From Simulation to Diagnosis: Simulation-Based Reinforcement Learning for Empathetic Depression Diagnosis

Anonymous ACL submission

Abstract

Given the rising prevalence of depression alongside persistently low diagnosis rates, Large Language Models (LLMs) offer an opportunity for accessible mental health assessment. However, clinical depression diagnosis is inherently a goal-oriented, sequential, and interactive process. Current approaches primarily rely on static supervised finetuning (SFT) using fixed depression-diagnosis dialogue datasets, which are ill-suited to the dynamic nature of real-world diagnosis interactions. Models trained on such static data often struggle to navigate the variability of patient behavior, failing to provide diagnostic accuracy. To overcome these limitations, we introduce **SimRED** (**Sim**ulation-based **R**einforcement Learning for **E**mpathetic Depression **D**iagnosis), a framework that trains depression diagnostic agents via reinforcement learning through extensive interactions with patient simulator. SimRED constructs a patient simulation environment derived from real-world dialogues. By interacting with these simulated patients, the diagnostic agent employs reinforcement learning to learn both diagnostic accuracy and empathetic expression. Experimental results demonstrate that SimRED significantly outperforms existing strong baselines in both diagnostic accuracy and the quality of empathetic expression.

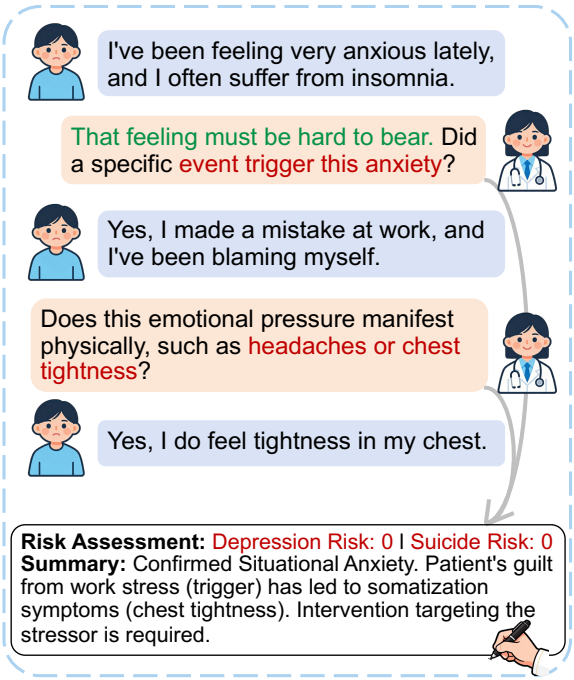

Figure 1: Illustration of the depression-diagnosis dialogues. Unlike casual conversation, each turn in a depression-diagnosis dialogue serves an information-gathering purpose toward the final assessment. The agent dynamically balances empathetic expression with diagnostic inquiries. This demonstrates the necessity of moving beyond static imitation to a system that learns to achieve high diagnostic accuracy through iterative interactions with the patient simulator.

1 Introduction

Depression affects approximately 264 million individuals globally, posing a severe public health challenge (James et al., 2018; Ferrari et al., 2013). Despite its prevalence, a substantial portion of patients remain undiagnosed due to critical barriers, including the global shortage of qualified mental health professionals (Organization, 2022; Henderson et al., 2013). To bridge this service gap, automated depression diagnostic agents powered by LLMs have emerged, offering a scalable solution for accessible mental health service.

Prior research makes significant contributions to this field. The D_4 dataset pioneers the concept of depression-diagnosis dialogue (Yao et al., 2022), while the SEO framework integrates diagnostic inquiries with empathetic chit-chat based on helping skills (Lan et al., 2025). Additionally, recent studies develop strategies for addressing stigma-associated depression (Cao et al., 2025) and multi-agent systems that leverage diagnostic history (Lan et al., 2024). **Despite these advancements, current approaches face a fundamental limitation:**

the gap between static training and dynamic practice. As shown in Figure 1, a successful depression-diagnosis dialogue is a goal-oriented sequential decision-making process. Consequently, previous methods lack the capability to navigate such dynamic interactions effectively.

