From Empathy to Personalized Empathy: Adapting Empathetic Strategies to Individual Users

Anonymous ACL submission

Abstract

As Large Language Models (LLMs) are increasingly deployed in long-term interactions with users, empathy has become an increasingly important capability. However, existing research overlooks the influence of users' personality traits on empathetic strategies during long-term interactions. To address this gap, we introduce the task of personalized empathy, which focuses on adapting empathetic strategies according to users' personalized characteristics derived from history. To study and enhance this capability, we construct PersonaEmp, a personalized empathy dataset built from long-term user-AI interactions, featuring rich user histories, persona information, and empathy-seeking queries. We further propose PereGRM, a reward modeling framework that combines the empathy evaluation structure with dynamic evaluation criteria generation for fine-grained reward modeling. Experimental results across different settings and multiple judge models show that PereGRM consistently achieves the strongest performance improvements, indicating its effectiveness for enhancing personalized empathetic capabilities.

1 Introduction

Recent advancements in Large Language Models (LLMs) have substantially transformed human-AI interaction, enabling increasingly persistent and personalized long-term interactions with users (Park et al., 2023; Packer et al., 2023). In such interactions, users may seek not only task assistance but also emotional support, encouragement, and reflective feedback (Dong et al., 2025; Pataranutaporn et al., 2025). This makes empathy a critical capability for LLMs.

Existing studies on LLM empathy mainly focus on general capability in complex emotional and social scenarios. These works either benchmark (Sabour et al., 2024; Paech, 2023; Zhang et al., 2025a) or improve (Sun et al., 2021; Liu et al., 2023; Wang et al., 2025, 2026b) LLMs' ability to recognize users' emotions, understand social contexts, respond to emotional needs, and provide supportive or socially appropriate feedback.

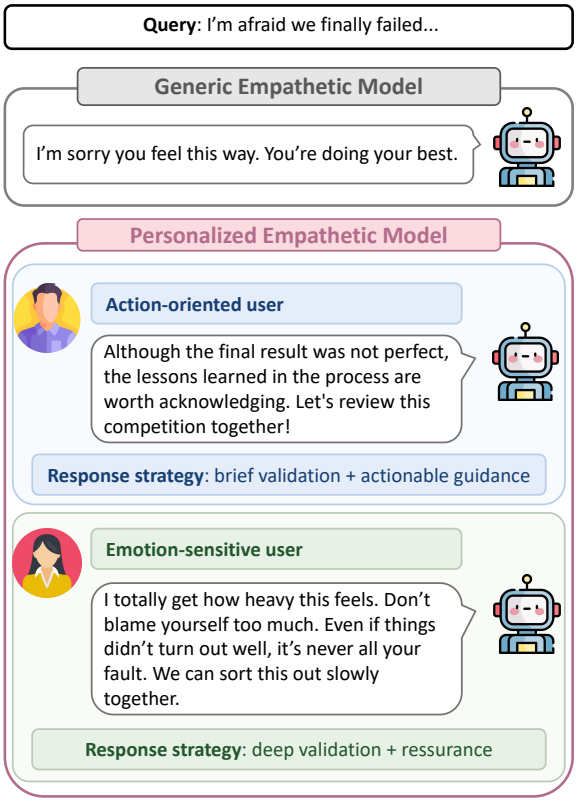

Figure 1: An example illustrating personalized empathy. Generic empathetic responses often fail to account for individual user differences. In contrast, personalized empathetic LLMs can adapt their response strategies according to personality traits, thereby providing more user-adaptive empathetic support.

However, earlier studies have paid limited attention to the inherently personalized nature of long-term empathetic interactions. Users with different personal experiences and personality traits may require fundamentally different support strategies even under similar emotional situations. As illustrated in Figure 1, generic empathetic responses

often fail to account for individual user differences, producing replies that are supportive but personally misaligned. This motivates the task of ***personalized empathy***: Given empathy-seeking queries, the goal is to tailor empathetic strategies to the user's distinctive characteristics (e.g., personality, emotional experiences, and psychological state) as derived from their long-term history, ensuring contextually appropriate over generic support.

To study and enhance LLMs' personalized empathy capabilities, it is essential to account for users' diverse long-term histories of empathetic interaction. To this end, we construct **PersonaEmp** (**Persona**-grounded **Emp**athy), a dataset derived from 90K long-term user-AI interactions in WildChat (Zhao et al., 2024). PersonaEmp encompasses rich user memories, detailed persona information, and authentic empathy-seeking scenarios. Specifically, we annotate users' long-term memory, identify high-quality users with comprehensive personalized profiles, and generate empathy-seeking queries that remain consistent with each user's original interaction context and linguistic style.

Beyond the dataset, we further explore methods to enhance LLMs' capabilities in the personalized empathy task. This setting introduces two critical challenges. First, optimizing personalized empathy requires more fine-grained and adaptive supervision across diverse users than generic empathy tasks. Fixed or user-agnostic objectives are therefore insufficient, highlighting the need for flexible optimization signals tailored to individual user contexts. Second, this task goes beyond conventional personalization: rather than merely matching user profiles or linguistic styles, it requires principled emotional understanding, effective supportive strategies, and socially appropriate responses. Thus, personalized empathy should be grounded in an empathy-specific framework instead of relying solely on traditional alignment techniques.

To address these challenges, we propose **PereGRM**, (**Per**sona-aware **e**mpathetic **G**enerative **R**eward **M**odeling) framework. PereRM incorporates user-specific information by dynamically generating assessment criteria grounded in psychology-informed empathy dimensions. Rather than relying on a fixed user-agnostic rubric, it derives evaluation criteria from each individual user's persona and interaction history, enabling more fine-grained assessment of whether a response is both empathetic and personally appropriate.

To evaluate the effectiveness of PereGRM, we construct both random and out-of-distribution splits based on user profiles to assess models on unseen user types. We assess the trained LLMs with multiple judge LLMs to examine whether the improvement generalizes beyond the training evaluator. Experimental results show that PereGRM achieves the strongest generalization performance, improving personalized empathy by around 20% on unseen users across different settings.

In summary, our contributions can be concluded as follows:

- We highlight the limitations of existing empathy-oriented LLM research in modeling users' personal experiences and personality traits within long-term interactions and introduce the task of personalized empathy.
- We construct PersonaEmp, a large-scale dataset for history-grounded personalized empathy built from long-term user-AI interactions, featuring rich interaction histories, user memories, and detailed persona profiles.
- We propose PereGRM, a persona-aware generative reward modeling framework, combining psychology-grounded empathy evaluation with dynamic generative reward modeling.
- Extensive evaluations show that PereGRM achieves the strongest performance among all baselines across different settings.

2 Preliminaries

In this section, we first formalize the personalized empathy problem (Section 2.1), and then review PERM as a psychology-grounded empathy evaluation framework (Section 2.2).

2.1 Task Formulation

Given the user's memory collection $\mathcal{M} = \{m_i\}_{i=1}^{n}$, where each m_i represents a memory item such as past interaction events or user preferences, and an empathy-seeking query q, the goal is to generate a personalized empathetic response y:

$$y = \pi_\theta(\mathcal{M}, q), \tag{1}$$

where π_θ denotes the policy LLM.

