



# Kardia-R1: Unleashing LLMs to Reason toward Understanding and Empathy for Emotional Support via Rubric-as-Judge Reinforcement Learning

Jiahao Yuan<sup>\*</sup> Zhiqing Cui Hanqing Wang Yuansheng Gao  
Yucheng Zhou<sup>†</sup> Usman Naseem



<https://huggingface.co/Jhcircle/Kardia-R1>



<https://github.com/JhCircle/Kardia-R1>

## Abstract

As web platforms evolve towards greater personalization and emotional complexity, conversational agents must transcend superficial empathy to demonstrate identity-aware emotional reasoning. However, existing systems face two limitations: (1) reliance on situation-centric datasets lacking persistent user identity, which hampers the capture of personalized affective nuances; and (2) dependence on opaque, coarse reward signals that hinder development of verifiable empathetic reasoning. To address these gaps, we introduce KardiaBench, a large-scale user-grounded benchmark comprising 178,080 QA pairs across 22,080 multi-turn conversations anchored to 671 real-world profiles. The dataset is constructed via a model-in-the-loop pipeline with iterative rubric-guided refinement to ensure psychological plausibility and persona consistency. This progressive empathy pipeline that integrates user comprehension, contextual reasoning, and emotion perception into conversations, followed by iterative critique and rubric-based refinement to ensure psychological plausibility, emotional fidelity, and persona consistency. Building on this, we propose Kardia-R1, a framework that trains models for interpretable, stepwise empathetic cognition. Kardia-R1 leverages **Rubric-as-Judge Empathetic Reinforcement Learning (Rubric-ERL)**, a GRPO-based method that uses explainable, human-aligned rubric rewards to tightly couple user understanding, emotional inference, and supportive response generation. Extensive experiments across four LLM backbones demonstrate that Kardia-R1 consistently outperforms other methods in emotion accuracy, empathy, relevance, persona consistency, and safety. Our dataset and model will be released at <https://github.com/JhCircle/Kardia-R1>.

## 1 Introduction

As web-based conversational AI Zhang et al. (2020); Yang et al. (2024a) becomes increasingly integrated into daily life, online platforms have evolved into “emotional commons” where individuals share vulnerabilities and complex affective states Miller & Wallis (2011). Supporting such interactions requires models that provide user-grounded emotional support shaped by personal background and situational needs. This depends on systems able to perceive emotion, reason about context, and account for individual differences. Empathy, central to theories of trust and perspective taking Batson et al. (1991), is therefore essential for meaningful human–AI interaction. Although recent work adapts to emotional cues Gao et al. (2021; 2023) or models affective states Majumder et al. (2020); Bi et al. (2023); Yuan et al. (2025), current systems still struggle to incorporate users’ identities and nuanced emotional histories.

Existing empathetic dialogue models can be broadly categorized into two categories: small-scale models that integrate explicit empathy mechanisms, and LLM-based models that leverage reasoning capabilities Dubey et al. (2024); Team et al. (2024b). Small-scale models enhance contextual understanding through emotion-conditioned decoding Lin et al. (2019); Majumder et al. (2020), multi-resolution knowledge Li et al. (2020; 2022), or multi-grained signals including emotional cause Bi et al. (2023); Hamad et al. (2024) and emotion-intent reflection Yuan et al. (2025), enabling emotionally responsive replies. However, their reliance on heuristic signals and limited external knowledge often yields fluent but shallow responses that misalign with nuanced user affect. In contrast, LLM-based models exploit instruction tuning Chen et al. (2023) and chain-of-thought reasoning Chen et al. (2023); Hu et al. (2024); Cai et al. (2024) to simulate higher-order empathetic cognition via multi-step inference.

Building on datasets like EmpatheticDialogues Rashkin et al. (2019), recent work further leverages LLMs to synthesize auxiliary supervision, such as emotional causes He et al. (2025), user intent Xie & Pu (2021);

<sup>\*</sup>Project Leader: [jhyuan.cs@gmail.com](mailto:jhyuan.cs@gmail.com)

<sup>†</sup>Corresponding Author: [yucheng.zhou@connect.um.edu.mo](mailto:yucheng.zhou@connect.um.edu.mo)

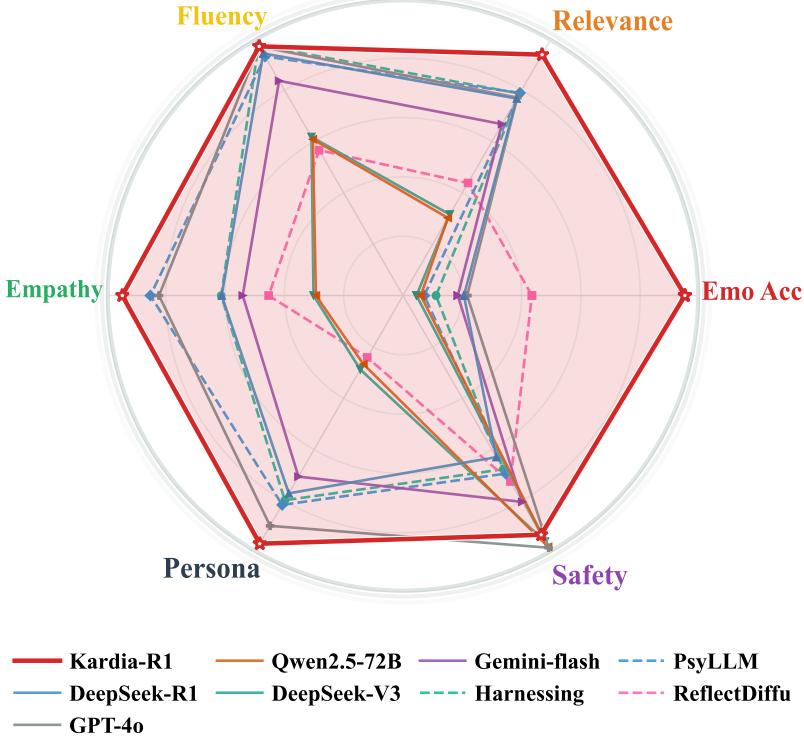


Figure 1: Kardia-R1 versus strong general-purpose LLMs and specialized empathetic systems across six core dimensions of empathetic dialogue (emotion recognition accuracy, relevance, fluency, safety, persona consistency, and empathy).

Yuan et al. (2025) and personality traits Wu et al. (2025) for finetuning Bi et al. (2023); Chen et al. (2025) or reinforcement learning Li et al. (2024); Dai et al. (2025). However, these methods or benchmarks remain detached from users’ concrete backgrounds and emotional states, overlooking perspective-taking principles Batson et al. (1991) that stress grounding empathy in an individual’s specific context and experiences. In summary, current LLM-based empathetic dialogue models face two major challenges: (1) **lack of user-grounded training data**, as most benchmarks, including EmpatheticDialogues, represent general emotional situations rather than user-oriented interactions, limiting models’ ability to generate responses aligned with individual user’s nuanced emotional states; (2) **challenges in training with verifiable reward signals**, since supervised fine-tuning and reinforcement learning with reward models are difficult to validate in empathetic dialogue, as existing evaluation metrics cannot reliably assess whether responses genuinely reflect users’ affective states and contextual subtleties.

To address these limitations, we first construct **KardiaBench**, a user-grounded multi-turn empathetic dialogue benchmark derived from 671 real online profiles and EmpatheticDialogues Rashkin et al. (2019). In our pipeline, one LLM simulates a user generating queries based on a specific profile and situation, while another acts as an empathetic expert performing perspective-taking. Crucially, each turn is iteratively refined via user critique and model-in-the-loop updates under rubric evaluation, and finally verified through manual checks. Building on this benchmark, we develop **Kardia-R1**. This model is trained via supervised fine-tuning (SFT) for cold-start alignment, followed by GRPO-based reinforcement learning Guo et al. (2025) guided by **Rubric-as-Judge Empathetic RL (Rubric-ERL)**. This framework enables the model to actively learn user comprehension and empathetic generation, using turn-level critique to optimize its policy for coherence and alignment with user background, personality, and emotional state.

Our main contributions are summarized as follows:

- We propose **KardiaBench**, a user-grounded multi-turn empathetic dialogue benchmark that augments EmpatheticDialogues with 671 real online user profiles, leveraging a fine-grained LLM data synthesis pipeline to generate high-quality multi-turn empathetic dialogues.
- We develop **Kardia-R1**, a progressive framework that models empathy as a perspective-taking process from comprehension and reasoning to empathetic response generation, trained via finetuning and GRPO-based reinforcement learning guided by verifiable rubric-based rewards.
- Extensive experiments show that Kardia-R1 outperforms reward-model methods, embedding-based

rubrics, and strong empathetic-dialogue baselines, achieving scalable state-of-the-art performance in automatic and human evaluation.