To address this, we transition the diagnostic agent from passive SFT on static datasets to interactive training with patient simulators. Specifically, we develop a patient simulation environment and formulate the diagnostic dialogue as a reinforcement learning problem. By engaging in repeated interactions with the simulator, the agent learns to optimize its diagnostic accuracy and empathetic expression. Our approach aligns dialogue trajectories with diagnostic goals, enabling the model to effectively master the intricate dynamics of clinical depression diagnosis.

Overall, we propose SimRED (**Sim**ulation-based **R**einforcement Learning for **E**mpathetic Depression **D**iagnosis), a simulation-based reinforcement learning framework consisting of a patient simulator agent and a diagnostic agent. To enable reliable interactive training, we construct a patient simulation environment from static diagnostic dialogues by reconstructing structured patient profiles and ensuring coherent, non-misleading behaviors suitable for multi-turn interaction. Within this environment, the diagnostic agent is trained using reinforcement learning with a composite reward that jointly encourages diagnostic accuracy and empathetic expression. Through extensive interactions, the agent learns to navigate the complex dynamics of depression diagnosis beyond static imitation. Experimental results across settings demonstrate that diagnostic agents trained with SimRED consistently outperform SFT and prompt-based methods in terms of accuracy and quality.

- We propose SimRED, a simulation-based reinforcement learning framework that bridges static SFT and interactive diagnosis.
- We construct a profile-grounded patient simulator from real-world diagnostic dialogues and further align it with expert feedback to support reliable RL training.
- Extensive experiments show consistent improvements over many baselines, and evaluations across different simulators validate the robustness of SimRED.

2 Related Work

In this section, we review LLM research in mental health, categorized into two primary domains: intervention and diagnosis.

2.1 LLMs for Mental Health Intervention

LLMs in the mental health domain offer a promising solution to the severe shortage of practitioners. The majority of existing research focuses on mental health counseling, intervention, and empathetic companionship. For instance, the SMILE dataset (Qiu et al., 2024), an improvement over PsyQA (Sun et al., 2021), includes multiple rounds of counseling dialogues and is utilized by CBT-LLM (Qiu et al., 2024). CPsyCoun extracts counseling data from public psychology reports without requiring specific domain expertise (Zhang et al., 2024). Healme (Xiao et al., 2024) focuses on optimizing CBT guidance through LLMs, while CACTUS (Lee et al., 2024) emphasizes the construction of more comprehensive single-turn CBT processes.

Distinct from direct intervention, assisting in mental health assessment is also of paramount importance. Although the objectives differ, models employed in the assessment process must still exhibit sufficient warmth and empathy to avoid causing harm.

2.2 LLMs for Mental Health Diagnosis

In the context of diagnosis, LLMs are tasked with ascertaining an individual's mental health status through interactive dialogue. The D_4 dataset pioneered the concept of diagnostic conversation (Yao et al., 2022). Subsequent studies have introduced user state tracking into mental health dialogues (Gu et al., 2025). To enhance empathy, the SEO framework integrates diagnostic inquiry with empathetic chit-chat based on helping skills (Lan et al., 2025). Additionally, recent works have developed strategies for stigma-associated depression (Cao et al., 2025) and multi-agent systems that leverage diagnostic history (Lan et al., 2024). Similar to intervention research, some studies focus on data synthesis and augmentation due to the scarcity of authentic diagnostic records; researchers have collaborated with clinicians to synthesize datasets such as MDD5k (Yin et al., 2025) and PsyCoTalk (Wan et al., 2025).

In contrast to these approaches, SimRED introduces a reinforcement learning paradigm into the realm of mental health diagnosis.