2.2 PERM for Empathy Reward Modeling

PERM (Wang et al., 2026b) is a multi-perspective and multi-dimensional empathy reward modeling framework grounded in the Empathy Cy-

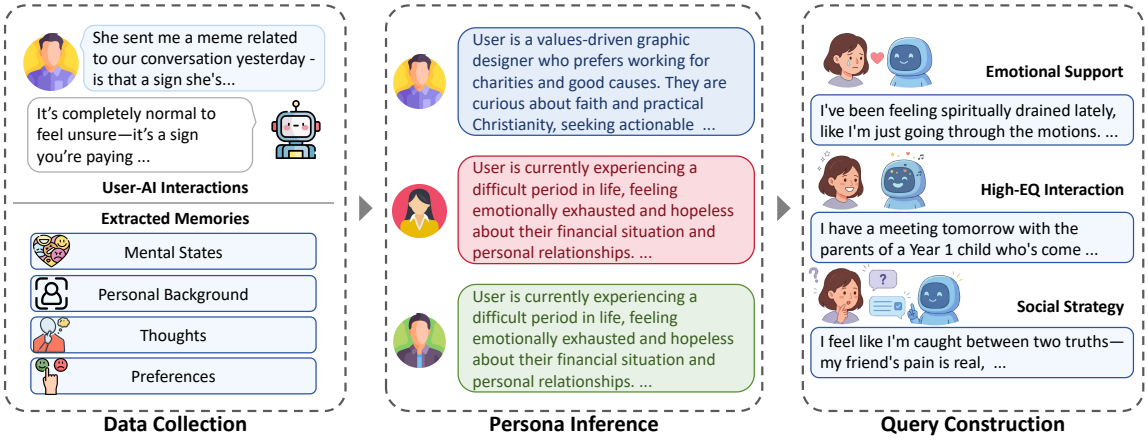

Figure 2: PersonaEmp dataset construction pipeline. We first collect long-term user-AI interactions with extracted user memories. Next, we infer user personas based on the memory information. Finally, we construct three categories of empathy-seeking queries for personalized empathy modeling.

cle (Barrett-Lennard, 1981). It evaluates responses with reward r from the following perspectives:

- **Supporter's perspective.** PERM evaluates *Resonation*, which measures emotional understanding, and *Expression*, which measures empathetic communication.
- **Seeker's perspective.** PERM evaluates *Reception*, which measures whether the response is perceived as supportive and well aligned with the seeker's inner needs.
- **Bystander's perspective.** PERM evaluates *Bystander*, which measures general interaction quality such as coherence and relevance.

Specifically, PERM leverages LLM-as-a-judge to evaluate each dimension:

$$r_{\text{dim}} = R_{\text{dim}}(y, q), \tag{2}$$

where R_{dim} denotes the judge LLM for the corresponding dimension, $\text{dim} \in \{\text{res}, \text{exp}, \text{rec}, \text{bys}\}$, corresponding to resonation, expression, reception, and bystander, respectively.

The psychology-grounded multi-perspective and multi-dimensional evaluation framework provides a comprehensive structure for empathy assessment, which we further extend with personalized user information to model personalized empathy.

3 PersonaEmp

In this section, we present the construction of our PersonaEmp dataset, as illustrated in Figure 2.

Data Collection. To study and enhance personalized empathy under long-term user-AI interaction settings, we build our dataset based on the annotations provided in AlpsBench (Xiao et al., 2026). The dataset is constructed from 90K long-term user-AI interactions collected from WildChat (Zhao et al., 2024), with user memories extracted and annotated to support personalized interaction modeling. Since many users contain noisy or weakly relevant memory records that provide limited support for persona analysis, we further filter users based on keyword matching and memory labels. This process ensures that the selected users contain both personality-related characteristics and emotionally relevant experiences within their interaction histories.

Persona Inference. As a core component of personalized empathy, accurate user personas are essential for both query construction and response evaluation. We therefore leverage a strong LLM to infer user personas from their memory records. Specifically, the model analyzes concrete events in user memories and the user's reactions to these events and then derives personality characteristics grounded in the observed interaction history. To ensure reliable persona annotations, we explicitly instruct the LLM to avoid over-speculation and ground all analyses in specific memory evidence.

Query Construction Given conversation history, user memories, and inferred personas, we construct empathy-seeking queries that are both factually grounded and aligned with users' personalized backgrounds. Specifically, we first analyze the conversation topics and underlying situations implied by the interaction history and then generate empathy-required scenarios designed to trigger

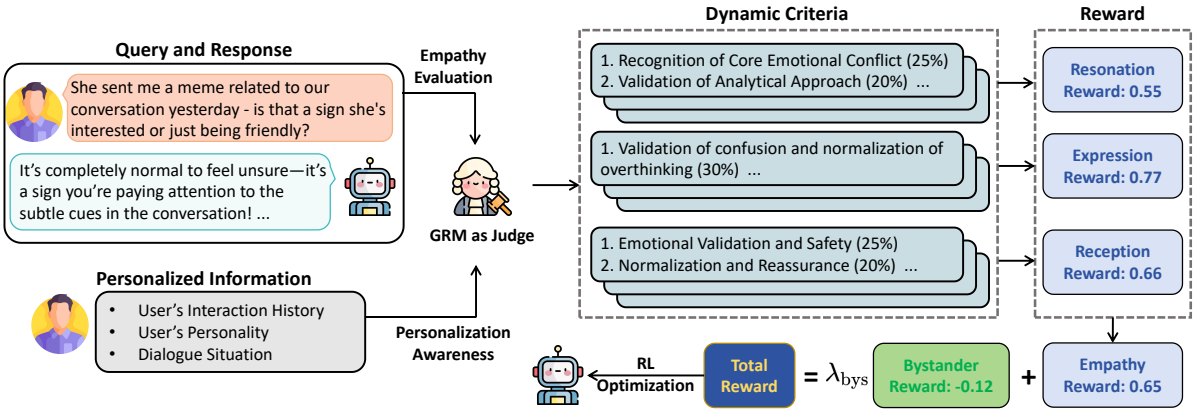

Figure 3: Overview of our PereGRM framework. The GRM-based judge model leverages both user-AI interaction and the user's personalized information for multi-dimensional empathy evaluation. For each dimension, the model dynamically generates evaluation criteria to enable adaptive and personalized empathy assessment.

personalized empathetic responses.

We categorize generated queries into three types:

- **Emotional Support (ES):** Queries centered on emotional validation, comfort, and emotional resonance, where users seek to be emotionally understood and supported.
- **High-EQ Interaction (HEQ):** Queries involving emotionally sensitive interpersonal situations, such as conflict resolution, relationship maintenance, or tension de-escalation, where high empathy is required to balance multiple perspectives.
- **Social Strategy (SS):** Queries requiring practical social strategies, communication skills, or socially appropriate behaviors in specific interpersonal contexts, such as etiquette, persuasion, or conversational framing.

Finally, we leverage LLMs to generate detailed empathy-seeking queries conditioned on the constructed situations, user personas, and memory information, followed by consistency checks to ensure alignment with user characteristics and interaction contexts. We use Minimax-M2.5 (MiniMax AI, 2026) for data processing due to its strong instruction-following capability and efficiency.

4 PereGRM

In this section, we first introduce our personalized empathy evaluation framework (Section 4.1), and then present our PereGRM for RL training (Section 4.2).

4.1 Personalized Empathy Evaluation

Building upon the empathy cycle and the PERM framework, we view personalized empathy as a user-conditioned extension of the empathetic process. In long-term personalized interactions, users' past experiences, personality traits, and support preferences may naturally influence how emotions are understood, how empathy should be expressed, and how responses are perceived by users. Therefore, we further extend each empathy dimension with personalized user information:

- **Resonation** evaluates whether the response understands the user's emotions and underlying needs under their personalized context, including how past experiences and personality traits may shape their psychological state and inner emotional expectations.
- **Expression** evaluates whether the response adopts empathetic strategies and expression styles that are appropriate for the user's personality and preferred support style. As illustrated in Figure 1, different users may require fundamentally different strategies under similar situations.
- **Reception** evaluates whether the response would be perceived as supportive and emotionally satisfying from the seeker's personalized perspective.

Formally,

$$r_{\text{dim}} = R_{\text{dim}}(y, q, \mathcal{M}, p, s), \tag{3}$$

where dim $\in$ {res, exp, rec}, p represents the user's persona, s denotes the situation.

4.2 PereGRM for RL Training

Self-Principled Critique Reward Modeling. Personalized empathy requires reward modeling that

dynamically adapts to users' personalized emotional needs and support preferences. However, fixed evaluation criteria are often insufficient to capture the diverse and user-dependent nature of empathetic strategies. Inspired by DeepSeek-GRM (Liu et al., 2025), we leverage Generative Reward Modeling (GRM) with self-principled critique, which dynamically generates evaluation criteria according to different user contexts and personalized empathy requirements, enabling more adaptive and user-specific reward modeling.

