6. *Dynamics of emotion in conversation: valence trajectory over four dialogue meetings in the anxiety scenario. Through empathetic response conditioning, the ESE gradually attains positive (0.25) valence as opposed to negative (-0.30) valence.*

Figure 7 shows the full VAD trajectory in five conversational turns as per the anxiety scenario. Arousal goes down as the system provides reassurance, yet valence and dominance increase. This illustrates emotional recovery and also growing the confidence of the user.

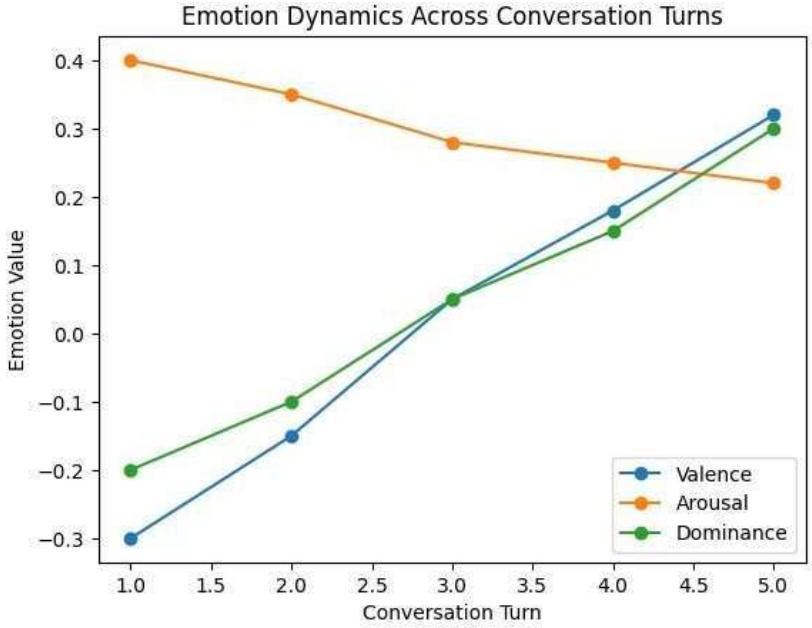

Fig. 7. *Emotion dynamics across conversation turns: trajectories of Valence, Arousal, and Dominance (VAD) over five dialogue turns. Valence and Dominance increase while Arousal decreases, consistent with successful empathetic engagement.*

The generation of emotion-conditioned response generated responses that were more consistent with the emotional context than the prompts generated with baseline conditions. The emotional state shifted toward confidence-building language (You're well-prepared, you've got this!), with the system switching to crisis-acknowledgment language (It's normal to feel nervous) in the course of the anxiety situation.

VIII. DISCUSSION

The proposed framework proves the idea that the dynamic approach to the modeling of emotional state leads to conversational coherence. In contrast to the conventional emotion-conscious dialogue models which are based on singly-turn emotion classification, the Emotional State Engine allows continuity of emotion by its temporal update process.

The style is consistent with psychological models of emotional processes, especially the circumplex model of affect [3], which considers emotions as dimensional instead of discrete. In VAD space, the ESE is inherently sensitive to the subtleties of changes between emotional states.

The exponential smoothing update formula gives a principled emotional memory update mechanism. The momentum coefficient α can be adjusted to suit various conversational situations: when therapeutic conversation is slow, the momentum coefficient should be low; when the conversation is fast and involves a task, it should be high.

IX. LIMITATIONS

This implementation currently has a number of limitations. First, emotional recognition is based entirely on textual cues, which do not necessarily show emotional expressions in speech prosody, facial expression and physiological indicators.

Second, simulated conversations were used in evaluations as opposed to large scale studies of human interaction. There is no crowd-sourced or clinical human evaluation of the results, which restricts generalizability of the reported results.

Third, the emotion reactivity coefficient of the fixed alpha does not adapt to individual differences in users. Individual learned momentum coefficient would enhance accuracy of the system.

X. FUTURE WORK

- The direction of future work will be into a number of directions:

Multimodal emotion recognition that combines speech, text, and facial expression signals in order to estimate the emotional state more richly.