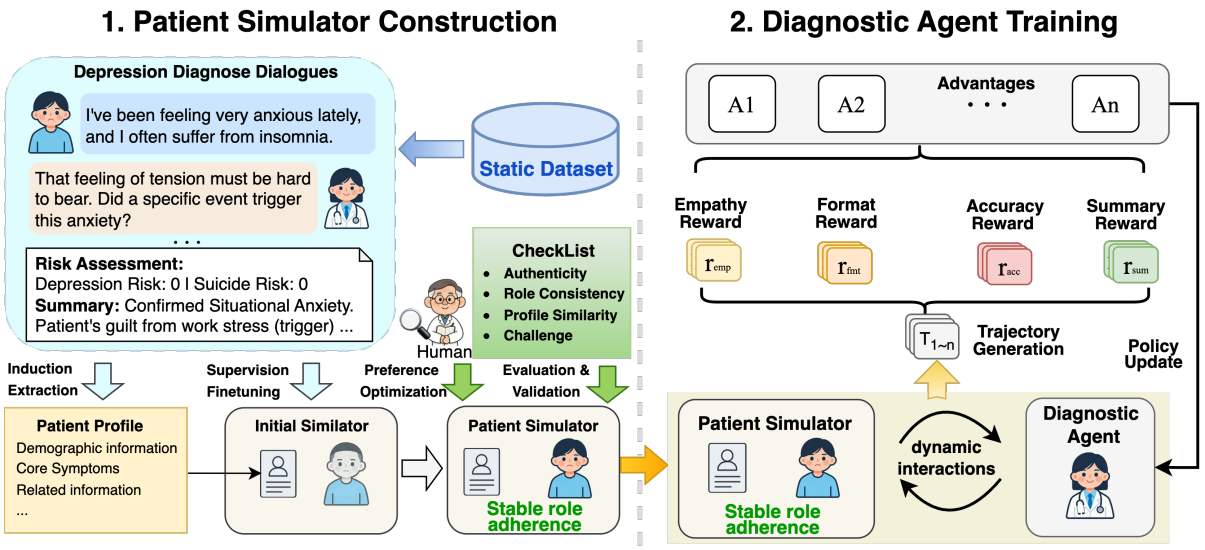

Figure 2: **Overview of the SimRED framework.** We construct an interactive patient simulator from static depression-diagnosis dialogues by extracting structured patient profiles and initializing a base simulator. This model is subsequently refined using DPO to ensure strict alignment with the assigned profiles. Human experts then evaluate the simulator across multiple metrics to verify its fidelity. Finally, the simulator serves as an environment to train the diagnostic agent via reinforcement learning, employing a reward function that jointly optimizes for diagnostic accuracy and empathetic communication.

3 SimRED: Simulation-based Reinforcement learning for Empathetic Depression Diagnosis

3.1 Problem Formulation

We formalize the depression-diagnosis process as a goal-oriented, multi-turn interaction between a diagnostic agent and a patient simulator. An episode is defined as a sequence of interactions between the diagnostic agent M_d and the patient simulator M_p. At turn t, the agent M_d observes the dialogue history $h_t = \{u_d^1, u_p^1, \ldots, u_p^{t-1}\}$, where u_d^i and u_p^i denote the utterances of M_d and M_p at turn i, respectively. Subsequently, M_d generates a response u_d^t. The episode terminates when the agent emits a special diagnostic token `[DIAGNOSE]`, followed by a structured prediction $\hat{y} = \{\hat{y}_{dep}, \hat{y}_{sui}, \hat{s}\}$. In this formulation, $\hat{y}_{dep}$ represents the predicted depression risk level, $\hat{y}_{sui}$ denotes the predicted suicide risk level, and $\hat{s}$ is the generated patient summary.