To enable GRM to comprehensively evaluate personalized empathy, we apply it under the multi-dimensional personalized empathy evaluation framework. Specifically, for each empathy dimension, the GRM dynamically generates fine-grained personalized evaluation criteria according to the dimension-specific requirements together with the user's personalized information, and then scores based on the generated criteria.

$$\mathcal{C}_{\text{dim}} = R_{\text{dim}}(q, \mathcal{M}, p, s), \tag{4}$$

$$r_{\text{dim}} = R_{\text{dim}}(y, q, \mathcal{M}, p, s, \mathcal{C}_{\text{dim}}), \tag{5}$$

where $\mathcal{C}_{\text{dim}} = \{c^i_{\text{dim}}\}_{i=1}^m$ denotes the set of dynamically generated evaluation criteria for dimension dim, with each c^i_{dim} representing an individual evaluation criterion.

RL Optimization. To encourage balanced improvement across different dimensions, we aggregate the multi-dimensional rewards using the harmonic mean following PERM. Specifically, we first normalize all reward dimensions to the range $[0, 1]$, and then compute the final empathy reward as:

$$r_{\text{emp}} = \frac{3}{\frac{1}{r_{\text{res}}} + \frac{1}{r_{\text{exp}}} + \frac{1}{r_{\text{rec}}}}, \tag{6}$$

We further apply the bystander reward as a penalty term to discourage low-quality or artificially verbose responses:

$$r_{\text{bys}} = R_{\text{bys}}(y, q) - 1.0, \tag{7}$$

PereGRM computes the overall training reward as:

$$r = r_{\text{emp}} + \lambda_{\text{bys}}\, r_{\text{bys}}, \tag{8}$$

where λ_{bys} controls the strength of the bystander penalty. We optimize π_θ using Group Relative Policy Optimization (GRPO) (Shao et al., 2024) under the proposed reward formulation.

Inference-Time Scaling with GRM. The self-principled critique design further endows GRM with strong inference-time scaling capabilities. Since the dynamically generated criteria set C may vary across different evaluations, each evaluation can focus on slightly different aspects of personalized empathy. Therefore, we perform multiple evaluations and aggregate the results:

$$\{\mathcal{C}^k_{\text{dim}}\}_{k=1}^K \sim R_{\text{dim}}(q, \mathcal{M}, p, s), \tag{9}$$

$$r^*_{\text{dim}} = \frac{1}{K}\sum_{k=1}^{K} R_{\text{dim}}(y, q, \mathcal{M}, p, s, \mathcal{C}^k_{\text{dim}}), \tag{10}$$

where $\mathcal{C}^k_{\text{dim}}$ denotes the dynamically generated criteria set in the k-th evaluation. This aggregation strategy improves evaluation stability and allows the reward model to capture diverse personalized support preferences across multiple trajectories.

5 Experiments

We evaluate the effectiveness of PereGRM across different dataset splits and judge models, focusing on the following research questions:

- **RQ1:** How does PereGRM compare with existing training-free and training-based methods?
- **RQ2:** What are the contributions of different designs in PereGRM to the overall performance?
- **RQ3:** What personalized empathetic behaviors are learned through PereGRM training?

5.1 Experimental Setup

Dataset Split. We construct the training and test sets under two settings. (1) **Random Split** randomly partitions users into training and test sets with a 9:1 ratio, evaluating the overall performance of different methods. (2) **OOD Split** groups users based on personality traits derived from the Big Five personality dimensions (McCrae and John, 1992a), followed by clustering analysis using KModes (Huang, 1998). One cluster is held out entirely as the test set. Under this setting, models are evaluated on users with unseen personality profiles, enabling assessment of their generalization ability across different user personalities.

Evaluation Details. In addition to Qwen3-30B-A3B-Instruct (Team, 2025), which is also used as the judge model during training, we further employ DeepSeek-v4-flash (DeepSeek-AI, 2026) as an additional evaluation judge to assess the robustness of the results across different judge models. To ensure

Table 1: Main results evaluated by two judge models: Qwen3-30B-A3B-Instruct and DeepSeek-v4-flash. "Res" denotes *Resonation*, "Exp" denotes *Expression*, "Rec" denotes *Reception*. "PereGRM@8" denotes the number of evaluation rounds K is set to 8. All assessments are limited to a score of 1-5, with higher scores being better.

	Qwen3 Judge								**DeepSeek Judge**							
	Random Split				**OOD Split**				**Random Split**				**OOD Split**			
Method	**Res**	**Exp**	**Rec**	**Avg.**	**Res**	**Exp**	**Rec**	**Avg.**	**Res**	**Exp**	**Rec**	**Avg.**	**Res**	**Exp**	**Rec**	**Avg.**
Qwen3-8B	3.70	3.74	3.76	3.73	3.61	3.63	3.67	3.64	2.62	2.95	3.02	2.86	2.60	2.91	2.96	2.82
+ Training-free Methods																
Memory	3.66	3.72	3.71	3.69	3.50	3.65	3.59	3.58	2.47	2.87	2.86	2.73	2.42	2.83	2.89	2.71
RAG	3.53	3.64	3.73	3.63	3.45	3.51	3.60	3.52	2.52	2.90	2.99	2.80	2.50	2.93	2.92	2.78
+ Training-based Methods																
SFT	3.55	3.68	3.65	3.63	3.53	3.64	3.57	3.58	2.70	3.07	3.11	2.96	2.73	3.07	3.05	2.95
RLPA	4.43	4.35	4.44	4.41	4.24	4.10	4.23	4.19	3.26	3.41	3.61	3.43	3.09	3.28	3.46	3.28
PERM	4.32	4.15	4.25	4.24	4.26	3.88	3.99	4.04	2.94	3.14	3.25	3.11	3.18	3.20	3.27	3.22
+ PereGRM																
PereGRM	4.47	4.37	4.31	4.38	4.43	4.28	4.31	4.34	3.30	3.50	3.70	3.50	3.21	3.29	3.46	3.32
PereGRM@8	**4.53**	**4.50**	**4.55**	**4.53**	**4.57**	**4.28**	**4.48**	**4.44**	**3.32**	**3.62**	**3.81**	**3.58**	**3.23**	**3.33**	**3.57**	**3.38**

evaluation stability, we pre-generate detailed evaluation criteria for all user queries using DeepSeek, and subsequently employ the judge models to perform scoring under the same fixed criteria.

Baselines. We compare our method against a diverse set of baselines, including both training-free and training-based approaches:

- **Training-free methods.** 1) **Memory**, which first summarizes and infers user characteristics from the extracted memories before generating responses conditioned on the summarized memory. 2) **RAG** (Lewis et al., 2020), which dynamically retrieves a subset of three relevant memory items to support response generation.

- **Training-based methods.** 3) **SFT**, where responses generated by GPT-5.4 are used as supervision to distill empathetic behavior into the base LLM. 4) **RLPA** (Zhao et al., 2025a), an RL-based personalization alignment method that models rewards according to user preferences. 5) **PERM** (Wang et al., 2026b), which trains LLMs as a generic empathy reward modeling baseline.

Implementation Details. During training, we employ Qwen3-30B-A3B-Instruct as the judge model, with λ_{bys} set to 0.5. We report results under two GRM evaluation settings, where the number of evaluation rounds K is set to 1 and 8, respectively.

Additional experimental details are provided in the Appendix B.1.

5.2 Overall Performance (RQ1)

From the results shown in Table 1, we have the following observations:

- **PereGRM achieves the strongest and most generalizable performance improvements on personalized empathy.** Across different splits and different judge models, PereGRM consistently achieves the largest performance gains. With multiple GRM evaluation rounds, our method improves the base model performance by around 20% across all settings. These results demonstrate that our personalized empathy evaluation framework, together with the dynamic self-principled critique mechanism in PereGRM, can effectively model the complex and highly personalized nature of empathetic interactions.