- Long-term emotional memory in a series of conversation sessions.
- Empathy optimization learning through reinforcers of user feedback.
- Learning of adaptive momentum coefficients depending on personal user emotional profile.
- Massive testing of people to confirm emotional consistency and user satisfaction.

XI. CONCLUSION

The paper has presented the Emotional State Engine, which is a framework that is aimed at modeling the emotional dynamics of conversational AI systems. The proposed system allows conversational agents to sustain emotional continuity by dynamically modeling emotional states, through the Valence- Arousal-Dominance representation, and generating emotion-conditioned dialogue by integrating emotion detection and emotion-conditioned dialogue generation.

The temporal state update mechanism offers an intellectual method of emotional memory that does not face the shortcomings of the static theories of emotion determination per utterance. The ESE is experimentally confirmed to produce emotionally consistent response trajectories on multi-turn conversations.

The Emotional State Engine is a step in the right direction towards more emotional and empathetic interactive systems, filling the gap between emotion recognition and conversational AI.

REFERENCES

[1] R. W. Picard, Affective Computing. Cambridge, MA: MIT Press, 1997.

[2] Y. Zhang, S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and B. Dolan, "DialoGPT: Large-scale generative pre-training for conversational response generation," in Proc. ACL, 2020, pp. 270–278.

[3] J. A. Russell, "A circumplex model of affect," J. Personality Social Psychol., vol. 39, no. 6, pp. 1161–1178, Dec. 1980.

[4] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," in Proc. ACL, 2019, pp. 5370–5381.

[5] S. Mohammad and P. Turney, "Crowdsourcing a word-emotion association lexicon," Comput. Intell., vol. 29, no. 3, pp. 436–465, 2013.

[6] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, Eds., A Practical Guide to Sentiment Analysis. Cham, Switzerland: Springer, 2017.

[7] V. Sethu, E. M. Provost, J. Epps, C. Busso, N. Cummins, and S. S. Narayanan, "The ambiguous world of emotion representation," arXiv preprint arXiv:1909.00360, 2019.

[8] Z. Lin, A. Xu, G. I. Winata, F. A. Cahyawijaya, Z. Liu, B. Xu, and P. Xu, "CAiRE: An end-to-end empathetic chatbot," in Proc. AAAI, 2020, pp. 13622–13623.

[9] Z. Lin, A. Madotto, J. Shin, P. Xu, and P. Fung, "MoEL: Mixture of empathetic listeners," in Proc. EMNLP-IJCNLP, 2019, pp. 121–132.

[10] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[11] I. V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues," in Proc. AAAI, 2017, pp. 3295–3301.

[12] H. Chen, X. Liu, D. Yin, and J. Tang, "A survey on dialogue systems: Recent advances and new frontiers," ACM SIGKDD Explor. Newsl., vol. 19, no. 2, pp. 25–35, Nov. 2017.

[13] P. Ekman, "An argument for basic emotions," Cognition Emotion, vol. 6, no. 3–4, pp. 169–200, 1992.

[14] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," Current Psychol., vol. 14, no. 4, pp. 261–292, 1996.

[15] A. Mehrabian and J. A. Russell, An Approach to Environmental Psychology. Cambridge, MA: MIT Press, 1974.

[16] B. Kort, R. Reilly, and R. W. Picard, "An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a learning companion," in Proc. IEEE ICALT, 2001, pp. 43–46.

[17] S. Roller, E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, E. M. Smith, Y.-L. Boureau, and J. Weston, "Recipes for building an open-domain chatbot," in Proc. EACL, 2021, pp. 300–325.

[18] M. Sap, R. Le Bras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, and Y. Choi, "ATOMIC: An atlas of machine commonsense for if-then reasoning," in Proc. AAAI, 2019, pp. 3027–3035.

[19] T. Wolf, V. Sanh, J. Chaumond, and C. Delangue, "TransferTransfo: A transfer learning approach for neural network based conversational agents," arXiv:1901.08149, 2019.

[20] Q. Jiang and Y. Huang, "Affective computing: Review and prospects," J. Intell. Fuzzy Syst., vol. 44, no. 4, pp. 5995–6009, 2023.