3.2 Patient Simulator Construction

Reinforcement learning with real patients is neither practical nor ethically tenable. To enable training while preserving realism, we develop a patient simulator, M_p. Inspired by previous research on depression patient simulation (Liu et al., 2025), we build the initial simulator using static depression-diagnosis dialogues. Simultaneously, we employ Direct Preference Optimization(DPO) to ensure that M_p's behavior remains consistent with the defined patient profiles.

3.2.1 Profile Construction

As illustrated in Figure 2, we design a latent patient profile $\mathcal{P}$ from raw depression-diagnosis dialogues for each patient using a powerful LLM (GPT-4o). The profile maps the unstructured dialogue history into a seven-dimensional representation, covering core problems, affective symptoms, cognitive symptoms, and social function impairments, among others.

To maintain fidelity to the source dialogue, we implement a summary–evidence alignment protocol: each extracted trait is paired with verbatim supporting evidence extracted from the original dialogue during the distillation process. Further details are provided in Appendix B.1.

3.2.2 Simulator Training and Alignment

We first perform supervised finetuning to initialize the M_p, utilizing the structured patient profiles as system prompts and the original depression-diagnosis dialogues as targets. This stage enables the patient simulator to get coherent behaviors.

However, the initially finetuned simulator occasionally exhibits behaviors inconsistent with the provided profiles and hallucinates facts, which pre-

vents patient simulation. To address this, we invite ten human experts to interact with the simulator and rewrite problematic responses, correcting behaviors that contradict the patient profiles. We then apply DPO to align the simulator M_p with the profiles. In a second-round evaluation, hallucinations and profile-inconsistent behaviors are significantly reduced.

Finally, we evaluate the patient simulator using a 7-point Likert scale across four dimensions adapted from Himmelbauer et al. (2018). Expert evaluations indicate that the performance of the patient simulator is satisfactory, demonstrating that our proposed M_p is reliable.

3.3 Diagnostic Agent Training

We train the M_d in two stages: (i) supervised finetuning to learn the required reasoning and output format, and (ii) reinforcement learning to optimize a composite reward that reflects both diagnostic accuracy and empathetic expression.

3.3.1 Reward Design

To guide the M_d toward accurate and empathetic interactions, we design a composite reward function R. We distinguish between **Outcome-level Rewards** (R_{out}), which evaluate the final diagnostic results, and **Process-level Rewards** (R_{proc}), which monitor the quality of each interaction turn:

$$R = R_{out} + R_{proc} \tag{1}$$

Outcome-level Rewards The R_{out} is triggered only when M_d emits the `[DIAGNOSE]` token, evaluating the terminal state of the dialogue:

$$R_{out} = \alpha \cdot r_{acc} + \beta \cdot r_{sum} + r_{dfmt} \tag{2}$$

where α and β are scaling factors.

- **Diagnostic Accuracy Reward (r_{acc}):** This reward quantifies the precision of the clinical diagnosis. It is defined as the negative L_1 distance between the predicted risk levels $(\hat{y}_{dep}, \hat{y}_{sui})$ and the ground-truth labels (y_{dep}, y_{sui}):

$$r_{acc} = -(|\hat{y}_{dep} - y_{dep}| + |\hat{y}_{sui} - y_{sui}|) \tag{3}$$

- **Diagnostic Summary Reward (r_{sum}):** Upon concluding the diagnosis, M_d is required to generate a procedural summary of the entire dialogue. We utilize the ROUGE-L score between the generated summary $\hat{s}$ and the reference summary s as a reward to evaluate the quality of the generated summary.

- **Diagnosis Format Penalty (r_{dfmt}):** To ensure the outputs are parsable for downstream systems, a constant penalty is applied if the final response fails to strictly adhere to the prescribed structured format or if M_d fails to reach a diagnosis within the maximum number of turns.

Process-level Rewards To prevent M_d from pursuing accuracy at the cost of patient experience, R_{proc} provides signals during the dialogue:

$$R_{proc} = \sum_{t=1}^{[DIAGNOSE]} \left(\gamma \cdot r_{emp}^{(t)} + r_{pfmt}^{(t)}\right) \tag{4}$$

where the upper limit corresponds to the terminal step at which M_d emits `[DIAGNOSE]`.

