Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.33965

Research and Analysis of Empathetic Dialogue Generation Method Based on Reinforcement Learning to Improve the Balance Between Emotion and Semantics

Zhixuan Li

XIPU International Business School, Xi'an Jiaotong-Liverpool University, Suzhou, China
Zhixuan.Li24@student.xjtlu.edu.cn

Abstract. With the advancement of AI, empathetic dialogue systems have made great progress, recognizing user feelings and providing responses that are suitable in different situations, which are becoming increasingly valuable. Nevertheless, there are still difficulties in specific areas, including mental health services and smart customer care. The main current limitations are the insufficient depth of emotional understanding in traditional AI, the inability to precisely detect the differences in the intensity of emotions, the inability to dynamically balance emotional responses to contextual semantics, and the inability to sustain emotional resonance in long-term dialogue. Moreover, the current approaches tend to use a single, fixed-weight reward system, which fails to meet the dynamics of the dialogue context. In the meantime, the vast majority of the models use single-turn responses as their optimization goal and ignore the role of long-term dependencies in the dialogue history in determining empathetic effectiveness. To address these challenges, this project suggests a machine learning-based empathetic dialogue generation framework that uses reinforcement learning. Using the "reward-penalty" mechanism and dynamic adjustment of weights to various goals, the system constructs a variety of reward signals to steer models towards sustained dialogue interactions instead of single-response patterns, adapting to various conversational scenarios and emotional variations.

Keywords: Empathetic dialogue generation, reinforcement learning, affective computing, dynamic weight adjustment, multi-objective optimization

1. Introduction

The initial AI-based systems were mostly based on a rule based question answering system. As large language models continue to develop and new areas of conversation are explored, the conversational intelligence capabilities of AI have greatly enhanced. Nowadays, AI-driven chatbots are used extensively in mental health assistance, virtual customer services, and intelligent assistants. Nevertheless, with the increased complexity of application scenarios, the ability to empathise within conversational systems has become increasingly important to provide more effective user experiences and interactions [1, 2]. No longer are user expectations of intelligent dialogue systems limited to mere question answering, but now they expect AI systems to be able to detect the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.33965

emotional undertones in a conversation, to demonstrate true empathy, and to be able to respond emotionally when required.

Psychologically, empathy refers to the capacity to understand others' emotions and to react to them; it has two facets: affective empathy (the capacity to experience the emotional conditions) and cognitive empathy (the capacity to interpret the circumstances and give appropriate responses) [3, 4]. Empathy has been an important factor in human-computer interaction in improving user experience and creating trust in systems. Empathy-driven design has been found to not only enhance user satisfaction but also user loyalty [5].

The research value of empathic dialogue generation can be exhibited in theoretical, technological, and practical aspects. In theory, it would improve the 'understanding of researchers on the emotional modules of artificial intelligence [6]. In terms of technology, this can be achieved through solving problems like real-time emotion recognition and adaptable dialogue strategies that can push the boundaries of the development of conversational systems further [7]. In practice, empathetic dialogue systems have been extensively used in mental health support, elder care [8], intelligent customer service, and education. The detailed study of these systems can promote the advancement of related areas. Nevertheless, the available models of empathetic dialogue continue to have such problems as the lack of emotional depth and the problem of balancing emotions and content expression. The existing AI-based emotion understanding is mostly based on the categorical classification (positive/negative/neutral), which does not distinguish the changes in the intensity of emotions and does not properly combine the emotions with cognitive processes. This leads to inefficient responses, for instance, AI systems might not be able to give emotionally warm responses that fit users semantic perceptions. Additionally, the traditional reinforcement learning and topic modeling methods find it difficult to balance between pure semantics processing and subtle emotional dynamics [4]. The existing methodologies do not have the ability to incorporate the dynamic context or resonate on the same emotional tone and, therefore, produce inconsistent or inappropriate responses [9].

The earliest studies of empathetic dialogue generation can be traced to the Empathetic Dialogues dataset released by Rashkin et al., which has been used as the reference point in most further studies. Many methodologies have been suggested based on this dataset. The emotion-labeling approach (MoEL) uses independent expert decoders on 32 emotions and combines them dynamically based on user distributions on emotions to produce more accurate empathetic dialogue generation, but it lacks adequate modeling of the intensity of emotional expressions [10]. The common-sense-based approach (CACKS) is a dynamically selected, context-aligned common-sense knowledge that improves cognitive consistency and emotional relevance [9]. The approach is based on the psychological theory (PERM), which measures the level of empathy in both directions, between the supporters 'and help-seeking users' [11]. The most applicable approach to this project was the reinforcement learning-based approach (RLBA), which was the first to quantify the level of emotion in empathetic dialogue generation (between 1 and 7), develop a two-dimensional reward function (emotional level alignment and semantic relevance), and optimize it with the PPO algorithm. The experimental findings reveal that RLES has a high performance compared to baselines like SFT and MoEL [4]. Nonetheless, the approach still cannot address such questions as fixed weights 'inability to adjust to different dialogue situations, single-round optimization tendencies to ignore long-term consequences, and performance in different emotional situations does not receive an analytical treatment. To tackle these issues, this paper presents four progressive experiments to assess the effectiveness of integrating reinforcement learning and dynamic multi-reward empathy dialogue

DOI: 10.54254/2755-2721/2026.33965

mechanisms in enhancing the balance between emotional and semantic aspects of empathetic dialogue generation.

