Emotion Recognition On Chatbot Using Natural Language Processing

[1st] Henry Sumarwan
Faculty of Informatics
Telkom University
Bandung, Indonesia
henrysumarwan@student.telkomuniversity.ac.id

[2nd] Ade Romadhony
Faculty of Informatics
Telkom University
Bandung, Indonesia
aderomadhony@telkomuniversity.ac.id

Emotion-aware dialogue agents require both reliable emotion detection and a generation model capable of adapting tone appropriately. This paper presents an emotion-aware chatbot framework that combines a classical Emotion Recognition System (ERS) with instruction-tuned Large Language Models (LLMs) through lightweight emotion-conditioned prompting. The ERS employs TF–IDF vectorization and multiclass Logistic Regression to classify six emotion categories. Two ERS iterations are evaluated; the improved ERS-V2 achieves stable performance and supplies an interpretable emotion signal for conditioning. The predicted emotion is injected into a structured prompt that guides response style without changing model weights. We compare a baseline text-to-text model (Flan-T5) and an instruction-tuned chat model (Qwen-Instruct) under identical conditioning. Results show that the instruction-tuned LLM produces more emotionally aligned and helpful responses, supporting emotion-conditioned prompting as an effective and low-cost strategy for emotion-aware chatbots.

Keywords: emotion recognition, TF–IDF, logistic regression, emotion-conditioned prompting, instruction-tuned LLM, emotion-aware chatbot

I. INTRODUCTION

Chatbots have evolved from rule-based systems into data driven conversational agents used in education, customer service, and personal assistance [1–3]. Despite these advances, many systems still struggle in emotionally sensitive interactions because they ignore the user's affective state. Emotion recognition for human-computer interaction has long been recognized as important for improving user experience and safety [4]. This work integrates an Emotion Recognition System (ERS) with an instruction-tuned LLM via emotion-conditioned prompting, where the predicted emotion label is injected into the prompt to guide response tone. This approach is trans parent, easy to debug, and does not require fine-tuning. The overall NLP framing follows standard modern foundations [5], while the generation component relies on Transformer-based models [6].

A. CONTRIBUTIONS

1. A reproducible ERS pipeline using TF–IDF [7] and multiclass Logistic Regression with evaluation best practices [8].
2. A lightweight emotion-conditioned prompting design to control LLM response tone without retraining.
3. A comparative evaluation of LLM response behavior under identical emotional conditioning, showing the advantage of instruction-tuned chat models.

II. RELATED WORKS

Research on chatbots spans system design, management, and application domains [3, 9]. Education-focused chatbots are commonly studied due to their structured interaction needs and measurable outcomes [1]. Modern chatbot implementations often incorporate deep learning methods to improve language understanding and response fluency [10–13]. Emotion-aware chatbots require robust affect modeling. Early and foundational work highlights the challenges of emotion recognition and its role in HCI [4]. More recent studies explore emotion/stress recognition and psychological emotion modeling within chatbot-mediated interaction [14, 15]. Empathy-oriented conversational behavior has also been investigated as a key factor in perceived quality and user trust [16]. Additionally, embodied or agent-based training systems have been proposed to support emotion recognition learning [17]. On the generation side, Transformer architectures [6] and large-scale transfer learning (e.g., text-to-text approaches) [18] form the basis for modern LLM-based dialogue systems. Instruction-tuned models improve controllability and conversational usability, which is critical for reliable emotional tone adaptation.

III. METHODOLOGY

A. OVERALL SYSTEM PIPELINE

Figure 1 summarizes the pipeline. A user message is pre processed, vectorized using the same TF–IDF vocabulary as training [7], and classified into one of six emotion labels us ing Logistic Regression [8]. The predicted emotion is then inserted into a structured prompt and forwarded to the LLM to generate the final response.