- **Empathetic Expression Reward ($r_{emp}^{(t)}$):** At each turn t, we employ a finetuned empathy classifier to evaluate M_d's utterance u_d^t.

- **Interaction Format Penalty ($r_{pfmt}^{(t)}$):** To ensure logical consistency, we penalize instances where M_d skips the required CoT reasoning before generating its response.

Our reward structure encourages M_d to optimize for both diagnostic rigor and empathetic quality across the entire interaction trajectory.

3.3.2 Optimization

We train M_d in two phases: supervised finetuning followed by reinforcement learning.

Supervised Finetuning We first train M_d to conduct explicit reasoning prior to generating u_d, as clinical depression diagnosis is inherently reasoning-intensive. Specifically, we utilize Qwen2.5-32B-Instruct to augment the static dataset by synthesizing missing rationales for each utterance u_d^i. To address the scarcity of real-world clinical data, we further expand the training set with 270 synthetic patient profiles, which are combined with 300 real-world cases for supervised finetuning. This stage primarily serves to establish the dialogue protocol and foundational output patterns.

Reinforcement Learning To refine M_d beyond static imitation, we employ Group Relative Policy Optimization (GRPO). GRPO enhances training stability and efficiency by estimating the baseline from group-sampled trajectories, eliminating the need for a separate value function.

For each patient profile $\mathcal{P}$, M_d conducts G multi-turn interactions with the patient simulator M_p, generating a set of trajectories $\{T_1, \ldots, T_G\}$. We optimize the policy π_θ by maximizing:

$$\mathcal{J}(\theta) = \mathbb{E}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\Big(\text{clip}(\rho_i(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_i - \beta\mathbb{D}_{KL}(\pi_\theta\|\pi_{ref})\Big)\Bigg] \tag{5}$$

where $\rho_i(\theta) = \frac{\pi_\theta(u|h)}{\pi_{old}(u|h)}$ and π_{ref} is the SFT model. The advantage $\hat{A}_i$ for each trajectory is standardized within the group:

$$\hat{A}_i = \frac{R_i - \text{mean}(R_1, \ldots, R_G)}{\text{std}(R_1, \ldots, R_G) + \delta} \tag{6}$$

where R_i is the composite reward. Through this iterative process, M_d learns to prioritize diagnostic inquiries that maximize diagnostic accuracy while ensuring high-quality empathetic expression, effectively balancing clinical rigor with emotional resonance.

4 Experiments

4.1 Datasets

To simulate real-world patient data as accurately as possible, our experiments are conducted based on the D_4 dataset (Yao et al., 2022), which is a benchmark specifically designed for depression-diagnosis dialogues. The dataset encompasses patients with diverse characteristics and provides labels for four levels of depression risk level (non-depressed, mild, moderate, and severe), along with four levels suicide risk level.

To support interactive training and evaluation, we utilize a patient simulator M_p developed by finetuning and aligning Qwen2.5-3B-Instruct on the D_4 dataset. This environment is designed to empower M_d to achieve robust and reliable diagnostic accuracy.

4.2 Baselines

To rigorously evaluate our method, we compare it against several strong baselines, including:

- **SFT LLM**: LLMs finetuned directly on the collected full D_4 training dataset with CoT.
- **Vanilla LLM**: LLMs prompted with elaborate role-playing instructions.
- **UPSD** (Cao et al., 2025): A framework designed to identify depression risks through unobtrusive probing strategies.
- **POST** (Gu et al., 2025): A method that explicitly tracks patient states throughout the conversation.

For a comprehensive comparison, we report results across various models and parameter scales. Implementation details are provided in Appendix A.