2. Manuscript preparation

2.1. Theoretical basis for empathetic dialogue generation

Empathic dialogue is aimed at allowing dialogue systems to identify user emotions and produce emotionally corresponding responses. Its mainly aims at attaining emotional consistency and semantic relevance. In psychological studies, empathy has two dimensions: affective and cognitive, which are associated with the aims of emotional matching and semantic understanding, respectively.

2.2. Reinforcement learning

Reinforcement learning is mainly based on a 'reward-penalty' system in order to allow AI to learn by trial and error. This method is especially applicable in the optimization of multi-objective dialogue tasks, the key elements of which are: agent, environment, state, action, and reward.

One of the most popular reinforcement learning algorithms is the Proximal Policy Optimization (PPO) algorithm. The continuous refinement of the policy of the model allows the model to attain better rewards in the execution of the task, thus enhancing its decision-making abilities in its interactions with the environment.

3. Research methods

3.1. Experimental design

Based on the aforementioned research questions, the study proposes four core hypotheses: (H1) A multi-reward mechanism integrating emotional feedback and semantic evaluation demonstrates superior performance in enhancing empathetic dialogue generation quality compared to single reward signals; (H2) In reinforcement learning optimization objectives, focusing on long-term dialogue accumulation proves more effective than prioritizing immediate rewards in current rounds for improving sustained empathetic dialogue quality; (H3) Dynamic weight adjustment configurations exhibit greater contextual adaptability than fixed weight settings; (H4) The proposed method is expected to maintain robust empathetic performance across diverse dialogue scenarios with varying emotional intensity and types.

This study involves four comparative models (Table 1). The dataset employed is the Empathetic Dialogues dataset, published on GitHub by Rashkin et al. for their empathy-driven dialogue generation research, and was divided into a training set (80%), a validation set (10%), and a test set (10%).

Table 1. Main models

model	Base model	act on
SFT	GPT-2	Verify the necessity of RL
MoEL	Transformer	Validation of RL advantages
RLES	T5	Verify dynamic weight advantage
This method	GPT-2	Core methodology of this study

DOI: 10.54254/2755-2721/2026.33965

3.2. Overall framework

To rigorously establish the effectiveness of various modules, this project developed four consecutive experiments to answer four fundamental questions: First, whether various signals of rewards can positively influence the quality of empathetic dialogue in comparison with a single reward. Second, whether cumulative rewards emphasizing long-term conversations are better than instant rewards emphasizing single conversations. Third, can dynamic weight adjustment show greater adaptability to changes in conversational scenarios than fixed weights. Lastly, whether the proposed method can be consistently performed in various emotional conditions.

3.3. Experimental content

Experiment 1 produced responses to all the models on the test set and computed six metrics (emotional difference, BERTScore, BLEU, PPL, Distinct-1↑, Distinct-2↑) to confirm that the method is comprehensive and outperforms the baseline on core metrics.

Experiment 2 trained four variants using the same base model to confirm the need for multi-reward signals and the effectiveness of dynamic weights. The four variants are designed in a specific way, which is presented in Table 2.

Table 2. Four variants

variant	Emotional rewards	Content rewards	dynamic weight	explain
V1 Full Version	√	√	√	This method
V2 only emotion	√	❌	√	Optimize emotions only
V3 content only	❌	√	√	Optimize content only
V4 fixed weight	√	√	❌	Fixed weight α=0.5

Experiment 3 evaluated two variants—timely optimization and long-term optimization—to determine whether the long-term cumulative reward incorporating dialogue history (γ=0.95) outperforms standalone independent optimization (γ=0).

Experiment 4 divided the samples into three cases according to the emotional intensity and BERTScore of the preceding rounds. In each situation, comparative experiments with dynamic and fixed weights were performed to show that dynamic weights work better than fixed weights in various conversational circumstances.

3.4. Primary evaluation outcomes

To compare and evaluate the performance of various models in the different dimensions, the main evaluation metrics that were set up during the experiment were the affective difference, BERTScore, BLEU, PPL, and Distinct (Table 3).