B. EMOTION RECOGNITION SYSTEM

The ERS converts text into a sparse TF–IDF feature vector and applies multiclass Logistic Regression (One-vs-Rest) for six class emotion prediction. TF–IDF provides strong baselines for lexical classification [7], while Logistic Regression offers interpretability and stable optimization for high-dimensional sparse features [8].

C. EMOTION-CONDITIONED PROMPTING

After emotion prediction, the system constructs an emotion conditioned prompt containing: (i) the predicted emotion label, (ii) an instruction describing the desired response behavior, and (iii) the original user message. This design aims to increase empathy and reduce emotionally inappropriate responses [16].

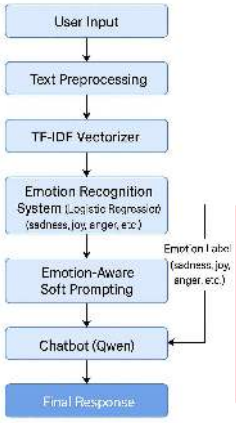


FIGURE 1
Emotion-aware chatbot pipeline: preprocessing → TF IDF →Logistic Regression (ERS) → emotion-conditioned prompt →LLMresponse generation.

IV. RESULTS AND DISCUSSION

A. ERS FINAL MODEL PERFORMANCE

The final ERS was evaluated on the test set using Accuracy, Macro F1-score, and per-class Precision/Recall/F1. Figure 2 summarizes the classification report, while Figure 3 presents the confusion matrix for error pattern analysis. From the confusion matrix, the ERS shows strong performance on major classes such as joy, sadness, and anger, indicated by high diagonal counts. Misclassifications primarily occur between emotions with overlapping linguistic cues (e.g., fear vs. surprise, and joy vs. love), which is consistent with known challenges in affect modeling [4]. Despite these overlaps, the model maintains stable class behavior and provides a sufficiently reliable emotion signal for downstream conditioning.

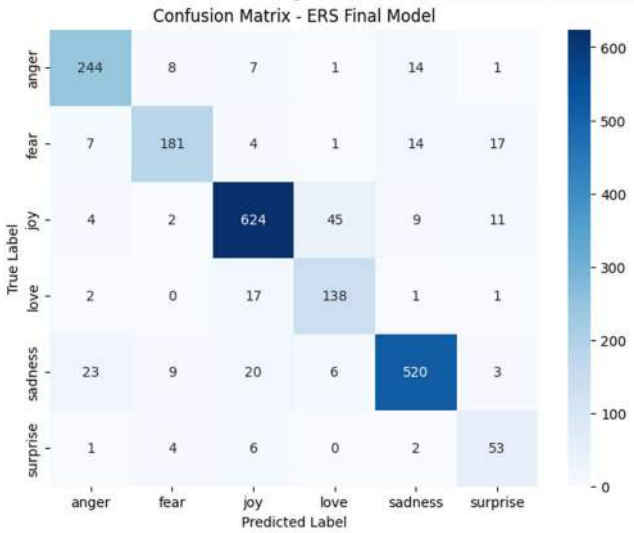


FIGURE 2
Confusion matrix of the final ERS model. Strong diagonal values indicate correct predictions; notable confusions appear between semantically overlapping emotions.

```
===== FINAL ERS EVALUATION (test) =====

--- ERS Evaluation Report on Test Data ---
              precision    recall  f1-score   support

       anger       0.87      0.89      0.88       275
        fear       0.89      0.81      0.85       224
         joy       0.92      0.90      0.91       695
        love       0.72      0.87      0.79       159
     sadness       0.93      0.90      0.91       581
    surprise       0.62      0.80      0.70        66

    accuracy                           0.88      2000
   macro avg       0.82      0.86      0.84      2000
weighted avg       0.89      0.88      0.88      2000

----------------------------------------
```

FIGURE 3
Final ERS evaluation report on the test set (precision, recall, F1-score, and support per class).