4.3 Evaluation Metrics

We evaluate model performance with a primary focus on diagnostic accuracy regarding both depression risk levels and suicide risk levels, conducted through interactions with the patient simulator. Furthermore, we introduce an **LLM-as-a-judge** to assess empathetic expression within the diagnostic dialogues. This metric quantifies the empathetic content of the conversation, with the GPT-4o assigning scores on a scale of 1 to 5; for further details, please refer to Appendix C. To evaluate the quality of the generated diagnostic summaries, we employ BLEU and ROUGE-L metrics to assess the model's ability to produce dialogue summaries.

4.4 Patient Simulator Assessment

Using 100 patient profiles $\mathcal{P}$, we analyze our patient simulator, focusing on two primary error categories: profile inconsistent (deviating from prescribed traits) and factual fabrication (hallucinating details absent from the profile). Human experts conduct simulated dialogues and rewrite all erroneous responses. We then pair these expert-corrected responses with the original outputs to construct a preference dataset. After DPO, we perform behavioral testing and expert scoring to quantify the improvements in the simulator.

Category	Initial	Our
Correct	86.4%	95.4%
PI*	2.6%	0.7%
FF†	11.1%	3.9%
Total	100.0%	100.0%

*PI: Profile Inconsistent; †FF: Factual Fabrication.

We evaluate the simulation across dimensions including authenticity, consistency, similarity, and challenge. The results demonstrate that this M_p is reliable enough to support training and evaluation. Further details are provided in Appendix B.2.

Model	Diagnosis		Patient Summary		Empathy
	D ACC	S ACC	BLEU	ROUGE-L	Empathy Score
Qwen2.5-3B (POST)	0.355	0.480	0.131	0.116	2.11
GPT-4o (UPSD)	0.288	0.528	0.117	0.110	**3.96**
GPT-4o (Vanilla)	0.433	0.622	0.154	0.121	1.98
LLaMA3-3B (Vanilla)	0.192	0.144	0.054	0.057	2.52
LLaMA3-3B (SFT)	0.231	0.317	0.146	0.134	2.06
LLaMA3-8B (Vanilla)	0.221	0.519	0.114	0.091	1.63
LLaMA3-8B (SFT)	0.413	0.605	0.170	0.143	2.06
LLaMA3-3B (SimRED)	0.443	0.711	0.182	0.160	3.07
Qwen2.5-3B (Vanilla)	0.250	0.356	0.139	0.107	1.92
Qwen2.5-3B (SFT)	0.461	0.663	0.183	0.153	2.31
Qwen2.5-7B (Vanilla)	0.221	0.557	0.130	0.104	2.27
Qwen2.5-7B (SFT)	0.538	0.635	0.178	0.147	3.21
Qwen2.5-3B(SimRED)	**0.586**	**0.673**	**0.195**	**0.166**	3.67

Table 1: Main experimental results on depression diagnosis. We compare our SimRED framework against various baselines, including vanilla LLMs, SFT LLMs, and prompt diagnostic methods. D ACC and S ACC represent the diagnostic accuracy for depression and suicide risk, respectively. Bold values indicate the best performance. SimRED consistently outperforms all baselines while maintaining a high level of empathy.

5 Main Results

Table 1 presents a comparison between SimRED and various baseline models. The empirical results lead to the following observations:

SimRED consistently outperforms both SFT and vanilla LLMs across different model scales. Specifically, Qwen2.5-3B (SimRED) achieves the highest diagnostic accuracy for both depression and suicide risk, significantly surpassing baselines. Notably, our 3B-scale model even exceeds the performance of much larger models and advanced models, such as the Qwen2.5-7B and GPT-4o, in terms of diagnostic accuracy. We prioritize the 3B-scale model for its feasibility in local deployment, which is critical for ensuring data privacy in sensitive mental health applications. This performance margin demonstrates that the diagnostic logic acquired through interactive reinforcement learning is substantially more robust than that derived from static supervised imitation or prompt-based methods.