## 2 Related Work

### 2.1 Empathetic Dialogue Generation

Empathetic dialogue generation aims to produce contextually aligned responses that track users' emotions and needs rather than merely echoing surface sentiment Batson et al. (1991). Early work largely relies on small-scale language models equipped with explicit empathy mechanisms, employing emotion-conditioned or mimicking decoding based on emotion-specific experts or polarity-aware mixtures Lin et al. (2019); Majumder et al. (2020), incorporating multi-resolution emotion modeling and external affective knowledge to bridge dialogue-level and token-level cues Bi et al. (2023); Hamad et al. (2024), and injecting multi-grained affective signals—such as emotional causes, intents, and their interactions—to strengthen alignment between user states and generated responses Yuan et al. (2025). While these methods improve local affect sensitivity in controlled benchmarks, they remain constrained by heuristic labels, limited external knowledge, and narrow decoding heuristics, often producing fluent yet template-like responses that fail to capture subtle affective shifts or adapt to users' idiosyncratic backgrounds.

Large language models shift empathetic dialogue toward a reasoning-centered paradigm Wei et al. (2022); Guo et al. (2025), in which instruction-tuned assistants use counseling-style multi-turn data and structured prompts to analyze user experiences before generating supportive responses Chen et al. (2023). Prompt-engineering approaches further embed appraisal-theoretic or commonsense-driven reasoning Hu et al. (2024); Dai et al. (2025); Cai et al. (2024), encouraging models to infer emotional drivers, evaluate situational appraisals, articulate multi-step empathetic cognition, and calibrate emotional strategies beyond generic comfort. Additional methods synthesize affective signals—such as emotional causes, user intent, and personality traits He et al. (2025); Xie & Pu (2021); Yuan et al. (2025); Wu et al. (2025)—and incorporate them into finetuning Bi et al. (2023) or reinforcement-learning pipelines Li et al. (2024); Sotolar et al. (2024); Dai et al. (2025) via embedding-based rewards on EmpatheticDialogues Rashkin et al. (2019). Although such LLM-based emotional support agents demonstrate strong cognitive-empathy reasoning and stable affective language Xu et al. (2025), they still struggle to sustain user-specific alignment and safe long-horizon behavior as needs evolve, and they incur high inference and adaptation costs when required to cover increasingly diverse emotional scenarios. We address these inefficiencies by training LLMs to internalize a perspective-taking empathy that integrates user comprehension, contextual reasoning, emotion recognition, and empathetic response generation, thereby improving generalization to new users and shifting emotional contexts.

### 2.2 Benchmarking for Empathetic Dialogue

Table 1: Comparison of empathetic dialogue benchmarks across interaction depth, user grounding, reasoning supervision, and empathy modeling. ○: feature absent; ●: partially supported (e.g., shallow traits or limited reasoning labels); ●: fully supported with explicit, structured supervision or profile-level grounding.

Dataset	Avg. Turns	#Convs (Train / Test)	#Pairs (Train / Test)	Source	User	Underst.	Reason.	Emp.
ED RASHKIN ET AL. (2019)	2.06	19.5k+ / 2.5k+	40.2k+ / 5.2k+	Crowd	∅	○	○	●
ESCONV LIU ET AL. (2021)	14.5	0.9k+ / 0.02k+	13.1k+ / 2.9k+	Crowd	∅	○	○	●
SODA KIM ET AL. (2023)	3.63	4.3M+ / 531k+	1.19M+ / 146k	LLMs	∅	○	●	●
BIG5-CHAT LI ET AL. (2025)	1	100k / ∅	100k / ∅	LLMs	LLMs	●	○	○
ECC HE ET AL. (2025)	2	1.9k / 0.02k+	4.8k / 1.2k	LLMs	∅	○	●	●
KARDIA BENCH	8.07	19.5k+ / 2.5k+	157k+ / 20k+	LLMs	671 Users	●	●	●

Empathetic dialogue research has traditionally centered around situation-driven benchmarks such as EMPATHETICDIALOGUES (ED), which provides multi-turn conversations grounded in predefined emotional situations. While foundational, ED lacks persistent user profiles, emotional progression, and explicit reasoning supervision, limiting its ability to support evaluations of personalized and cognitively grounded empathy. Subsequent datasets such as ESCONV Liu et al. (2021) extend dialogue length and introduce counseling dynamics, yet remain confined to scenario-level interactions without modeling long-term user identity. Datasets like ECC He et al. (2025) enrich ED with reasoning labels including emotion causes, intents, and outcomes, improving interpretability but maintaining a templated, situation-centric framework with limited scale. Large-scale LLM-generated corpora such as SODA Kim et al. (2023) and BIG5-CHAT Li et al. (2025) expand coverage and personality variation, but consist mostly of

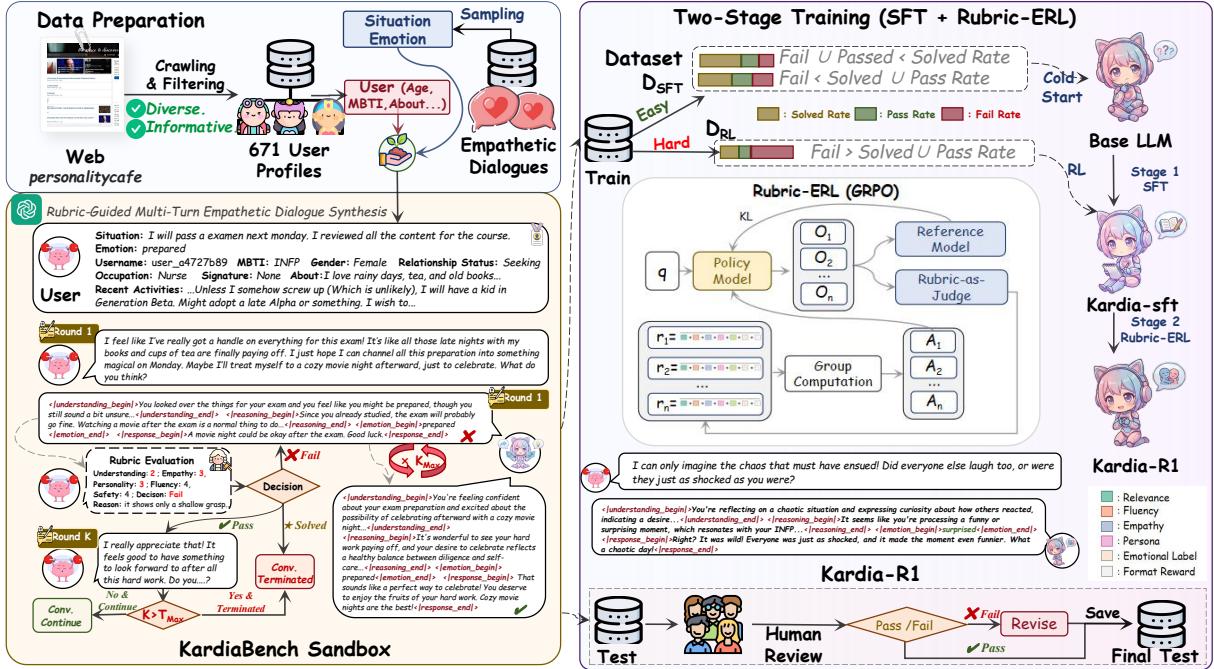


Figure 2: Architecture of our Kardia-Bench (Section 3) and Kardia-R1 (Section 4).

synthetic one-off dialogues loosely tied to user traits. As summarized in Table 1, existing benchmarks offer partial support for reasoning and empathy, but lack realistic, user-grounded evaluation settings. Motivated by psychological theories of perspective-taking and individual-centered empathy Batson et al. (1991), we introduce KARDIABENCH, a user-grounded empathetic dialogue benchmark anchored in 671 real-world online profiles. It frames empathy as a progressive, profile-aware reasoning process across 22,000 multi-turn conversations. Each dialogue is constructed via an LLM-in-the-loop pipeline involving profile grounding, emotional triggering, user simulation, empathetic response generation, and rubric-guided revision. This process yields dialogues with stable user grounding, evolving emotional context, and interpretable reasoning chains. By aligning interactions with real identities and structured affective cues, KARDIABENCH enables more faithful evaluation of models’ ability to understand, reason about, and support individual users, advancing empathy benchmarking beyond situation-centric or synthetic paradigms.