Table 3. Five experimental evaluation indicators

metric	Measurement dimension	explain
Emotional differences	Emotional Compatibility Score	The normalized value of emotional intensity difference between user and response, the smaller the better.

DOI: 10.54254/2755-2721/2026.33965

Table 3. (continued)

BERTScore	Content relevance	For BERT-based semantic similarity calculation, the higher the value, the better.
BLEU	n-gram overlap degree	The higher the lexical match with the standard response, the better.
PPL	Language fluency	The smaller the confusion level, the more natural the language.
Distinct	lexical diversity	The only n-gram ratio; a higher value indicates richer vocabulary.

4. Experimental results

4.1. Experiment 1

Table 4. Comparison of model performance effects

model	Emotional differences ↓	BERTScore↑	BLEU↑	PPL↓	Distinct-1↑	Distinct-2↑
SFT	0.6899	0.2459	0.0924	2.51	0.2523	0.6373
MoEL	0.6351	0.2489	0.0970	2.66	0.2769	0.6722
RLES	0.6118	0.2134	0.0870	5.19	0.3025	0.7396
This method	0.6287	0.2219	0.0857	5.41	0.2813	0.7178

The results of the experiment (Table 4) indicate that the RL-based solution (RLES + the approach of the paper) performs significantly better than non-RL methods (SFT + MoEL): RLES yields a 11.3% lower error rate than SFT, and the approach of the paper yields 8.9% higher than SFT, which proves the usefulness of the RL training in boosting the quality of the empathetic dialogue. Moreover, BLEU measure was optimized with smoothing methods in experiments to ensure that all the models are within the 0.086-0.097 range, which is a standard range where the performance of open-domain dialogue is measured.

When this method is compared to RLES, the emotional difference score of this method is slightly greater than the score of RLES (0.629 and 0.612, respectively). Nevertheless, this strategy is better because it has better BERTScore (0.222 vs 0.213) and the mean response length (12.1 vs 11.1), which means that dynamic weights lead to a more balanced trade-off between emotional and content.

The high PPL (5.19-5.41) in the course of experiments is a standard RL training effect, which is caused by the model leaving local optima in SFT to venture into a wider range of representations.

4.2. Experiment 2

Table 5. Results of experiment 2

variant	feeling	content	trends	Emotional differences ↓	BERTScore↑	PPL↓	Distinct-2↑
V1 Full Version	√	√	√	0.6287	0.2219	5.41	0.7178
V2 only emotion	√	❌	√	0.6133	0.2177	5.30	0.7513
V3 Content Only	❌	√	√	0.5931	0.2189	5.24	0.7513
V4 Fixed Weight	√	√	❌	0.6172	0.1926	5.12	0.7731

DOI: 10.54254/2755-2721/2026.33965

The findings of Experiment 2 (Table 5) show that V4 (fixed weights) had the lowest BERTScore (0.193), which means that fixed weights do not dynamically distribute resources based on the needs of a given context, and the quality of content is compromised. V1 had the highest BERTScore (0.222), and its full version had the best BERTScore performance and reasonable emotional variance, which achieved the best overall balance. Although V3 demonstrated the least emotional variation (0.593), it was due to the style drift, as the RL training was oriented on the content rewards, but not on the improved emotional model. Lastly, that the BERTScore has improved by 15.2% between V1 (0.222) and V4 (0.193) confirms that dynamic weights retain the quality of content more effectively, which underscores their applicability.

4.3. Experiment 3

Table 6. Comparison of results between timely optimization and long-term optimization

variant	γ	Emotional differences ↓	BERTScore↑	PPL↓	Distinct-1↑
Instant Optimization	0.0	0.6120	0.2145	4.98	0.2946
Long-term optimization	0.95	0.6703	0.2196	5.41	0.2845

As shown in Table 6, the comparison between real-time optimization and the long-term optimization shows that the long-term optimization has a better performance in BERTScore (0.220 vs. 0.215), which shows that integrating future reward is much better than real-time optimization to generate semantically relevant responses. Real-time optimization has a less emotional divergence (0.612 vs. 0.670), indicating that it is an aggressive optimizer of single-round emotional matching. The long-term optimization elicits longer responses (12.7 words vs. 11.8 words), which means that the model has developed more conversational strategies. Nevertheless, because of the lack of long dialogues (≥5 rounds) in the test set, other measures, such as multi-round emotional total score and topic coherence, are not validated. Enlarging the test dataset may be useful in future validation.