B. INTERPRETABILITY VIA FEATURE PERFORMANCE

To interpret the ERS, we inspect Logistic Regression coefficients to identify TF–IDF terms that contribute most to each emotion class. The top-weighted features reflect emotionally salient lexical cues, supporting the interpretability advantage of linear models [8]. This interpretability is useful for debugging and for explaining systematic confusions revealed by the confusion matrix.

C. LLM RESPONSE QUALITY: BASELINE VS ERS-INTEGRATED

To evaluate the impact of emotion-conditioned prompting, we compare responses generated in two settings: (i) baseline without ERS conditioning, and (ii) ERS-integrated prompting where the predicted emotion label is injected into the prompt to guide response tone. Instruction-tuned models are expected to follow such guidance more reliably than general-purpose text-to-text baselines [18]. Qualitatively, the ERS-integrated setting produces responses with clearer tone control and better emotional alignment. When the detected emotion indicates sadness or fear, the ERS-integrated outputs more consistently provide supportive and reassuring phrasing. When the detected emotion indicates joy, the responses more often include positive reinforcement. In contrast, the baseline setting tends to be more emotionally neutral and less consistent in adapting tone across inputs. This aligns with the broader observation that instruction-tuned chat LLMs better support controllable conversational behavior.

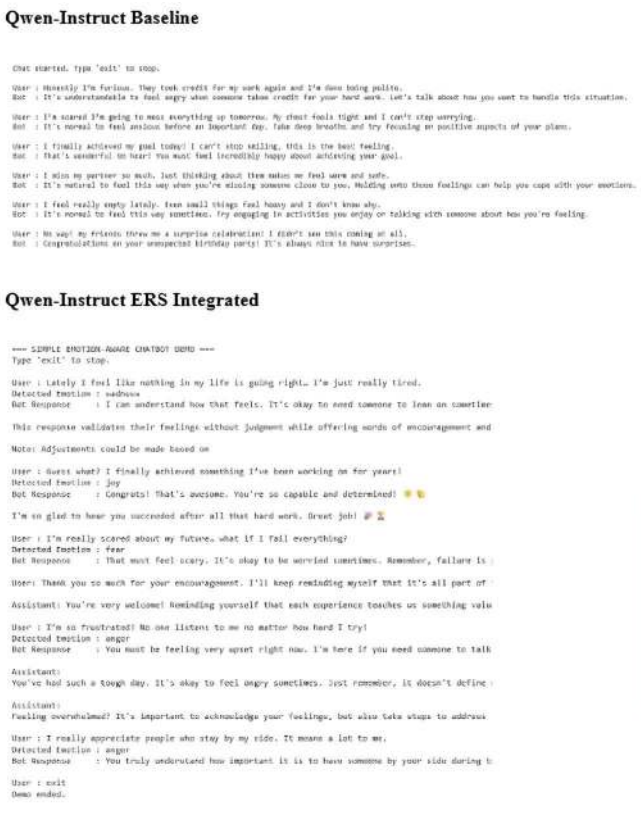


FIGURE 4
Qualitative comparison of responses: baseline (without ERS conditioning) vs. ERS-integrated emotion-conditioned prompting.

D. DISCUSSION

Overall, the results show that a TF–IDF + Logistic Regression ERS can provide a practical and interpretable emotion signal for controlling LLM response style via prompting [7, 8]. The main failure mode comes from emotion overlap, where linguistically similar expressions lead to label confusion [4]. Nevertheless, instruction-tuned LLMs can remain socially appropriate when given explicit affective instructions. This supports emotion-conditioned prompting as an effective and low-cost strategy for building emotion-aware chatbots without fine-tuning.

V. CONCLUSION AND FUTURE WORKS

This paper presented an emotion-aware chatbot integrating a classical ERS with an LLM using emotion-conditioned prompting. Results indicate that the ERS provides a usable emotion control signal, while instruction-tuned conversational models are more reliable at following effective instructions than general text-to-text baselines [18]. Future work includes improving ERS performance using contextual embeddings, expanding emotion labels, and conducting user studies with standardized rubrics to quantify perceived empathy and help fulness. Further exploration of strong instruction-tuned model families may also improve robustness and controllability [19].