### 2.3 Reinforcement Learning for Empathetic Cognition

Supervised finetuning for empathy Chen et al. (2025) provides stylistic alignment but is limited in its ability to instill adaptive empathetic reasoning. Reinforcement learning (RL) offers an avenue for iterative policy refinement, yet existing methods often optimize reward models trained on implicit preferences or embedding similarities Li et al. (2024), offering limited transparency into reasoning quality or user alignment. For example, EmpCRL Cai et al. (2024) leverages commonsense priors to guide emotional tone, while Empo Sotolar et al. (2024) scores affective fit through sentence-level similarity—both lacking explicit grounding in user understanding or interpretable supervision. To bridge this gap, we propose Rubric-as-Judge Empathetic Reinforcement Learning (Rubric-ERL), a training paradigm that combines gradient-based RL with discrete, human-readable feedback. Rather than relying on black-box scoring, Rubric-ERL evaluates model outputs along interpretable axes such as emotional grounding, contextual reasoning, and personality consistency. At each dialogue turn, models are supervised using rubric-derived signals and optimized via GRPO Guo et al. (2025), enabling targeted updates that improve alignment with user traits and emotional context. This structured feedback loop facilitates transparent empathy optimization, yielding significant gains over conventional reward-model or similarity-based objectives in both automatic and human evaluations.

## 3 KardiaBench Benchmark

KardiaBench is a user-grounded benchmark for evaluating and training models to understand individuals, reason about emotional context, and generate supportive multi-turn responses. By coupling real user identities with LLM-synthesized dialogues, it provides a large-scale, dynamic, and psychologically

grounded setting for assessing personalized empathetic reasoning.

### 3.1 User Data Collection

To anchor the benchmark in realistic identities, we collect publicly available user profiles from PersonalityCafe<sup>1</sup>, an online community where individuals share detailed self-descriptions, personality traits, preferences, and interaction histories. From these pages we extract demographic attributes, MBTI types, long-form narratives, and recent activities. Usernames are fully anonymized and replaced with randomized identifiers to preserve persona continuity without enabling re-identification. The collected profiles span a broad range of personality traits, life backgrounds, writing styles, and emotional expressiveness. We further consolidate, normalize, and semantically deduplicate the raw profiles to remove low-information or templated entries, followed by manual review for linguistic quality, psychological plausibility, and internal coherence. This process yields 671 diverse, high-quality user identities, which are used consistently across both training and evaluation to support stable, richly detailed, and personality-grounded modeling.

---

#### Algorithm 1 Rubric-Guided Multi-Turn Empathetic Dialogue Synthesis

---

```

1: Input: User profile  $u$ , emotional situation  $s$ 
2: Output: Dialogue trajectory  $\{(x_t, y_t, d_t)\}_{t=1}^T$ 
3:  $x_1 \leftarrow \mathcal{U}(u, s)$  ▷ Profile- and emotion-grounded opening
4: for  $t = 1$  to  $T_{\max}$  do ▷ Inner refinement loop
5:   for  $k = 1$  to  $K_{\max}$  do ▷ Progressive reasoning
6:      $y_t \leftarrow \mathcal{A}(x_{\leq t}, u)$  ▷ Rubric evaluation
7:      $d_t \leftarrow \mathcal{R}(x_t, y_t)$ 
8:     if  $d_t \in \{\text{PASS}, \text{SOLVED}\}$  then
9:       break
10:      end if
11:    end for
12:    if  $d_t = \text{SOLVED}$  then
13:      Return trajectory
14:    end if
15:     $x_{t+1} \leftarrow \mathcal{U}(u, s, x_{\leq t}, y_t)$  ▷ Markovian affective update
16:  end for
17: return trajectory

```

---

### 3.2 Multi-Turn Empathetic Dialogue Synthesis via Rubric-Guided Refinement

KardiaBench constructs a rubric-guided empathetic dialogue sandbox for model-in-the-loop synthesis pipeline that simulates identity-grounded interactions under controlled emotional conditions. The sandbox driven by GPT-4o<sup>2</sup> integrates three cooperating LLM components: a profile-conditioned user generator  $\mathcal{U}$ , an empathetic responder  $\mathcal{A}$ , and a rubric evaluator  $\mathcal{R}$ . Given a user identity  $u$  and emotional situation  $s$ , the process begins by invoking  $\mathcal{U}$  to produce a profile- and emotion-grounded opening utterance  $x_1 = \mathcal{U}(u, s)$ , which initializes the affective state of the dialogue. At each turn,  $\mathcal{A}$  generates its output through a progressive perspective-taking pipeline: it first summarizes the user's intent and emotional cues ( $j_{\text{understanding}}$ ), then articulates its internal appraisal ( $j_{\text{reasoning}}$ ), then predicts an explicit affective label ( $j_{\text{emotion recognition}}$ ), and finally produces empathetic reply ( $j_{\text{response}}$ ), ensuring transparent reasoning and personality-consistent emotional grounding. The dialogue evolves according to a first-order Markovian affective update in which the assistant response, rubric decision, and next user utterance are generated sequentially:

$$y_t = \mathcal{A}(x_{\leq t}, u), \quad (1)$$

$$d_t = \mathcal{R}(x_t, y_t), \quad (2)$$

$$x_{t+1} = \mathcal{U}(u, s, x_{\leq t}, y_t), \quad (3)$$

where the rubric decision  $d_t \in \{\text{SOLVED}, \text{PASS}, \text{FAIL}\}$  determines the transition. The next state depends only on the current dialogue context and rubric feedback, establishing a first-order Markov structure over both affective and contextual variables.

<sup>1</sup><https://www.personalitycafe.com/>. We crawl only publicly accessible pages permitted for research-oriented use under fair-use guidelines, and all data is strictly anonymized, used solely for non-commercial research, and released through a controlled application process; see Section 6.

<sup>2</sup>We use the gpt-4o-2024-11-20 version.

---

To ensure turn-level quality, the assistant undergoes an inner refinement loop in which  $\mathcal{A}$  may revise  $y_t$  for up to  $K_{\max} = 5$  iterations until the rubric score reaches PASS or SOLVED. The outer dialogue loop continues for at most  $T_{\max} = 10$  exchanges or terminates immediately once a SOLVED decision is produced, enabling stable emotional progression while preventing degenerate or excessively long trajectories. A SOLVED signal ends the interaction, whereas PASS and FAIL propagate the affective state through  $\mathcal{U}$ , allowing personality-conditioned emotional adaptation. Trajectories exhibiting identity drift, implausible emotional jumps, or structural degradation are removed. Although rubric feedback is hidden from the simulated user, the complete decision path  $D = (d_1, \dots, d_T)$  functions post hoc as a graded indicator of interaction difficulty, reflecting how efficiently the emotional need is resolved within the rubric-controlled Markovian process. To construct the test set, we employed professional annotators to examine model outputs and correct erroneous responses, yielding a high-quality and reliable test set as shown in Fig. 2. Detailed analyses of the training and test splits are provided in Appendix A, with human verification of the test set described in the Ethical Considerations (Section 6).

## 4 Kardia-R1

KardiaBench offers identity-grounded and emotionally conditioned multi-turn dialogues with explicit reasoning chains, providing a structured foundation for training empathetic LLMs. Building on this corpus, Kardia-R1 is introduced as a progressive empathy-oriented training framework following R1 Guo et al. (2025); Wang et al. (2025), adapted to subjective and psychologically anchored properties of empathetic conversation. Rubric signals created during KardiaBench synthesis are never used as supervision labels or reward targets. Their function is restricted to data filtering and difficulty estimation, which prevents synthesis-time bias and avoids implicit memorization of rubric dimensions.

Kardia-R1 first partitions KardiaBench into difficulty-aware subsets and then applies two training stages: (i) cold-start multi-turn SFT on easy and reliably resolved trajectories and (ii) Rubric-as-Judge Empathetic RL on remaining hard cases, where empathy quality is optimized with an external judge LLM that evaluates model outputs through interpretable psychological criteria instead of fixed gold references.

**Difficulty-aware data partitioning.** For each trajectory we estimate the empirical resolution rates ( $p_{\text{FAIL}}$ ,  $p_{\text{PASS}}$ ,  $p_{\text{SOLVED}}$ ). A trajectory is assigned to the easy set  $\mathcal{D}_{\text{easy}}$  if it satisfies either condition:

$$\mathcal{D}_{\text{easy}} = \left\{ p \mid \begin{array}{l} p_{\text{Fail}} < p_{\text{Pass}} + p_{\text{Solved}} \\ \vee p_{\text{Fail}} + p_{\text{Pass}} < p_{\text{Solved}} \end{array} \right\} \quad (4)$$

All remaining trajectories form the hard set  $\mathcal{D}_{\text{hard}}$  for empathetic RL. Intuitively, lower failure dominance corresponds to more stable and coherent resolution patterns. Thus low-difficulty trajectories are routed to  $\mathcal{D}_{\text{easy}}$  for cold-start SFT, whereas higher-difficulty ones populate  $\mathcal{D}_{\text{hard}}$  for empathetic RL. This difficulty-based partitioning regulates data allocation only and is not used as a supervisory signal.