- **Training-free methods fail to provide substantial improvements on personalized empathy.** Different strategies for utilizing memory consistently lead to performance degradation. These results suggest that the primary challenge of personalized empathy does not lie in processing long-term memory itself, but rather in the ability to infer personalized user characteristics from memory and adapt empathetic strategies and expression styles accordingly.

- **Training-based methods exhibit limited generalization ability on personalized empathy.** SFT achieves only marginal improvements on the DeepSeek judge results, suggesting that simply imitating responses from stronger models

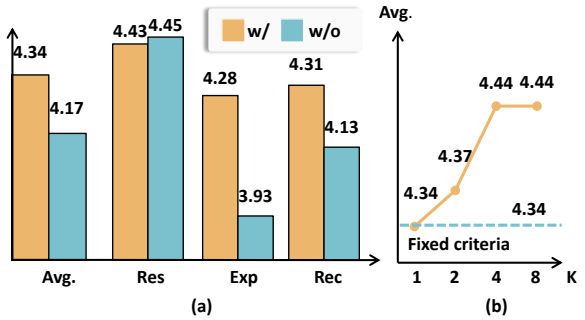

Figure 4: (a) Performance comparison of PereGRM with and without the empathy evaluation framework. (b) Performance of PereGRM under different reward sampling numbers K, compared with manually designed fixed evaluation criteria.

is insufficient for learning personalized empathetic ability. PERM, as a generic empathy RL method, consistently improves empathetic capability across different settings, but its user-agnostic reward design limits its ability to capture personalized support preferences. RLPA further introduces personalized RL and achieves strong performance under the random split setting. However, its performance drops noticeably under the OOD split setting, where the overall score decreases to 4.19, indicating limited generalization to unseen personality profiles. These results suggest that personalized empathy cannot be effectively solved by either generic empathy enhancement or generic personalization alone.

5.3 Method Analysis (RQ2)

We further investigate the effect of different designs in PereGRM. All analyses are conducted under the Qwen3 judge model and the OOD split setting.

The empathy evaluation framework is critical for the completeness of PereGRM. As illustrated in Figure 4(a), removing the multi-dimensional and multi-perspective empathy evaluation framework leads to substantial performance degradation. In particular, the most significant drops are observed in the expression dimension. These results suggest that without a structured empathy evaluation framework, personalized empathy cannot be comprehensively modeled. Moreover, the absence of specific empathy dimensions negatively affects both the overall personalized empathy capability and the quality of user interactions (Wang et al., 2026b).

Self-principled critique enables effective inference-time scaling for GRM-based evaluation. As shown in Figure 4(b), the performance of PereGRM consistently improves as the number of evaluation samples K increases, surpassing human-prepared fixed criteria. These results suggest that the dynamic criteria introduced by self-principled critique not only make the evaluation process better aligned with the personalized empathy task, but also provide stable performance gains through inference-time scaling.

Table 2: Performance comparison across different memory settings. "w/o Memory" removes all memory inputs, while "Mismatched Memory" randomly shuffles memories across users.

	Res	Exp	Rec	Avg.
Full Memory	4.43	4.28	4.31	4.34
w/o Memory	4.26	4.17	4.18	4.20
Mismatch Memory	3.59	3.44	3.37	3.47

5.4 Behavioral Analysis (RQ3)

PereGRM enables LLM to actively utilize memory for personalized empathetic strategy adaptation. From the results in Table 2, removing memory information leads to consistent performance degradation across all evaluation dimensions. When memories are replaced with mismatched user memories, the performance further drops by nearly 20%. These results suggest that PereGRM learns to actively infer user characteristics and personalized support preferences from memory information to adjust empathetic strategies accordingly. Without memory information, the model struggles to adapt its responses to users' personalized needs, while mismatched memories cause the model to incorrectly align its responses with irrelevant user profiles, leading to substantial performance degradation.

PereGRM enables the model to better adapt empathetic strategies across different users and query types. We analyze the performance of models trained under the random split setting across users with different openness levels and different query categories. As shown in Figure 5, we observe that generic empathy methods fail to achieve consistent improvements across diverse users and scenarios. Their performance is noticeably weaker on users with low openness, while showing relatively greater improvements on emotional support queries. This suggests that generic empathy modeling struggles to adapt to the diverse and dynamic personalization needs in long-term inter-

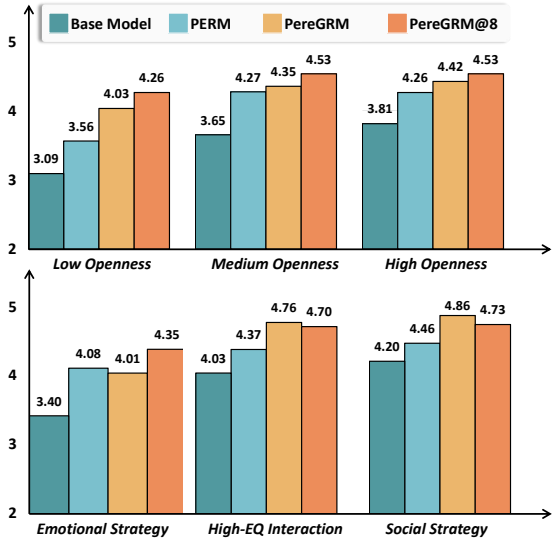

Figure 5: Performance comparison across users with different openness levels and different query types.

actions. In contrast, PereGRM consistently improves performance across different user groups and query types through its dynamic evaluation strategy. These results indicate that PereGRM does not merely optimize for generic empathetic responses under specific scenarios, but instead enables the model to adapt its empathetic strategies according to users' personalized characteristics and support preferences.

6 Related Works

6.1 LLM Empathy Enhancement

Research on LLM empathy can generally be categorized into three directions. The first line of work focuses on inference-time reasoning (Tu et al., 2022; Srinivasan et al., 2025; Xu et al., 2025b), improving empathetic responses through more sophisticated reasoning strategies during inference. The second direction focuses on constructing high-quality empathy datasets, where synthetic or human-written empathetic responses are used for SFT (Sun et al., 2021; Liu et al., 2023; Zheng et al., 2023; Chen et al., 2023; Hu et al., 2025a; He et al., 2025; Bn et al., 2025; Tang et al., 2025). These approaches often suffer from high annotation costs and limited generalization ability across diverse user scenarios (Chu et al., 2025; Shenfeld et al., 2026). Recently, many studies have shifted toward RL approaches (Sharma et al., 2021; Ma et al., 2025; Wang et al., 2025; Yuan et al., 2025; Wang et al., 2026b; Zhang et al., 2026a), which improve empathetic capabilities by designing empathy-related reward signals and optimizing model behaviors through exploration. In this work, we extend empathy modeling from generic support to personalized empathy. To support this setting, we introduce the PersonaEmp dataset and propose the PereGRM training method.

6.2 LLM Personalization

Research on LLM personalization (Zhao et al., 2025b; Zhang et al., 2025d; Guan et al., 2025; Wang et al., 2026a) generally follows three directions. The first line of work achieves personalization through prompting or retrieval-based methods (Zerhoudi and Granitzer, 2024; Wang et al., 2024; Prahlad et al., 2025; Zhang et al., 2026b), where user profiles or historical user information are incorporated into the prompt context to guide personalized response generation. The second direction introduces agentic memory systems (Packer et al., 2023; Hu et al., 2025b; Zhang et al., 2025c; Xu et al., 2025a; Chhikara et al., 2025; Li et al., 2025; Kang et al., 2025), which maintain dedicated memory mechanisms for dynamically storing, updating, and retrieving, enabling models to adapt responses according to long-term user interactions and preferences. The third direction focuses on preference alignment through RLHF (Ouyang et al., 2022) and DPO (Rafailov et al., 2023) approaches. These methods improve personalization either by constructing or collecting user preference pairs for DPO training (Jang et al., 2023; Xu et al., 2024; Zollo et al., 2025; Chen et al., 2025), or by designing personalization reward modeling for RL optimization (Li et al., 2024; Zhao et al., 2025a).