4.4. Experiment 4

Table 7. Results of experiment 4

scene	trends emo↓	fixed emo↓	Dynamic advantage	trends BERT	fixed BERT
The user is sad (48)	0.6946	0.7582	↓8.4%	0.2264	0.1687
Regular Chat (15)	0.5868	0.5813	↑0.9%	0.2161	0.1567
Topic conversion (50)	0.8418	0.7430	↑13.3%	0.2219	0.1748

Experiment 4 mainly confirmed Hypotheses H3 and H4, as reflected in Table 7. Dynamic weights were shown to be more effective in emotional differentiation (↓8.4%) with larger BERTScore values (0.226 vs. 0.169, ↑34.3%) in the cases of user distress. In regular conversation, the difference between the emotions was comparatively balanced (gap of 0.9), but the dynamic weights were much more superior in BERTScore (0.216 vs. 0.157, ↑37.8). Dynamic weights were more emotionally differentiated (↑13.3%) and performed better on BERTScore (0.222 vs. 0.175, ↑27.0%) during topic transitions, although fixed weights still outperformed dynamic weights. Comparative analysis between such scenarios shows that dynamic weights always perform better than fixed weights in

DOI: 10.54254/2755-2721/2026.33965

BERTScore under all conditions (27%-38%), suggesting that dynamic adjustment is effective to protect the quality of content.

5. Conclusion

This paper validates the effectiveness of a dynamic multi-reward empathy dialogue generation method based on reinforcement learning using four progressive experiments. In particular, RL training exhibits much better performance in comparison to non-RL methods. Multi-reward signals demonstrate greater increases in content relevance in comparison to single-reward signals, and emotional and content coordination is essential. With dynamic weights, the relevance of content is always enhanced in all situations, which is better than fixed weights. Long-term optimization helps in improving the quality of content, and future rewards help in generating semantically relevant responses.

The multi-round emotional total score and topic coherence measures were not, however, completely validated due to a lack of long-sample dialogues and limited computational resources in this research. Future work improvements can be conducted by increasing training datasets, including more long-dialogue samples to ensure experimental integrity, and adding human evaluation methods.

References

[1] Beale, R., & Creed, C. (2009). Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies*, *67*(9), 755-776.

[2] Guo, S., & Ning, B. (2024). A Review of Empathetic Conversational Systems. 2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP), Intelligent Computing and Signal Processing (ICSP), 2024 9th International Conference On, 1279–1286. https: //doi.org/10.1109/ICSP62122.2024.10743383

[3] Yu, C. L., & Chou, T. L. (2018). A Dual Route Model of Empathy: A Neurobiological Prospective. Frontiers in Psychology, 9. https: //doi.org/10.3389/fpsyg.2018.02212

[4] Cheng, J., Jiang, Z., Chen, Z., & Han, D. (2026). Empathetic Response Generation via Reinforcement Learning with Empathy Level Alignment and Semantic Relevance. Symmetry (20738994), 18(1), 148. https: //doi.org/10.3390/sym18010148

[5] Ma, N., Khynevych, R., Hao, Y. Q. & Wang, Y. H., (2025). The impact of anthropomorphism and perceived intelligence in visual design of chatbot avatars on user experience: accounting for perceived empathy and trust. Frontiers in Computer Science, 7. https: //doi.org/10.3389/fcomp.2025.1531976 User loyalty

[6] Lee, Y.-C., Zhang, J., Song, T., & Tan, Y. (2026). Conversational AI for Social Good (CAI4SG): An Overview of Emerging Trends, Applications, and Challenges.

[7] Mai, K. Y., & Le, T. H. (2025). Talk to Me: A Preliminary Review on the Evolution and Impact of Emotional AI [TREO Poster Paper]. Proceedings of the Americas Conference on Information Systems (AMCIS) 2025. https: //aisel.aisnet.org/treos_amcis2025/194/

[8] Leema, P. B., & Sangaiah, A. K. (2025). Empathetic AI: Multimodal Interaction Framework for Emotional Resilience and Cognitive Engagement in Elderly Care. 2025 International Conference on Next Generation Information System Engineering (NGISE), Next Generation Information System Engineering (NGISE), 2025 International Conference On, 1, 1–6. https: //doi.org/10.1109/NGISE64126.2025.11085267

[9] Wang, Y., & Feng, J. (2025). CACKS: Context-Adaptive Commonsense Knowledge Selection for Empathetic Dialogue Generation. 2025 44th Chinese Control Conference (CCC), Chinese Control Conference (CCC), 2025 44th, 9052–9058. https: //doi.org/10.23919/CCC64809.2025.11178883

[10] Lin, Z., Madotto, A., Shin, J., Xu, P., & Fung, P. (2019). MoEL: Mixture of Empathetic Listeners. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 121–132, Hong Kong, China. Association for Computational Linguistics.

[11] Wang, C., Zheng, W., Zhang, Y., Zhu, F., Cheng, J., Xie, Y., Wang, W., & Feng, F. (2026). PERM: Psychology-grounded Empathetic Reward Modeling for Large Language Models.