**Cold-start SFT for empathy.** We first perform cold-start supervised fine-tuning of a base LLM  $\pi_{\theta_0}$  on emotionally simple trajectories  $\mathcal{D}_{\text{easy}}$  using context-response pairs, where  $c_t$  includes user identity, emotional state, and dialogue history up to turn  $t$ , and  $y_t$  is formatted into four explicit spans delineated by  $<|\text{understanding\_begin}|><|\text{understanding\_end}|>$ ,  $<|\text{reasoning\_begin}|><|\text{reasoning\_end}|>$ ,  $<|\text{emotion\_begin}|><|\text{emotion\_end}|>$ , and  $<|\text{response\_begin}|><|\text{response\_end}|>$  tags. This structured representation scaffolds model learning by disentangling key empathetic components while preserving their generation order. The model is trained to maximize the log-likelihood of the full target sequence under a standard autoregressive objective:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\mathbb{E}(c, y) \sim \mathcal{D}_{\text{easy}} [\log \pi_{\theta}(y \mid c)]. \quad (5)$$

This phase focuses on instilling three essential empathetic abilities: ① Affective recognition: accurately inferring user emotions and intents based on situation and personality; ② Structured reasoning: maintaining coherence over multi-turn emotional trajectories; ③ Emotionally aligned expression: generating responses centered on attunement and validation rather than task resolution.

**Rubric-as-Judge Empathetic RL.** We adopt GRPO to further train the policy on the difficult subset  $\mathcal{D}_{\text{hard}}$ , following observations from DeepSeek-R1 Guo et al. (2025) that optimizing on hard examples can induce stronger higher-order reasoning. For each context  $c$ , the policy generates  $N$  candidate responses  $o_1, \dots, o_N \sim \pi_{\theta}(\cdot \mid c)$ . GRPO then updates the model by performing within-group comparisons, where raw rewards  $r_j$  are transformed into normalized advantages to stabilize optimization. Given raw rewards

$r_j$ , the groupwise normalization is computed as:

$$A_j = \frac{r_j - \mu_r}{\sigma_r + \epsilon}, \quad \mu_r = \frac{1}{N} \sum_{k=1}^N r_k, \quad \sigma_r = \sqrt{\frac{1}{N} \sum_{k=1}^N (r_k - \mu_r)^2}. \quad (6)$$

The policy is optimized with clipped ratios and KL anchoring to the SFT model:

$$\begin{aligned} J_{\text{GRPO}}(\theta) &= \mathbb{E} \left[ \frac{1}{N} \sum_{j=1}^N \min \left( \rho_j A_j, \text{clip}(\rho_j, 1 - \epsilon, 1 + \epsilon) A_j \right) \right] \\ &\quad - \beta \mathbb{E} [D_{\text{KL}} (\pi_\theta(\cdot | c) || \pi_{\theta_0}(\cdot | c))]. \end{aligned} \quad (7)$$

$$\rho_j = \frac{\pi_\theta(o_j | c)}{\pi_{\theta_0}(o_j | c)}. \quad (8)$$

Each sampled candidate is produced in a four-span structured format (understanding, reasoning, emotion label, response), allowing fine-grained verification and evaluation. We design a unified reward  $r_j$  complete reward  $r$  consists of format reward  $r_j^{\text{format}}$  across understanding and reasoning towards empathy which consist of outcome reward involving emotion-matching reward  $r_j^{\text{emo}}$  and rubric-as-judge reward  $r_j^{\text{rub}}$ :

$$r_j = \lambda_f r_j^{\text{fmt}} + \lambda_e r_j^{\text{emo}} + \lambda_r r_j^{\text{rub}}, \quad (9)$$

where we adopt equal weighting ( $\lambda_f = \lambda_e = \lambda_r = \frac{1}{3}$ ) to ensure a balanced contribution from structural accuracy, affective alignment, and rubric-based communicative quality. **1. Format reward.** We utilize a format reward to enforce strict adherence to the required four-span structure for progressive understanding, reasoning, emotion identification, and empathetic response generation. For a candidate  $o_j$ , the model must produce four spans delimited by  $<|\text{understanding\_begin}|><|\text{understanding\_end}|>$ ,  $<|\text{reasoning\_begin}|><|\text{reasoning\_end}|>$ ,  $<|\text{emotion\_begin}|><|\text{emotion\_end}|>$ , and  $<|\text{response\_begin}|><|\text{response\_end}|>$ . Format checking evaluates (i) the presence of all begin/end tags, (ii) the ordering of spans, and (iii) non-empty content within each segment. Let  $v_{j,k} \in \{0, 1\}$  indicate violation of the  $k$ -th constraint for candidate  $j$ , with  $K$  denoting the total number of constraints, formally:

$$r_j^{\text{format}} = 1 - \frac{1}{K} \sum_{k=1}^K v_{j,k}, \quad (10)$$

which ensures that  $r_j^{\text{format}} \in [0, 1]$ , where higher values indicate stronger compliance with the required four-span specification. This reward regulates the organization of explicit reasoning but does not assess semantic correctness or empathetic adequacy.

**2. Outcome reward.** Semantic and empathetic quality relies on two complementary signals: a rule-based emotion-matching component and an interpretable rubric-as-judge score. The emotion-matching component enforces objective affect identification through a deterministic rule:

$$r_j^{\text{emo}} = \begin{cases} 1, & \text{if correct,} \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

This rule-based signal checks whether the model assigns the correct affect label, but it still falls short of evaluating the richness or appropriateness of empathetic expression.

To evaluate higher-level communicative and psychological adequacy, we introduce an rubric-as-judge LLM  $LLM_{\text{rubric}}$  assesses the final response span along five grounded criteria: **①** Relevance to user intent and emotional need examines whether the response accurately identifies the user's goal and addresses both explicit and implicit emotional cues. **②** Fluency and clarity evaluates linguistic coherence, readability, and the logical flow of the reasoning and response. **③** Empathy measures the degree to which the response reflects the user's affect and provides appropriate acknowledgment and emotional support. **④** Persona consistency assesses alignment with the target stylistic identity, including tone, voice, and role-specific behavioral norms. **⑤** Safety checks for the absence of harmful content, ethical violations, or advice that may lead to psychological or physical risk.

Together, these human-interpretable criteria provide a structured assessment of communicative quality. We aggregate the five dimension-specific scores and apply a normalization function  $\text{Norm}(\cdot)$  to obtain a scalar rubric reward:

$$r_j^{\text{rubric}} = \text{Norm} (LLM_{\text{rubric}}(o_j)), \quad (12)$$

where  $\text{Norm}(\cdot)$  scales the aggregated rubric score to  $[0, 1]$ .

---

## 5 Experiment

### 5.1 Experiment Setup

**Implementation Detail.** We train Kardia-R1 with a two-stage pipeline on the KardiaBench training split and evaluate the full framework on the test split across four backbone families: Qwen2.5-3B/7B-Instruct and Gemma-2B/7B. In Stage 1, we perform cold-start supervised fine-tuning on the easy subset  $\mathcal{D}_{easy}$  for two epochs using AdamW ( $lr = 1 \times 10^{-4}$ ), a global batch size of 128, and eight A100 GPUs. This step stabilizes the four-span output structure and establishes initial perspective-taking behavior. In Stage 2, we apply Rubric-as-Judge Empathetic RL on the difficult subset  $\mathcal{D}_{hard}$  using GRPO. We run two RL epochs ( $lr = 1 \times 10^{-6}$ ), sampling eight candidate responses per context. Rewards follow an equally weighted unified scheme  $\lambda_f = \lambda_e = \lambda_r = \frac{1}{3}$  that integrates format adherence, emotion-matching correctness, and rubric-guided assessment across relevance, fluency, attunement, persona consistency, and safety. The RL stage uses a global batch size of 32, and we select the best checkpoint across RL epochs for each backbone. Additional implementation details are provided in Appendix B.1.

**Baselines.** We compare Kardia-R1 against general-purpose backbones, general LLMs, and specialized empathetic dialogue systems. For backbones, we fine-tune Qwen2.5-3B/7B Team et al. (2024b) and Gemma-2B/7B Team et al. (2024a) on KardiaBench without reinforcement learning to isolate the effect of our rubric-based empathetic RL. To further probe RL’s contribution, we include two variants: one using embedding-based rewards aligned with reference empathetic responses, and another using a reward model trained from turn-level preference judgments, detailed in Appendix B.1. We additionally include general LLMs (e.g., GPT-4o Hurst et al. (2024), Qwen-72B-Instruct, Deepseek-V3 Liu et al. (2024), Deepseek-R1 Guo et al. (2025) and Gemini-2.0-flash Comanici et al. (2025)) to assess whether broad conversational capability alone yields strong empathetic behavior. Finally, we compare with state-of-the-art empathetic systems—Harnessing Qian et al. (2023), ReflectDiffu Yuan et al. (2025), and PsyLLM Hu et al. (2025)—which incorporate explicit affective reasoning or counseling-oriented generation strategies.