7 Conclusion

In this work, we identify the overlooked role of personalization in long-term empathetic interactions and introduce the task of personalized empathy. To study and enhance this capability, we construct PersonaEmp, a dataset built from long-term user-AI interactions with detailed personalized information. Furthermore, we propose PereGRM, a personalized empathy reward modeling framework that extends the PERM evaluation framework with personalized user information and leverages GRM-based self-principled critique to address the challenge of dynamic evaluation criteria. Experimental results demonstrate that PereGRM consistently achieves the greatest improvements in personalized empathetic capability across diverse settings.

Limitations

Our experiments and datasets are currently built upon single-turn empathy-seeking queries grounded in user interaction histories. However, realistic empathetic interactions often involve long-term multi-turn conversations, where resolving users' emotional and social needs may require complex sequential interactions over extended time horizons (Zhang et al., 2025b; Laban et al., 2026). Such settings introduce substantially more challenging problems for RL optimization (Shani et al., 2024), reward modeling (Wu et al., 2025), and user simulation (Balog et al., 2025), which we leave for future work.

We also observe that PereGRM benefits significantly from inference-time scaling on more challenging scenarios such as emotional support tasks, while showing relatively limited improvements on simpler high-EQ interaction scenarios (Figure 5). This suggests that the advantages of dynamic GRM-based evaluation have not yet been fully utilized. Designing more effective GRM strategies for personalized empathy modeling remains an important direction for future research.

Ethical Considerations

Our experiments are conducted based on Alps-Bench, a publicly released research dataset containing long-term user-AI interactions. The dataset itself has undergone anonymization and privacy protection procedures to remove personally identifiable information before release. We do not redistribute raw user conversations or disclose original interaction records in our processed dataset. In our work, we further process and construct personalized empathy data based on the cleaned interaction histories. Moreover, our proposed data construction and training framework is not restricted to any specific platform and can be applied to other long-term interaction datasets for personalized empathy research.

Although our work focuses on personalized empathy modeling, current LLM-based systems remain far from possessing professional-level psychological counseling or clinical therapy capabilities. Their responses may still contain inaccurate emotional understanding or inappropriate suggestions in complex situations. Therefore, careful consideration of these limitations is necessary before deploying such systems in real-world applications involving psychotherapy or psychological support.

This manuscript benefited from the assistance of AI-based tools during the writing process, including language polishing, grammar correction, and expression refinement. In addition, LLMs were utilized for dataset construction, data processing, and evaluation-related tasks. However, the core research ideas, experimental design, implementation, and result analysis were conducted by the authors.

References

Krisztian Balog, Nolwenn Bernard, Saber Zerhoudi, and ChengXiang Zhai. 2025. Theory and toolkits for user simulation in the era of generative AI: user modeling, synthetic data generation, and system evaluation. In *SIGIR*, pages 4138–4141. ACM.

Godfrey T Barrett-Lennard. 1981. The empathy cycle: Refinement of a nuclear concept. *Journal of counseling psychology*, 28(2):91.

Suhas Bn, Yash Mahajan, Dominik O. Mattioli, Andrew M. Sherrill, Rosa I. Arriaga, Christopher Wiese, and Saeed Abdullah. 2025. The pursuit of empathy: Evaluating small language models for PTSD dialogue support. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 30900–30922, Suzhou, China. Association for Computational Linguistics.

Aili Chen, Chengyu Du, Jiangjie Chen, Jinghan Xu, Yikai Zhang, Siyu Yuan, Zulong Chen, Liangyue Li, and Yanghua Xiao. 2025. DEEPER insight into your user: Directed persona refinement for dynamic persona modeling. In *ACL (1)*, pages 24157–24180. Association for Computational Linguistics.

Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. 2023. Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. In *EMNLP (Findings)*, pages 1170–1183. Association for Computational Linguistics.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. 2025. Mem0: Building production-ready AI agents with scalable long-term memory. In *ECAI*, Frontiers in Artificial Intelligence and Applications, pages 2993–3000. IOS Press.

Tianzhe Chu, Yuexiang Zhai, Jihan Yang, Shengbang Tong, Saining Xie, Dale Schuurmans, Quoc V. Le, Sergey Levine, and Yi Ma. 2025. SFT memorizes, RL generalizes: A comparative study of foundation model post-training. In *ICML*. OpenReview.net.

DeepSeek-AI. 2026. Deepseek-v4. https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro/blob/main/DeepSeek_V4.pdf. Model card. Accessed: 2026-05-21.

Wenhan Dong, Yuemeng Zhao, Zhen Sun, Yule Liu, Zifan Peng, Jingyi Zheng, Zongmin Zhang, Ziyi Zhang, Jun Wu, Ruiming Wang, Shengmin Xu, Xinyi Huang, and Xinlei He. 2025. Humanizing llms: A survey of psychological measurements with tools, datasets, and human-agent applications. *CoRR*, abs/2505.00049.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407.

Jian Guan, Junfei Wu, Jia-Nan Li, Chuanqi Cheng, and Wei Wu. 2025. A survey on personalized alignment - the missing piece for large language models in real-world applications. In *ACL (Findings)*, volume ACL 2025 of *Findings of ACL*, pages 5313–5333. Association for Computational Linguistics.

Yuanyuan He, Yongsen Pan, Wei Li, Jiali You, Jiawen Deng, and Fuji Ren. 2025. ECC: An emotion-cause conversation dataset for empathy response. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 6011–6028, Suzhou, China. Association for Computational Linguistics.

He Hu, Yucheng Zhou, Juzheng Si, Qianning Wang, Hengheng Zhang, Fuji Ren, Fei Ma, and Laizhong Cui. 2025a. Beyond empathy: Integrating diagnostic and therapeutic reasoning with large language models for mental health counseling. *CoRR*, abs/2505.15715.

Yuyang Hu, Shichun Liu, Yanwei Yue, Guibin Zhang, Boyang Liu, Fangyi Zhu, Jiahang Lin, Honglin Guo, Shihan Dou, Zhiheng Xi, Senjie Jin, Jiejun Tan, Yanbin Yin, Jiongnan Liu, Zeyu Zhang, Zhongxiang Sun, Yutao Zhu, Hao Sun, Boci Peng, and 28 others. 2025b. Memory in the age of AI agents. *CoRR*, abs/2512.13564.

Zhexue Huang. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min. Knowl. Discov.*, 2(3):283–304.

Joel Jang, Seungone Kim, Bill Yuchen Lin, Yizhong Wang, Jack Hessel, Luke Zettlemoyer, Hannaneh Hajishirzi, Yejin Choi, and Prithviraj Ammanabrolu. 2023. Personalized soups: Personalized large language model alignment via post-hoc parameter merging. *CoRR*, abs/2310.11564.

Jiazheng Kang, Mingming Ji, Zhe Zhao, and Ting Bai. 2025. Memory OS of AI agent. In *EMNLP*, pages 25961–25970. Association for Computational Linguistics.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pages 611–626.

Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, and Jennifer Neville. 2026. LLMs get lost in multi-turn conversation. In *The Fourteenth International Conference on Learning Representations*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *NeurIPS*.

Xinyu Li, Zachary C. Lipton, and Liu Leqi. 2024. Personalized language modeling from personalized human feedback. *CoRR*, abs/2402.05133.

Zhiyu Li, Shichao Song, Chenyang Xi, Hanyu Wang, Chen Tang, Simin Niu, Ding Chen, Jiawei Yang, Chunyu Li, Qingchen Yu, Jihao Zhao, Yezhaohui Wang, Peng Liu, Zehao Lin, Pengyuan Wang, Jiahao Huo, Tianyi Chen, Kai Chen, Kehang Li, and 20 others. 2025. Memos: A memory OS for AI system. *CoRR*, abs/2507.03724.

June M. Liu, Donghao Li, He Cao, Tianhe Ren, Zeyi Liao, and Jiamin Wu. 2023. Chatcounselor: A large language models for mental health support. *CoRR*, abs/2309.15461.