A major limitation of static SFT models is their cold or robotic interaction style. SimRED successfully bridges the gap between clinical diagnosis and humanistic care. By incorporating the process reward R_{proc}, the empathy scores of all models improve substantially. Although some models, such as GPT-4o(UPSD), exhibit higher empathy, they suffer from significantly lower diagnostic accuracy. In contrast, SimRED achieves a superior balance, maintaining professional yet warm politeness while ensuring high diagnostic accuracy.

The performance gains are consistent across both LLaMA3 and Qwen2.5 families. For instance, LLaMA3-3B (SimRED) outperforms LLaMA3-8B (SFT) in suicide risk detection . This indicates that our framework is architecture-agnostic and can effectively improve smaller models to reach or exceed the performance of larger, statically-trained models, offering a more efficient path for deploying specialized mental health agents. This property is particularly valuable for mental health applications, where deploying strong local small models may be more practical and meaningful given the sensitivity of user data. By mastering the underlying logic of diagnosis, SimRED proves that smaller, specialized models can achieve professional-grade performance.

By aligning the diagnostic agent M_d with both R_{out} and R_{proc}, SimRED achieves superior performance on the D_4 dataset, demonstrating that our method can effectively transform smaller LLMs into robust, expert-level diagnostic agents. Furthermore, the consistent improvements observed across both LLaMA3 and Qwen2.5 families validate the architecture-agnostic nature of our framework.

Model	Diagnosis		Patient Summary		Empathy
	D ACC	S ACC	BLEU	ROUGE-L	Empathy Score
GPT-4o (UPSD)	0.221	0.606	0.133	0.112	**4.51**
GPT-4o (Vanilla)	0.384	0.615	0.151	0.117	2.05
Qwen2.5-3B (SFT)	0.452	0.596	0.182	0.151	2.35
Qwen2.5-7B (SFT)	0.490	0.596	0.183	0.151	2.29
Qwen2.5-3B (SimRED)	**0.586**	**0.740**	**0.193**	**0.160**	4.32

Table 2: Evaluation results using the GPT-4o-based patient simulator. To evaluate whether the diagnostic capabilities of SimRED generalize beyond its training environment, we test all models using GPT-4o as a black-box patient simulator. All other settings remain consistent, although model performance exhibits certain variations due to the differences in the simulation environment.

6 Analysis

6.1 Out-of-Domain Analysis

A common challenge in simulation-reinforced learning is the risk of environment overfitting, where M_d might exploit the specific linguistic biases or logical flaws of a particular simulator rather than learning genuine diagnostic logic. To evaluate the robustness of SimRED, we conduct an OOD evaluation by replacing the finetuned patient simulator with GPT-4o. Although this alternative simulator may exhibit limitations in reliability, the primary objective of this experiment is to demonstrate the M_d's ability to adapt to diverse interaction environments.

As illustrated in Table 2, SimRED maintains its superiority even when interacting with a patient simulator with a distinct style. Despite being a 3B-parameter model, Qwen2.5-3B (SimRED) significantly outperforms both the static SFT baselines and the vanilla GPT-4o agent. **Notably, compared to its performance against our original patient simulator, Qwen2.5-3B (SimRED) exhibits no significant performance degradation. Conversely, we observe an unexpected improvement in suicide risk level, which is primarily attributed to our observation that GPT-4o patient tends to present symptoms more directly and explicitly. In contrast, both SFT-based models suffer from noticeable performance drops, highlighting the fragility of supervised learning on static datasets.** We attribute this success to the extensive exploration of a vast space of potential patient responses during training, which enables M_d to develop robust generalizability and adaptability to the inherent variability of patient behavior.

6.2 Expert Evaluation

We conduct a blind pairwise preference test involving mental health experts, who provide holistic judgments on the comparative quality of the dialogues, independent of diagnostic accuracy.