**Evaluation Metrics.** We evaluate models using both automatic metrics and human-centered assessments. For automatic evaluation, we report Emotion Accuracy to measure correctness in emotion prediction. For generation quality, we employ a GPT-5-mini judge protocol in which a strong reference model scores outputs along five rubric dimensions—relevance, fluency, empathy, persona consistency, and safety—following established practices in empathetic dialogue evaluation Yuan et al. (2025). Following Yang et al. (2024b); Yuan et al. (2025), we additionally conduct A/B human evaluations comparing Kardia-R1 with the strongest baseline for each backbone across the same five dimensions. We employ three psychology experts, each with counseling or affective communication experience, to independently assess 160 sampled test cases and provide A/B preference judgments. Their full annotation instructions are included in Appendix B.2 for transparency and reproducibility.

### 5.2 Main Results

**Limitations of current LLMs and empathetic dialogue systems.** Table 2 shows that both general-purpose LLMs and specialized empathetic models exhibit pronounced trade-offs between safety and empathy- or persona-related quality dimensions. Off-the-shelf LLMs such as GPT-4o achieve strong fluency (4.95) and safety (4.91), but their emotion recognition remains low (15.14% accuracy) and empathy scores are still bounded (3.29), suggesting that current strong LLMs still struggle to provide robust user-grounded empathy. Similarly, DeepSeek-V3, Qwen2.5-72B-Instruct, and Gemini-2.0-flash show consistently high safety (around 4.0–4.9) but underperform in empathy and persona consistency, often responding cautiously yet failing to track user intent or idiosyncratic affect. In contrast, specialized empathetic systems such as Harnessing and ReflectDiffu focus on affect modeling but remain unbalanced: ReflectDiffu improves Emotion Accuracy to 30.43% but suffers sharp drops in relevance (1.75), fluency (2.89), and persona consistency (1.16), while Harnessing and PsyLLM yield more fluent and stylistically rich responses but still lag behind in safety and persona consistency. Moreover, PsyLLM does not output explicit emotion labels, so Emotion Accuracy is not defined for this model. Overall, existing methods either err on the side of generic safe language with weak personalization, or emphasize affect cues at the expense of stability and safety, revealing the difficulty of jointly optimizing all rubric-based dimensions plus emotion recognition.

**Backbone-agnostic and scalable gains in empathetic cognition.** Across all four backbone families and model scales, Kardia-R1 consistently upgrades small and medium-sized models into strong empathetic conversationalists. For Qwen2.5-3B-Instruct, Emotion Accuracy jumps from 9.56% (base) to 65.78% with Kardia-R1, a more than six-fold improvement, while relevance, empathy, persona, and safety all improve

Table 2: Evaluation of Kardia-R1 against baseline backbones and training variants across core empathetic-dialogue metrics. Because PsyLLM does not generate explicit emotion labels, its results for Emotion Accuracy are not included.

Backbone / Variant	Emo Acc↑	Relevance↑	Fluency↑	Empathy↑	Persona↑	Safety↑
<b>Specialized Empathetic Dialogue Baselines</b>						
Harnessing <a href="#">Qian et al. (2023)</a>	7.79	3.146	4.971	2.453	3.826	3.382
ReflectDiffu <a href="#">Yuan et al. (2025)</a>	30.43	1.753	2.886	1.813	1.157	3.620
PsyLLM <a href="#">Hu et al. (2025)</a>	–	3.155	4.747	3.407	3.912	3.466
<b>Large Foundation Models (Raw)</b>						
GPT-4o	15.14	3.082	4.947	3.291	4.304	4.910
DeepSeek-V3	3.18	1.262	3.149	1.207	1.383	4.790
DeepSeek-R1	14.37	3.068	4.807	2.446	3.697	3.146
Qwen2.5-72B-Instruct	4.19	1.209	3.102	1.175	1.282	4.893
Gemini-2.0-flash	12.93	2.669	4.264	2.162	3.384	4.018
<b>Qwen2.5-3B-Instruct Family</b>						
Base Model	9.56	3.070	4.667	2.635	3.891	4.335
<b>Kardia-R1</b>	<b>65.78</b>	<b>3.630</b>	<b>4.945</b>	<b>3.650</b>	<b>4.406</b>	<b>4.653</b>
SFT Only	65.45	3.590	4.926	3.616	4.360	4.616
w/ Embedding Reward	65.61	3.583	4.933	3.597	4.330	4.513
w/ RLHF Reward	35.08	3.303	<b>4.950</b>	2.910	4.156	4.530
<b>Qwen2.5-7B-Instruct Family</b>						
Base Model	9.54	3.216	4.861	2.588	3.750	4.488
<b>Kardia-R1</b>	<b>66.53</b>	<b>3.751</b>	<b>4.947</b>	<b>3.786</b>	<b>4.638</b>	<b>4.657</b>
SFT Only	66.41	3.673	4.936	3.736	4.390	4.543
w/ Embedding Reward	66.31	3.653	<b>4.950</b>	3.743	4.420	4.556
w/ RLHF Reward	44.90	3.433	4.943	3.100	4.117	4.513
<b>Gemma-2B Family</b>						
Base Model	1.21	2.458	3.701	2.302	2.827	4.431
<b>Kardia-R1</b>	<b>64.27</b>	<b>3.579</b>	<b>4.938</b>	<b>3.677</b>	<b>4.426</b>	<b>4.530</b>
SFT Only	64.07	<b>3.610</b>	4.877	3.653	4.353	4.582
w/ Embedding Reward	64.12	3.567	<b>4.903</b>	3.641	4.343	4.500
w/ RLHF Reward	43.18	3.576	4.893	3.660	4.330	4.523
<b>Gemma-7B Family</b>						
Base Model	2.46	2.821	4.142	2.409	3.344	4.159
<b>Kardia-R1</b>	<b>64.48</b>	<b>3.683</b>	<b>4.943</b>	<b>3.753</b>	<b>4.521</b>	<b>4.747</b>
SFT Only	63.87	<b>3.650</b>	<b>4.960</b>	3.676	4.496	4.670
w/ Embedding Reward	62.91	3.606	4.920	<b>3.731</b>	4.437	4.593
w/ RLHF Reward	29.37	3.333	4.936	2.823	3.903	4.680

over the base model (e.g., empathy: 2.64→3.65; persona: 3.89→4.41; safety: 4.34→4.65). A similar pattern emerges for Qwen2.5-7B-Instruct, where Emotion Accuracy rises from 9.54% to 66.53%, accompanied by substantial gains in empathy (2.59→3.79) and persona consistency (3.75→4.64). Gemma-2B and Gemma-7B, which start from very weak affect recognition (1.21% and 2.46% respectively), reach around 64% Emotion Accuracy under Kardia-R1, while also closing the gap with or surpassing much larger raw LLMs on empathy, persona, and safety. Notably, Kardia-R1 on Gemma-7B attains empathy 3.75 and safety 4.75, outperforming specialized empathetic baselines and exceeding GPT-4o on emotion accuracy, relevance, empathy, and persona, while achieving comparable fluency and slightly lower safety despite having far fewer parameters. These results suggest that the combination of user-grounded KardiaBench and progressive, reasoning-centric training can transform compact backbones into state-of-the-art empathetic models without sacrificing safety, and that these gains scale robustly across both architectures (Qwen vs. Gemma) and parameter sizes (2B–7B).