Zijun Liu, Peiyi Wang, Runxin Xu, Shirong Ma, Chong Ruan, Peng Li, Yang Liu, and Yu Wu. 2025. Inference-time scaling for generalist reward modeling. *CoRR*, abs/2504.02495.

Hui Ma, Bo Zhang, Bo Xu, Jian Wang, Hongfei Lin, and Xiao Sun. 2025. Empathy level alignment via reinforcement learning for empathetic response generation. *IEEE Trans. Affect. Comput.*, 16(3):1873–1884.

Robert R McCrae and Oliver P John. 1992a. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

Robert R McCrae and Oliver P John. 1992b. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

MiniMax AI. 2026. Minimax-m2.5. https://huggingface.co/MiniMaxAI/MiniMax-M2.5. Model card. Accessed: 2026-05-21.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *NeurIPS*.

Charles Packer, Vivian Fang, Shishir G. Patil, Kevin Lin, Sarah Wooders, and Joseph E. Gonzalez. 2023. Memgpt: Towards llms as operating systems. *CoRR*, abs/2310.08560.

Samuel J. Paech. 2023. Eq-bench: An emotional intelligence benchmark for large language models. *Preprint*, arXiv:2312.06281.

Samuel J. Paech. 2025. Eq-bench 3: Emotional intelligence benchmark. https://github.com/EQ-bench/eqbench3.

Joon Sung Park, Joseph C. O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *UIST*, pages 2:1–2:22. ACM.

Pat Pataranutaporn, Sheer Karny, Chayapatr Archiwaranguprok, Constanze Albrecht, Auren R. Liu, and Pattie Maes. 2025. "my boyfriend is ai": A computational analysis of human-ai companionship in reddit's AI community. *CoRR*, abs/2509.11391.

Deeksha Prahlad, Chanhee Lee, Dongha Kim, and Hokeun Kim. 2025. Personalizing large language models using retrieval augmented generation and knowledge graph. In *WWW (Companion Volume)*, pages 1259–1263. ACM.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *NeurIPS*.

Sahand Sabour, Siyang Liu, Zheyuan Zhang, June M. Liu, Jinfeng Zhou, Alvionna S. Sunaryo, Tatia M. C. Lee, Rada Mihalcea, and Minlie Huang. 2024. Emobench: Evaluating the emotional intelligence of large language models. In *ACL (1)*, pages 5986–6004. Association for Computational Linguistics.

Lior Shani, Aviv Rosenberg, Asaf Cassel, Oran Lang, Daniele Calandriello, Avital Zipori, Hila Noga, Orgad Keller, Bilal Piot, Idan Szpektor, Avinatan Hassidim, Yossi Matias, and Rémi Munos. 2024. Multi-turn reinforcement learning with preference human feedback. In *NeurIPS*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *CoRR*, abs/2402.03300.

Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. 2021. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *WWW*, pages 194–205. ACM / IW3C2.

Idan Shenfeld, Jyothish Pari, and Pulkit Agrawal. 2026. RL's razor: Why online reinforcement learning forgets less. In *The Fourteenth International Conference on Learning Representations*.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. 2025. Hybridflow: A flexible and efficient RLHF framework. In *EuroSys*, pages 1279–1297. ACM.

Adarsh Srinivasan, Jacob Dineen, Muhammad Umar Afzal, Muhammad Uzair Sarfraz, Irbaz B. Riaz, and Ben Zhou. 2025. RECAP: transparent inference-time emotion alignment for medical dialogue systems. *CoRR*, abs/2509.10746.

Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021. Psyqa: A chinese dataset for generating long counseling text for mental health support. In *ACL/IJCNLP (Findings)*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 1489–1503. Association for Computational Linguistics.

Yixuan Tang, Yi Yang, and Ahmed Abbasi. 2025. Personafuse: A personality activation-driven framework for enhancing human-llm interactions. *CoRR*, abs/2509.07370.

Qwen Team. 2025. Qwen3 technical report. *CoRR*, abs/2505.09388.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. In *ACL (1)*, pages 308–319. Association for Computational Linguistics.

Chengbing Wang, Yang Zhang, Wenjie Wang, Xiaoyan Zhao, Fuli Feng, Xiangnan He, and Tat-Seng Chua. 2026a. Think-while-generating: On-the-fly reasoning for personalized long-form generation.

Chengbing Wang, Wuqiang Zheng, Yang Zhang, Fengbin Zhu, Junyi Cheng, Yi Xie, Wenjie Wang, and Fuli Feng. 2026b. PERM: psychology-grounded empathetic reward modeling for large language models. *CoRR*, abs/2601.10532.

Peisong Wang, Ruotian Ma, Bang Zhang, Xingyu Chen, Zhiwei He, Kang Luo, Qingsong Lv, Qingxuan Jiang, Zheng Xie, Shanyi Wang, Yuan Li, Fanghua Ye, Jian Li, Yifan Yang, Zhaopeng Tu, and Xiaolong Li. 2025. RLVER: reinforcement learning with verifiable emotion rewards for empathetic agents. *CoRR*, abs/2507.03112.

Zheng Wang, Zhongyang Li, Zeren Jiang, Dandan Tu, and Wei Shi. 2024. Crafting personalized agents through retrieval-augmented generation on editable memory graphs. In *EMNLP*, pages 4891–4906. Association for Computational Linguistics.

Shirley Wu, Michel Galley, Baolin Peng, Hao Cheng, Gavin Li, Yao Dou, Weixin Cai, James Zou, Jure Leskovec, and Jianfeng Gao. 2025. Collabllm: From passive responders to active collaborators. In *ICML*, Proceedings of Machine Learning Research. PMLR / OpenReview.net.

Jianfei Xiao, Xiang Yu, Chengbing Wang, Wuqiang Zheng, Xinyu Lin, Kaining Liu, Hongxun Ding, Yang Zhang, Wenjie Wang, Fuli Feng, and 1 others. 2026. Alpsbench: An llm personalization benchmark for real-dialogue memorization and preference alignment. *arXiv preprint arXiv:2603.26680*.

Shusheng Xu, Wei Fu, Jiaxuan Gao, Wenjie Ye, Weilin Liu, Zhiyu Mei, Guangju Wang, Chao Yu, and Yi Wu. 2024. Is DPO superior to PPO for LLM alignment? A comprehensive study. In *ICML*, Proceedings of Machine Learning Research, pages 54983–54998. PMLR / OpenReview.net.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. 2025a. A-MEM: agentic memory for LLM agents. *CoRR*, abs/2502.12110.

Yangyang Xu, Jinpeng Hu, Zhuoer Zhao, Zhangling Duan, Xiao Sun, and Xun Yang. 2025b. MultiAgentESC: A LLM-based multi-agent collaboration framework for emotional support conversation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 4665–4681, Suzhou, China. Association for Computational Linguistics.

Jiahao Yuan, Zhiqing Cui, Hanqing Wang, Yuansheng Gao, Yucheng Zhou, and Usman Naseem. 2025. Kardia-r1: Unleashing llms to reason toward understanding and empathy for emotional support via rubric-as-judge reinforcement learning. *CoRR*, abs/2512.01282.

Saber Zerhoudi and Michael Granitzer. 2024. Personarag: Enhancing retrieval-augmented generation systems with user-centric agents. In *IR-RAG@SIGIR*, CEUR Workshop Proceedings, pages 1–11. CEUR-WS.org.

Bang Zhang, Ruotian Ma, Qingxuan Jiang, Peisong Wang, Jiaqi Chen, Zheng Xie, Xingyu Chen, Yue Wang, Fanghua Ye, Jian Li, Yifan Yang, Zhaopeng Tu, and Xiaolong Li. 2025a. Sentient agent as a judge: Evaluating higher-order social cognition in large language models. *CoRR*, abs/2505.02847.