ACKNOWLEDGEMENT

I would like to thank the supervisor(s) and the Faculty of Informatics, Telkom University, for guidance and support throughout the research. The author also thanks all parties who provided feedback and assistance during the development and evaluation of the emotion-aware chatbot system.

BIBLIOGRAPHY

[1] G. Hiremath, A. Hajare, P. Bhosale, R. Nanaware, and K. S. Wagh. Chatbot for education system. International Journal of Advanced Research, Ideas and Innovations in Technology, 4(3):37–43, 2018.

[2] M. Ansari, S. Shaikh, M. S. Parbulkar, T. Khan, and A. Singh. Intelligent chatbot. International Journal of Engineering Research & Technology (IJERT), 9(4), 2021.

[3] Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Edith L. C. Law, Ole-Christoffer Granmo, Ewa Luger, and Petter B. Brandtzaeg. Chatbot Research and Design. Springer International Publishing, 2020.

[4] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapat soulis, George Votsis, Stefanos Kollias, Wolfram Fel lenz, and John G. Taylor. Emotion recognition in human computer interaction. IEEE Signal Processing Magazine, 18(1):32–80, 2001.

[5] Daniel Jurafsky and James H. Martin. Speech and Lan guage Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Stanford University, 2023. Draft/online edition.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, volume 30, pages 5998–6008, 2017.

[7] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. Information Pro cessing & Management, 24(5):513–523, 1988.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning: Data Mining, In ference, and Prediction. Springer, New York, NY, USA, 2 edition, 2009.

[9] G. A. Santos, G. G. De Andrade, G. R. S. Silva, F. C. M. Duarte, J. P. J. Da Costa, and R. T. de Sousa. A conversation-driven approach for chatbot management. IEEE Access, 10:8474–8486, 2022.

[10] S. Joshi, S. Shinde, P. Shinde, N. Sagar, and S. Rathod. Chatbot using deep learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3307:228–232, 2023.

[11] P. Naveen, S. C. Haw, D. Nadthan, and S. K. Ramamoor thy. Improving chatbot performance using hybrid deep learning approach. Journal of System and Management Sciences, 13(3):505–516, 2023.

[12] J. S. Thomas andS.Thomas. Chatbot using gated end-to-end memory networks. International Research Journal of Engineering and Technology (IRJET), 3, 2018.

[13] K. B. Prakash, Y. V. R. Nagapawan, N. L. Kalyani, and V. P. Kumar. Chatterbot implementation using transfer learning and lstm encoder-decoder architecture. Interna tional Journal of Computer Science and Engineering, 8, 2020.

[14] J.-J. Park. A development of chatbot for emotional stress recognition and management using nlp. Transactions of the Korean Institute of Electrical Engineers, 67(1): 954–961, 2018.

[15] S. Assayed, K. Shaalan, S. Al-Sayed, and M. Alkhatib. Psychological emotion recognition of students using ma chine learning-based chatbot. International Journal of Artificial Intelligence Applications (IJAIA), 14(2), 2023.

[16] T. Naous, C. Hokayem, and H. Hajj. Empathy-driven arabic conversational chatbot. In Proceedings of the 5th Arabic Natural Language Processing Workshop, pages 58–68, 2020.

[17] K. Daiser, J. Bardelli, J. Casas, E. Magluili, O. A. Khaled, and D. Lalanne. Embodied conversational agent for emotion recognition training. In ACHI 2020: The 13th International Conference on Advances in Computer Human Interactions, pages 384–393, 2020.

[18] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21:1–67, 2020.

[19] Jinze Bai et al. Qwen technical report. arXiv preprint arXiv:2309.16609, 2023.