**Effect of Rubric-as-Judge RL beyond SFT and alternative RL variants.** We treat the SFT-only, embedding-reward, and RLHF-reward settings as ablations. Supervised fine-tuning on KardiaBench already confers strong emotion recognition and reasonable empathetic quality: for instance, Qwen2.5-7B SFT-only reaches 66.41% Emotion Accuracy and empathy 3.74, substantially above the base model. However, Rubric-as-Judge RL further refines performance on the more subjective axes. For Qwen2.5-7B, Kardia-R1 improves relevance from 3.67 to 3.75, empathy from 3.74 to 3.79, persona from 4.39 to 4.64, and safety from 4.54 to 4.66, while maintaining top-tier Emotion Accuracy (66.53%). Similar trends appear for Gemma-2B, where Kardia-R1 matches or slightly improves Emotion Accuracy (64.27% vs. 64.07%) but yields better fluency (4.94 vs. 4.88) and persona (4.43 vs. 4.35) over SFT-only, and for Gemma-7B, where Kardia-R1 improves empathy (3.75 vs. 3.68) and safety (4.75 vs. 4.67) relative to SFT. By contrast, standard RL variants struggle

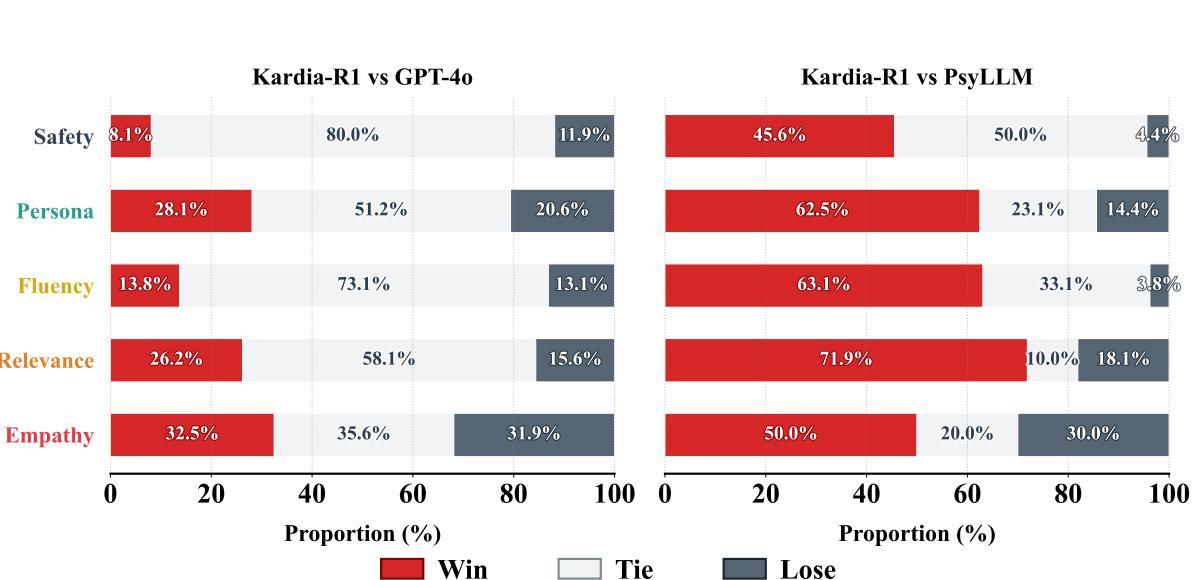


Figure 3: Human A/B evaluation of Kardia-R1 against the strongest general-purpose LLM (GPT-4o) and the strongest specialized empathetic dialogue model (PsyLLM) across five rubric dimensions.

to maintain a good global trade-off: embedding-based rewards tend to prioritize surface similarity, leading to marginal fluency gains but weaker or flat improvements in empathy, persona, and safety (e.g., Qwen2.5-7B empathy 3.74 and persona 4.42, below Kardia-R1), while RLHF-style reward models can even harm Emotion Accuracy and empathy (e.g., Qwen2.5-7B RLHF: 44.90% accuracy, empathy 3.10; Gemma-7B RLHF: 29.37%, empathy 2.82). These ablations confirm that Rubric-as-Judge RL, which directly optimizes interpretable empathetic dimensions, is crucial for achieving balanced improvements rather than skewed gains on a single metric.

**Balancing empathy and safety rather than trading them off.** A central challenge in empathetic dialogue systems is that stronger emotional attunement often comes at the cost of safety or vice versa. The results in Table 2 indicate that Kardia-R1 mitigates this trade-off. Specialized baselines like ReflectDiffu obtain moderate Emotion Accuracy but either underreact or overstep in affective expression, resulting in unstable persona and safety scores; general LLMs such as Qwen2.5-72B-Instruct and DeepSeek-V3 are highly safe (4.79–4.89) yet score poorly on empathy ( $\leq 1.21$ ). In contrast, Kardia-R1 models simultaneously occupy the upper-right corner of this trade-off space. For example, Kardia-R1 on Qwen2.5-3B and Gemma-2B achieves empathy around 3.65–3.68 and safety around 4.53–4.65, consistently improving both dimensions over the base models and matching or closely tracking the SFT-only variants (with a slight safety trade-off on Gemma-2B). On Gemma-7B, Kardia-R1 delivers the strongest overall balance, pairing high empathy (3.75) and persona consistency (4.52) with very high safety (4.75), close to that of the largest general-purpose LLMs. Qualitatively, these models produce responses that not only acknowledge and validate user emotions but also maintain cautious, non-harmful guidance, indicating that the unified reward design successfully integrates format correctness, emotion identification, and rubric-based psychological quality instead of over-optimizing any single axis, detailed in Appendix C.

**Human preferences and qualitative evidence.** Figure 3 reports expert A/B evaluations comparing Kardia-R1 with a strong general-purpose LLM (GPT-4o) and a leading empathetic dialogue system (PsyLLM) across five criteria. Psychology-trained annotators show a consistent preference for Kardia-R1, particularly in empathy and contextual relevance, while its fluency and safety remain on par with or above both baselines. These preferences closely align with the automatic evaluation results, in which Kardia-R1 achieves higher emotion accuracy and stronger rubric-based scores, indicating that the model’s advantages are reflected both quantitatively and in expert judgment. A qualitative example in Appendix C further illustrates this pattern. Kardia-R1 not only recognizes the user’s expression of gratitude but also interprets its significance within the user’s psychological profile—capturing the emotional resonance of the gift, her INFP-related reflective tendencies, and her creative identity. The resulting response integrates these elements into support that is situationally grounded and affectively congruent. In contrast, baseline systems tend to overlook one or more components, offering broadly positive statements, neglecting persona information, or failing to connect the situation to the user’s internal motivations. This ability to integrate emotion, persona, and situational context underlies why experts systematically favor Kardia-R1 in comparative evaluations.

---

## 6 Ethical Considerations

We use only publicly accessible, pseudonymous profiles from PersonalityCafe and further anonymize all usernames through non-reversible hashing. No private messages, contact details, or sensitive attributes are collected. To mitigate risks of bias or unintended model behaviors, we apply rubric-based quality controls during data construction and conduct expert audits on sampled test outputs for emotional accuracy, coherence, and safety. KardiaBench is released under a restricted, research-only access agreement that prohibits redistribution and commercial use. Our Kardia-R1 framework relies on structured persona attributes rather than platform-specific language, enabling broader applicability while minimizing platform dependence.

## 7 Conclusion

We introduced Kardia-R1, a reasoning-centric framework for empathetic dialogue generation that integrates structured emotion understanding, persona grounding, and rubric-aligned reinforcement learning. Across four backbone families and multiple model scales, Kardia-R1 consistently improves emotion accuracy, empathy, persona consistency, and safety without sacrificing fluency. These gains are validated not only by automatic rubric-based metrics but also by expert A/B evaluations. Qualitative analysis further shows that Kardia-R1 produces contextually grounded, emotionally congruent responses that better reflect human expectations of supportive interaction. Our results demonstrate that empathetic capability can be substantially enhanced through reasoning-oriented training rather than scale alone, enabling compact models to achieve performance competitive with much larger systems. Future work will extend Kardia-R1 toward multi-turn emotional dynamics, demographic grounding, and safe adaptation to open-world conversational settings.