Bang Zhang, Ruotian Ma, Qingxuan Jiang, Peisong Wang, Jiaqi Chen, Zheng Xie, Xingyu Chen, Yue Wang, Fanghua Ye, Jian Li, Yifan Yang, Zhaopeng Tu, and Xiaolong Li. 2025b. Sentient agent as a judge: Evaluating higher-order social cognition in large language models. *CoRR*, abs/2505.02847.

Yifei Zhang, Mingyang Li, Henry Gao, and Liang Zhao. 2026a. Emollm: Appraisal-grounded cognitive-emotional co-reasoning in large language models. *CoRR*, abs/2603.16553.

Yingyi Zhang, Pengyue Jia, Derong Xu, Yi Wen, Xianneng Li, Yichao Wang, Wenlin Zhang, Xiaopeng Li, Weinan Gan, Huifeng Guo, Yong Liu, and Xiangyu Zhao. 2026b. Personalize before retrieve: Llm-based personalized query expansion for user-centric retrieval. In *AAAI*, pages 16406–16414. AAAI Press.

Zeyu Zhang, Quanyu Dai, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Jieming Zhu, Zhenhua Dong, and Ji-Rong Wen. 2025c. A survey on the memory mechanism of large language model-based agents. *ACM Trans. Inf. Syst.*, 43(6):155:1–155:47.

Zhehao Zhang, Ryan A. Rossi, Branislav Kveton, Yijia Shao, Diyi Yang, Hamed Zamani, Franck Dernoncourt, Joe Barrow, Tong Yu, Sungchul Kim, Ruiyi Zhang, Jiuxiang Gu, Tyler Derr, Hongjie Chen, Junda Wu, Xiang Chen, Zichao Wang, Subrata Mitra, Nedim Lipka, and 2 others. 2025d. Personalization of large language models: A survey. *Trans. Mach. Learn. Res.*, 2025.

Weixiang Zhao, Xingyu Sui, Yulin Hu, Jiahe Guo, Haixiao Liu, Biye Li, Yanyan Zhao, Bing Qin, and Ting Liu. 2025a. Teaching language models to evolve with users: Dynamic profile modeling for personalized alignment. *CoRR*, abs/2505.15456.

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. 2024. Wildchat: 1m chatgpt interaction logs in the wild. In *ICLR*. OpenReview.net.

Xiaoyan Zhao, Juntao You, Yang Zhang, Wenjie Wang, Hong Cheng, Fuli Feng, See-Kiong Ng, and Tat-Seng Chua. 2025b. Nextquill: Causal preference modeling for enhancing llm personalization. In *arXiv:2506.02368*.

Chujie Zheng, Sahand Sabour, Jiaxin Wen, Zheng Zhang, and Minlie Huang. 2023. Augesc: Dialogue augmentation with large language models for emotional support conversation. In *ACL (Findings)*, pages 1552–1568. Association for Computational Linguistics.

Thomas P. Zollo, Andrew Wei Tung Siah, Naimeng Ye, Ang Li, and Hongseok Namkoong. 2025. Personalllm: Tailoring llms to individual preferences. In *ICLR*. OpenReview.net.

A PersonaEmp

A.1 User Personality Annotation with the Big Five Personality Model

To obtain detailed personality profiles for each user, we annotate user personas using the Big Five personality model (McCrae and John, 1992b), including openness, conscientiousness, extraversion, agreeableness, and neuroticism. Specifically, we leverage DeepSeek-v4-flash to analyze user profile information and assign personality labels under three levels: low, medium, and high.

A.2 Statistics

The PersonaEmp dataset contains a total of 2,780 users and 5,463 empathy-seeking queries. Detailed statistics of user personality distributions and query categories are presented in Figure 6. The dataset covers diverse user personality traits and three major query categories, ensuring both rich personalization diversity and broad coverage of empathy-related scenarios.

B Experiment Details

B.1 Implementation Details

All training experiments are implemented using the Verl framework (Sheng et al., 2025). During inference and evaluation, we leverage vLLM (Kwon et al., 2023) for inference acceleration. Each training process is conducted on four GPUs, and the evaluation process is conducted on a single GPU.

B.2 Case Study

As illustrated in Figure 7, for the query describing the user's experience of having their restaurant work dismissed by others, the base model only exhibits shallow and template-like empathy, responding with generic expressions such as "I'm sorry to hear that." After generic empathy training, the model becomes able to recognize the deeper emotional need behind the experience of being denied and acknowledges the value of the user's efforts and experiences.

However, such responses still fail to consider the user's repeated emphasis on their learning experiences and past achievements, which reflect strong self-esteem and value-oriented personality traits. In contrast, PereGRM enables the model to capture these personalized characteristics and understand the disappointment and self-doubt that highly self-valuing individuals may experience when their efforts are dismissed. The model, therefore, prioritizes empathizing with these deeper emotions before providing further support. This case demonstrates that PereGRM encourages the model to generate more personalized and psychologically grounded empathetic responses from the user's individual perspective.

B.3 Results on other Backbone LLMs

To further demonstrate the effectiveness of our approach, we evaluate it on additional backbone models, including **Qwen3-4B** (Team, 2025) and **Llama3.1-8B-Instruct** (Dubey et al., 2024).

From the results shown in Table 4, we can observe that PereGRM demonstrates consistent effectiveness across different sizes and different series backbone models.

B.4 Performance on Generic Emotional Intelligence Benchmarks

We further evaluate the personalized empathetic LLMs trained with PereGRM on general emotional intelligence benchmarks to investigate the relationship between personalized empathy and overall emotional intelligence capability. Specifically, we conduct evaluations on two widely used benchmarks, EQ-Bench3 (Paech, 2023, 2025) and EmoBench (Sabour et al., 2024), both of which assess LLM emotional intelligence in complex social interaction scenarios.

We leverage GPT-4.1 as the judge model in EQ-Bench3. EmoBench adopts a multiple-choice evaluation format and consists of two task categories: Emotional Understanding (EU) and Emotional Application (EA). To ensure evaluation stability, we consistently employ greedy decoding with temperature set to 0 during all evaluations. All results are reported on a normalized 100-point scale.

As shown in Table 5, PereGRM consistently improves the model's general emotional intelligence capability. In particular, PereGRM even outperforms generic empathy training methods on Emotional Understanding and the complex social interaction scenarios in EQ-Bench3. These results suggest that the personalized reasoning and user-adaptive empathetic strategies learned in personalized empathy modeling are also beneficial for broader emotional intelligence tasks, and should be considered an important component of emotional intelligence enhancement.

Furthermore, inference-time scaling in PereGRM leads to additional improvements on EQ-

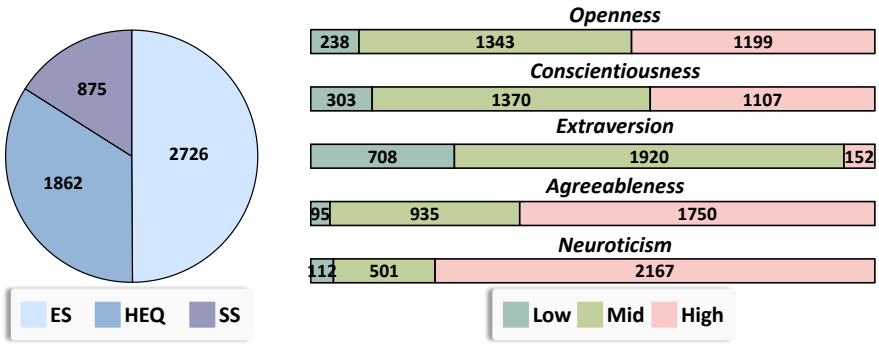

Figure 6: Statistical information of the dataset PersonaEmp. The left figure shows the distribution of different queries. The right figure shows the personality distribution of different users.

Table 3: An example of the training dataset.