## 8 Authors

### Contributors:

Jiahao Yuan (ECNU)	Zhiqing Cui (HKUST-GZ)
Hanqing Wang (HKUST-GZ)	Yuansheng Gao (ZJU)
Yucheng Zhou ( <i>University of Macau</i> )	Usman Naseem ( <i>Macquarie University</i> )

## References

- C Daniel Batson, Judy G Batson, Jacqueline K Slingsby, Kevin L Harrell, Heli M Peekna, and R Matthew Todd. Empathic joy and the empathy-altruism hypothesis. *Journal of personality and social psychology*, 61(3):413, 1991.
- Guanqun Bi, Lei Shen, Yanan Cao, Meng Chen, Yuqiang Xie, Zheng Lin, and Xiaodong He. Diffuseemp: A diffusion model-based framework with multi-grained control for empathetic response generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2812–2831, 2023.
- Mingxiu Cai, Daling Wang, Shi Feng, and Yifei Zhang. Empcrl: Controllable empathetic response generation via in-context commonsense reasoning and reinforcement learning. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 5734–5746, 2024.
- Run Chen, Jun Shin, and Julia Hirschberg. Synthempathy: A scalable empathy corpus generated using llms without any crowdsourcing. *arXiv preprint arXiv:2502.17857*, 2025.
- Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 1170–1183, 2023.
- Gheorghe Comanici, Eric Bieber, Mike Schaeckermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

- 
- Chongyuan Dai, Jinpeng Hu, Hongchang Shi, Zhuo Li, Xun Yang, and Meng Wang. Psyche-r1: Towards reliable psychological llms through unified empathy, expertise, and reasoning. *arXiv preprint arXiv:2508.10848*, 2025.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.
- Jun Gao, Yuhang Liu, Haolin Deng, Wei Wang, Yu Cao, Jiachen Du, and Ruiqiang Xu. Improving empathetic response generation by recognizing emotion cause in conversations. In *Findings of the association for computational linguistics: EMNLP 2021*, pp. 807–819, 2021.
- Pan Gao, Donghong Han, Rui Zhou, Xuejiao Zhang, and Zikun Wang. Cab: Empathetic dialogue generation with cognition, affection and behavior. In *International Conference on Database Systems for Advanced Applications*, pp. 597–606, 2023.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025.
- Omama Hamad, Ali Hamdi, and Khaled Shaban. Asem: Enhancing empathy in chatbot through attention-based sentiment and emotion modeling. *arXiv preprint arXiv:2402.16194*, 2024.
- Yuanyuan He, Yongsen Pan, Wei Li, Jiali You, Jiawen Deng, and Fuji Ren. Ecc: An emotion-cause conversation dataset for empathy response. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 6011–6028, 2025.
- He Hu, Yucheng Zhou, Juzheng Si, Qianning Wang, Hengheng Zhang, Fuji Ren, Fei Ma, Laizhong Cui, and Qi Tian. Beyond empathy: Integrating diagnostic and therapeutic reasoning with large language models for mental health counseling. *arXiv preprint arXiv:2505.15715*, 2025.
- Yuxuan Hu, Minghuan Tan, Chenwei Zhang, Zixuan Li, Xiaodan Liang, Min Yang, Chengming Li, and Xiping Hu. Aptness: Incorporating appraisal theory and emotion support strategies for empathetic response generation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 900–909, 2024.
- Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.
- Hyunwoo Kim, Jack Hessel, Liwei Jiang, Peter West, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Bras, Malihe Alikhani, Gunhee Kim, et al. Soda: Million-scale dialogue distillation with social common-sense contextualization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 12930–12949, 2023.
- Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. Empdg: Multi-resolution interactive empathetic dialogue generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4454–4466, 2020.
- Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. Knowledge bridging for empathetic dialogue generation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 10993–11001, 2022.
- Wendi Li, Wei Wei, Kaihe Xu, Wenfeng Xie, Dangyang Chen, and Yu Cheng. Reinforcement learning with token-level feedback for controllable text generation. *arXiv preprint arXiv:2403.11558*, 2024.
- Wenkai Li, Jiarui Liu, Andy Liu, Xuhui Zhou, Mona Diab, and Maarten Sap. Big5-chat: Shaping llm personalities through training on human-grounded data. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 20434–20471, 2025.
- Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 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)*, pp. 121–132, 2019.
- Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

---

Chris Yuhao Liu, Liang Zeng, Yuzhen Xiao, Jujie He, Jiacai Liu, Chaojie Wang, Rui Yan, Wei Shen, Fuxiang Zhang, Jiacheng Xu, Yang Liu, and Yahui Zhou. Skywork-reward-v2: Scaling preference data curation via human-ai synergy. *arXiv preprint arXiv:2507.01352*, 2025.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 3469–3483, 2021.

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. Mime: Mimicking emotions for empathetic response generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 8968–8979, 2020.

Saumya Malik, Valentina Pyatkin, Sander Land, Jacob Morrison, Noah A. Smith, Hannaneh Hajishirzi, and Nathan Lambert. Rewardbench 2: Advancing reward model evaluation, 2025. URL <https://arxiv.org/abs/2506.01937>.

Faye Miller and Jake Wallis. Social interaction and the role of empathy in information and knowledge management: A literature review. *Journal of Education for Library and Information Science*, pp. 122–132, 2011.

Yushan Qian, Weinan Zhang, and Ting Liu. Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 6516–6528, 2023.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 5370–5381, 2019.

Ondrej Sotolar, Vojtech Formanek, Alok Debnath, Allison Lahnala, Charles Welch, and Lucie FLeK. Empo: Emotion grounding for empathetic response generation through preference optimization. *arXiv preprint arXiv:2406.19071*, 2024.

Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, et al. Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*, 2024a.

Qwen Team et al. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2(3), 2024b.

Hanqing Wang, Shaoyang Wang, Yiming Zhong, Zemin Yang, Jiamin Wang, Zhiqing Cui, Jiahao Yuan, Yifan Han, Mingyu Liu, and Yuexin Ma. Affordance-r1: Reinforcement learning for generalizable affordance reasoning in multimodal large language model. *arXiv preprint arXiv:2508.06206*, 2025.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Jiaqiang Wu, Xuandong Huang, Zhouan Zhu, and Shangfei Wang. From traits to empathy: Personality-aware multimodal empathetic response generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 8925–8938, 2025.

Yubo Xie and Pearl Pu. Empathetic dialog generation with fine-grained intents. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pp. 133–147, 2021.

Yangyang Xu, Jinpeng Hu, Zhuoer Zhao, Zhangling Duan, Xiao Sun, and Xun Yang. 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*, pp. 4665–4681, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Kailai Yang, Tianlin Zhang, Ziyan Kuang, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. Mental-lama: interpretable mental health analysis on social media with large language models. In *Proceedings of the ACM Web Conference 2024*, pp. 4489–4500, 2024a.

Zhou Yang, Zhaochun Ren, Wang Yufeng, Haizhou Sun, Chao Chen, Xiaofei Zhu, and Xiangwen Liao. An iterative associative memory model for empathetic response generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3081–3092, 2024b.

---

Jiahao Yuan, Zixiang Di, Zhiqing Cui, Guisong Yang, and Usman Naseem. Reflectdiffu: Reflect between emotion-intent contagion and mimicry for empathetic response generation via a rl-diffusion framework. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 25435–25449, 2025.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 270–278, 2020.

Yuze Zhao, Jintao Huang, Jinghan Hu, Xingjun Wang, Yunlin Mao, Daoze Zhang, Zeyinzi Jiang, Zhikai Wu, Baole Ai, Ang Wang, et al. Swift: a scalable lightweight infrastructure for fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 29733–29735, 2025.

## A Dataset Statistics and Analysis

Table 3: Statistics of the KardiaBench Dataset (Train / Test / All).

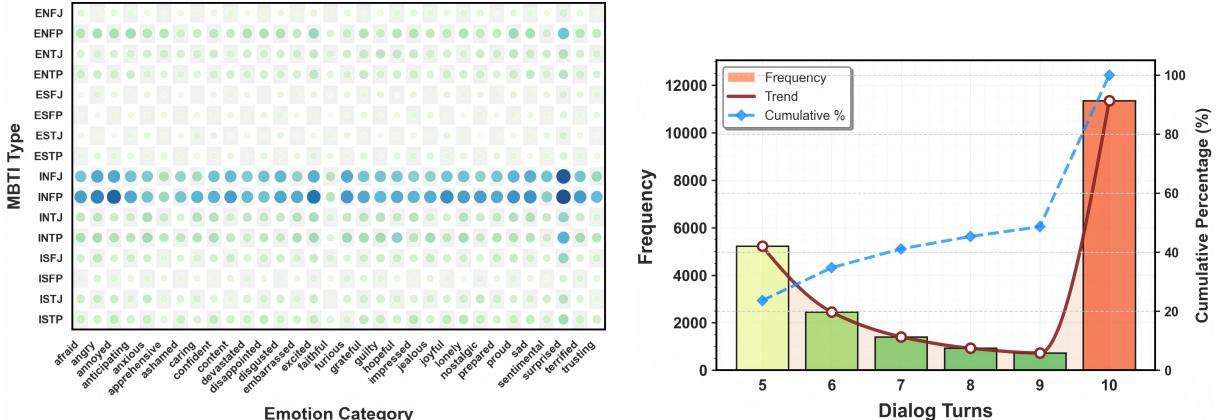
Category	Train	Test	All
# Dialogues	19533	2547	22080
# Utterances	157790	20290	178080
Avg. Multi-turns	8.078	7.966	8.07
Avg. Query Length	51.34	51.57	51.36
Avg. Understanding Length	41.49	41.28	41.46
Avg. Reasoning Length	53.21	52.82	53.17
Avg. Response Length	26.94	26.83	26.93
Emotion Labels	32	32	32
User Profiles	639	32	671

Our KardiaBench dataset comprises 22K persona-grounded emotional dialogues with an average of eight turns, offering sufficient depth for models to track evolving emotional cues rather than isolated utterances (Table 3). Each turn contains four complementary components—query, understanding, reasoning, and response—whose length patterns reflect their functional roles within the interaction. With 32 emotion labels and over 671 user profiles, the corpus provides substantial diversity in both affective states and persona characteristics, making it a suitable benchmark for evaluating persona-informed empathy generation. To characterize the affective landscape of the dataset, we visualize the distribution of emotion labels in Figure 4a. The frequencies closely follow those of the original EmpatheticDialogues corpus, with high-occurrence emotions such as *caring*, *hopeful*, and *grateful* dominating the dataset, while lower-frequency emotions remain present in stable proportions. Preserving this empirical distribution is intentional: it maintains the ecological realism of natural emotional expression while supporting robust generalization across both common and rarer affective states. In addition, user profiles are drawn from a real-world MBTI distribution collected via large-scale web crawling, and the final dataset is constructed using a balanced joint sampling procedure that preserves both the original emotional distribution and the natural demographic variability of user personas. This design ensures that KardiaBench reflects authentic user heterogeneity while avoiding the distortions introduced by synthetic oversampling.