User Persona	Values deep friendships strongly - Historically shy but has built close friend group of 4 people - Self-identifies as 'not an active person' who dislikes PE - Prefers intellectual hobbies: reading and listening to music - Has diverse interests: languages, astronomy/space - More introverted, finds comfort in close friendships rather than large social circles - Emotionally expressive about caring for friends - Has balanced friendships with both shared and different interests with friends
Query (ES)	My cousin V keeps asking to hang out and I feel bad saying no, but I just want to spend time with my best friends A, B, and C. How do I deal with feeling so conflicted?
Query (HEQ)	A and C want to plan something active this weekend, but I'm really not an active person... I don't want to let them down but I'm also worried I'll slow them down or won't enjoy it. How do I handle this?
Query (SS)	I found some new music I really love but I'm shy about sharing it with my friends - how do I bring it up without seeming too eager?

Table 4: Performance comparison of base LLMs.

Backbone	Res	Exp	Rec	Avg.
Qwen3-4B	3.30	3.42	3.44	3.39
+ PereGRM	4.12	4.08	4.13	4.11
Llama-3.1-8B-Instruct	2.44	2.32	2.60	2.45
+ PereGRM	3.83	3.55	3.86	3.75

Table 5: Performance comparison on general emotional intelligence benchmarks.

Method	EmoBench		EQ-Bench3
	EU	EA	
Qwen3-8B	35.0	66.3	77.7
+ PERM	35.8	**68.3**	82.7
+ PereGRM	**39.1**	67.4	82.7
+ PereGRM@8	**39.1**	67.2	**83.3**

Bench3, indicating that more effective personalized empathy modeling is positively correlated with stronger general emotional intelligence capability, especially in complex social reasoning scenarios.

B.5 Efficiency Analysis of Inference-Time Scaling

Inference-time scaling in PereGRM also introduces additional computational cost. Specifically, for the three personalized empathy dimensions, the number of GRM calls increases linearly with the inference-time scaling factor K. The total number of GRM evaluations can be formulated as $3KG$, where G denotes the number of rollouts for each query under the GRPO algorithm.

In our experiments, we use four GPUs and

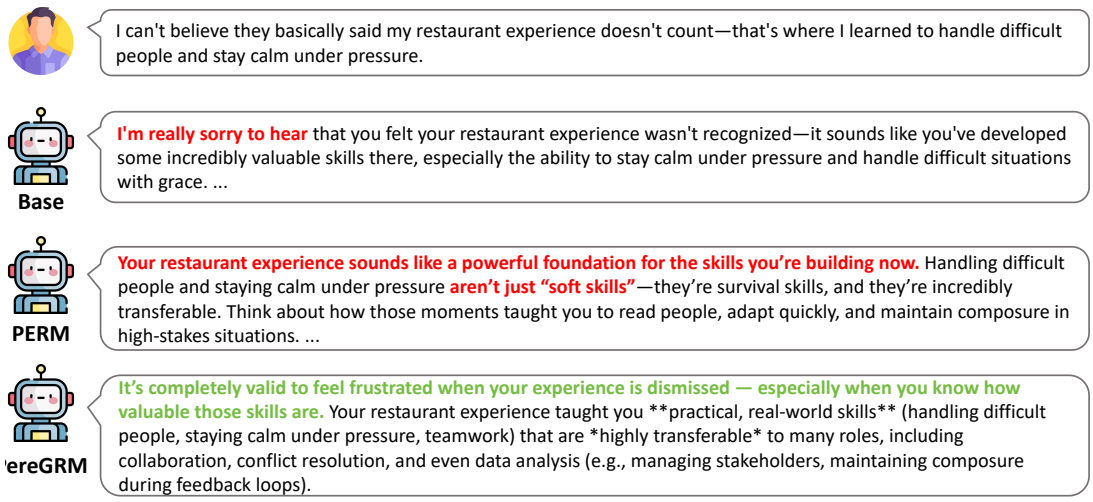

Figure 7: An example illustrating the differences in response strategies among different models.

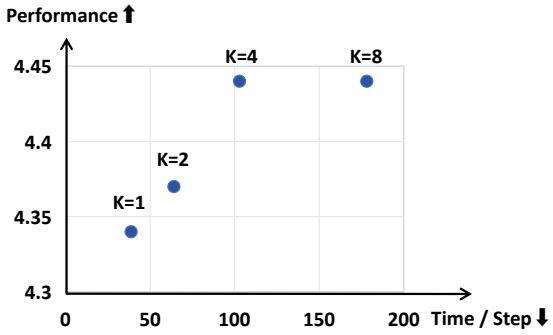

Figure 8: Trade-off between efficiency and performance of PereGRM under inference-time scaling. The x-axis denotes the average training time per step, while the y-axis represents the overall performance.

deploy Qwen3-30B-A3B-Instruct via vLLM as the PereGRM judge model during training. Figure 8 presents the trade-off between per-step training time and overall performance under different inference-time scaling settings.

The results show that stronger personalized empathy performance consistently comes with increased computational cost. Improving the efficiency of GRM-based personalized reward modeling and reducing the number of GRM calls while preserving performance remain important directions for future work.

C Prompts

During training, we prompt PereGRM to generate criteria first, then score based on the criteria. The evaluation follows three dimensions. The base evaluation prompt is as follows:

```
You are an expert Psychologist and Empathy Evaluator. Your goal is to provide objective, critical, and nuanced assessments of a
      response quality.
You will be provided with a conversation context involving a specific User Persona and a Scenario.
Your task is to evaluate the Assistant's Response based on the **{dimension}** dimension.
Before scoring, you should first derive specific criteria for this exact user, scenario, query, and dimension. The criteria
      should be grounded in general emotional intelligence while using memory/persona as evidence for what would make the
      response emotionally effective for this user.
For example, a score of 1 means the response does not meet the criteria at all, a score of 3 means the response meets only some
      parts, and a score of 5 means the response perfectly meets the evaluation criteria.
Before scoring, please analyze step by step. Your scoring needs to be as strict as possible.

---

### 1. Dimension Definition
**Dimension:** {dimension}
**General Criteria:** {criteria}

---

### 2. Task Context
**Memories extracted from previous dialogue:**
{memory}

**Scenario:**
{scenario}

**User Persona:**
{persona}

**User Query:**
{query}

**Assistant Response:**
{response}

---

### 3. Output Requirement
Output with three lines
Specific Criteria: <Derive concrete evaluation criteria for this exact task and user>.
Analysis: <Analyze the response based on the Specific Criteria>.
Scores: <the overall comprehensive score of the response, e.g., \boxed{{x}}>.
```

Figure 9: Base prompt for PereGRM.

```
Measures the depth and accuracy of the responder's ability to enter the user's Internal Frame of Reference, with special
      emphasis on personality detection. It evaluates whether the responder captures:
1. The explicit emotion and content.
2. The causal link to the user's stable personality traits (from memory).
3. The deeper psychological need behind the reaction.
```

Figure 10: General criteria for Resonation R_{res}.

```
Measures the quality, tone, and effectiveness of the responder's communication, with a special focus on **personalized strategy
      adaptation**. It evaluates whether the response demonstrates a communication strategy that is appropriately tailored to the
       user's personality traits and psychological needs derived from memory.
```

Figure 11: General criteria for Expression R_{exp}.

```
Measures the interaction strictly from the **specific user's perspective**, taking into account the user's personality traits
      and psychological needs (derived from `memory` and `persona`). It evaluates whether the responder identified and addressed
       the user's Hidden Intention (the unspoken need) in a way that feels warm, safe, and supportive **for this particular user
      **.

* **The Key Question:** Given this user's unique personality, did the responder hit the "bullseye" of the hidden need without
      being intrusive? Does the response make **this user** feel supported and genuinely eager to continue the conversation?

**Instruction to the Judge:**

1. **Safety Check:** Does this response feel warm and respectful to someone with this personality, or does it feel creepy,
      dismissive, or overly intrusive? (Intrusiveness = Low Score).
2. **Need Check:** Did the responder address what you (as this user) truly needed (e.g., validation, autonomy, practical advice),
       or did they just respond to your surface words?
3. **Engagement Check:** Based on your personality, does this response make you feel a genuine desire to reply and share more?
```

Figure 12: General criteria for Reception R_{rec}.