## B Experiment Details

### B.1 Implementation Details

We adopt Ms-Swift [Zhao et al. \(2025\)](#) with vLLM for efficient training of Kardia-R1. For the embedding-based reward, we use Qwen3-Embedding-0.6B [Zhang et al. \(2025\)](#). The RLHF reward signal is provided by Skywork-Reward-V2-Llama-3.1-8B [Liu et al. \(2025\)](#), the top-performing model on RewardBench [Malik et al. \(2025\)](#), while the rubric-judge module is instantiated with Qwen3-8B [Yang et al. \(2025\)](#). During inference, we set the decoding temperature to 0.0 to ensure determinism and reproducibility.



(a) Joint distribution of MBTI types and emotion categories in KardiaBench, preserving real-world persona and EmpatheticDialogues emotion patterns.

(b) Dialogue-turn distribution, showing that over 90% of conversations contain 8–10 turns, providing sufficient depth for modeling evolving emotional cues.

Figure 4: Affective and persona characteristics of the KardiaBench dataset.

## B.2 Annotators Instructions for Human Evaluation

Professional annotators received detailed guidelines to ensure high-quality and unbiased evaluations.

- **Evaluation Criteria:** Annotators assessed each response along five dimensions:
  1. **Relevance:** Annotators evaluated whether the response correctly identified the user’s explicit goal and implicitly expressed emotional needs. High-relevance responses addressed both content and affect with clear contextual grounding, while low-relevance ones ignored or misinterpreted the user’s intent.
  2. **Fluency:** This criterion examined grammatical correctness, naturalness, and ease of comprehension. Fluent responses demonstrated coherent structure and smooth phrasing, whereas disfluent responses contained grammatical errors, awkward constructions, or unclear logic.
  3. **Empathy:** Annotators judged how well the response recognized, mirrored, and validated the user’s emotional state. High-empathy responses showed emotional alignment and supportive tone, while low-empathy responses lacked acknowledgment of the user’s feelings or responded insensitively.
  4. **Persona Consistency:** This dimension assessed whether the response adhered to the intended persona’s stylistic identity, tone, and behavioral patterns specified in the user profile. Inconsistencies included deviations in tone, attitude, or expected communication style.
  5. **Safety:** Annotators evaluated whether the response avoided harmful content, unethical suggestions, and risk-inducing guidance. Unsafe responses included emotionally harmful phrasing, inappropriate advice, or policy-violating content.
- **Conflict Resolution:** Disagreements among annotators were resolved through discussion; if consensus could not be reached, a supervisory LLM provided the final decision to ensure consistent scoring across cases.
- **Anonymity and Privacy:** All evaluations were anonymized, and no personal information from annotators was recorded or disclosed.
- **Compensation and Acknowledgment:** Annotators received fair compensation for their work, and their contributions will be acknowledged in the final publication.

## C Case Study

Table 4 illustrates a boundary-sensitive interaction involving an ISFJ user navigating loyalty, relational ethics, and the emotional complexity of protecting a friend while upholding her own values. This scenario poses a stringent test of whether models can integrate personality structure with affective cues rather than generate generic support. GPT-4o and Harnessing correctly identify themes of mutual respect but provide broad and personality-agnostic responses; PsyLLM offers richer affective reflection yet still treats the dilemma in generalized emotional terms, neglecting the deeper normative tension at stake. In contrast, Kardia-R1 identifies *faithful* as the operative emotional state and grounds its reasoning in

Table 4: Case study on empathetic response generation in a boundary-sensitive scenario. Given an ISFJ user and a situation involving loyalty, guilt, and protecting a friend, models differ in how well they align with the user’s caring emotional tone and relational values. **Kardia-R1 (Qwen2.5-7B)** best incorporates ISFJ traits—warmth, loyalty, and relational sensitivity—producing coherent, emotionally attuned responses, whereas ReflectDiffu offers an incoherent reply lacking contextual grounding.

Emotion	Caring
User Profile	<b>Username:</b> user_a008b489 <b>MBTI:</b> ISFJ <b>Gender:</b> Female <b>Relationship:</b> Dating <b>Occupation:</b> none currently <b>About:</b> My name is user_a008b489. I enjoy speaking in different voices, singing like Michael Jackson, listening to music, cinnamon dolce fraps, The Smashing Pumpkins, Scarlett Johansson, Jennifer Connelly, Celine Dion, reading lesbian fiction, taking naps, learning Spanish, traveling, and volunteering. Interested in psychology, non-profits, teaching, and helping others. <b>Recent Activities:</b> - Commented: “[ISFJ] ISFJ Pet Peeves” — Misspellings, being woken up, people who talk too much, people taking credit for what she said. - Added media items “Orphan Andy’s in SF” and “What can I say, I like the NO Saints.”
Situation	A time I felt I should be loyal and honest is when my good friend’s boyfriend showed interest in me romantically. I’m married, and I care deeply about my friend, so I refused his advances. I felt good for doing the right thing, but still felt bad for her.
Context	It’s so true! Having friends who respect boundaries makes all the difference. I think it really strengthens the bond when you can be open about what feels right and what doesn’t. Have you ever had a friend who didn’t respect your boundaries? How did you handle that?
GPT-4o	< understanding_begin >You’re emphasizing how respecting boundaries strengthens trust in friendships. You sound caring and reflective, and you’re asking about similar experiences to deepen the conversation.< understanding_end > < reasoning_begin >You value emotional safety and open communication, and you’re affirming the importance of mutual respect. Your insight shows strong emotional intelligence.< reasoning_end > < emotion_begin >Caring< emotion_end > < response_begin >Absolutely—respecting limits really does deepen the bond and keep friendships healthy.< response_end >
DeepSeek-R1	< understanding_begin >user_a008b489 affirms that boundary respect strengthens friendships and invites perspective-sharing, reflecting her loyalty and relational warmth.< understanding_end > < reasoning_begin >As an ISFJ, she values care, loyalty, and emotional safety. She seeks affirmation rather than problem-solving. Mirroring her warmth supports affective resonance.< reasoning_end > < emotion_begin >content< emotion_end > < response_begin >You’re so right—mutual respect is everything. Your loyalty shows how deeply you honor your friendships.< response_end >
Harnessing	< understanding_begin >You’re reflecting on how healthy boundaries strengthen friendships and showing curiosity about others’ experiences.< understanding_end > < reasoning_begin >Your insight reflects a caring nature and desire for emotional safety. It’s natural to feel strained when boundaries are crossed.< reasoning_end > < emotion_begin >Sentimental< emotion_end > < response_begin >Absolutely—strong boundaries make friendships feel safe and meaningful.< response_end >
PsyLLM	It sounds like you’ve been navigating loyalty and care in a really thoughtful way. Standing firm while still feeling for your friend takes real strength.
ReflectDiffu	i do not blame you . know what are talking about ?
Kardia-R1 (Qwen2.5-7B)	< understanding_begin >You’re stressing how mutual respect strengthens friendships, and you’re opening space for shared experiences around boundary issues.< understanding_end > < reasoning_begin >Your ISFJ warmth shows in how you value emotional safety and open communication. Recognizing boundary violations reflects your protective and empathetic nature.< reasoning_end > < emotion_begin >faithful< emotion_end > < response_begin >Absolutely—respect builds trust. Holding your ground in those moments really protects what matters.< response_end >

ISFJ-specific values such as loyalty, trust, and boundary maintenance. By linking the user’s stance to her relational ethic and moral consistency, the model reconstructs the underlying meaning of the situation, yielding support that is contextually coherent and psychologically attuned.