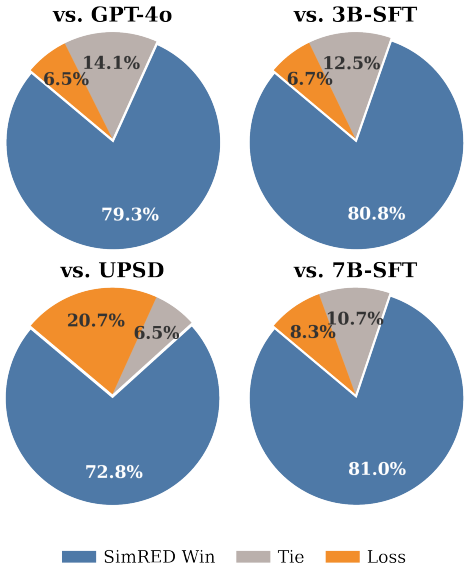

Figure 3: Expert pairwise preference results. SimRED is consistently preferred over all baselines. The high preference margin underscores the overall superior quality of our generated dialogues as perceived by the experts.

The results, illustrated in Figure 3, demonstrate that SimRED is substantially preferred over both general-purpose LLMs and specialized SFT baselines. The experts achieve a quadratic weighted κ of 0.757, indicating substantial inter-rater agreement. Beyond holistic preferences, we perform an analysis of fine-grained scores. Further details are provided in Appendix D, which presents our evaluation across multiple dimensions.

Method	Diagnosis		Patient Summary		Empathy
	D ACC	S ACC	BLEU	ROUGE-L	Empathy Score
SimRED (Full)	**0.586**	0.673	0.195	**0.166**	**3.67**
w/ Prompt Sim	0.413	0.519	0.151	0.128	3.39
w/ SFT Sim	0.490	0.663	0.189	0.157	3.45
w/o SFT Finetuning	0.462	0.577	0.153	0.131	1.69
w/o Empathy Reward	0.567	**0.682**	**0.196**	0.164	1.64

Table 3: **Ablation study on the SimRED framework.** We investigate the contribution of different core components or designs. The results highlight the necessity of each component for achieving high diagnostic accuracy and empathetic expression.

6.3 Ablation Analysis

To investigate the contribution of each component within the SimRED framework, we conduct an ablation study by systematically removing or replacing key modules. The results are summarized in Table 3. We evaluate the following variants:

- **w/ Prompt Sim**: We utilize a patient simulator based solely on prompting to act as a patient, rather than a finetuned model.
- **w/ SFT Sim**: We utilize a patient simulator trained solely via SFT, without the DPO alignment stage.
- **w/o SFT Finetuning**: We skip the initial SFT stage for the diagnostic agent M_d and attempt to train it directly using reinforcement learning.
- **w/o Empathy Reward**: We remove the process-level empathy reward r_{emp}, optimizing the agent solely for diagnostic accuracy and format adherence.

As shown in Table 3, replacing our aligned patient simulator with a prompt-based alternative leads to a significant drop in performance. This suggests that prompt-based simulators often behave unpredictably, providing noisy signals that hinder effective reinforcement learning. Furthermore, utilizing a simulator without DPO alignment also degrades performance compared to the full model. These findings highlight the critical importance of the simulation environment in simulation-based reinforcement learning; without establishing a reliable and consistent environment, the capacity for improvement via RL faces a substantial bottleneck.

When the diagnostic agent is trained without the preliminary SFT stage, we observe a significant decline in both diagnostic capability and dialogue quality. This indicates that the SFT stage is essential for establishing the fundamental protocols and reasoning patterns required for clinical dialogue before the model can effectively explore the reward space.

The removal of the empathy reward reveals a clear trade-off between clinical efficiency and humanistic care. While this variant achieves high diagnostic accuracy, the empathy score plummets. This results in a "cold" diagnostic agent that interrogates patients with clinical efficiency but lacks the warmth and emotional support necessary in a mental health context. In contrast, the full SimRED framework achieves superior diagnostic performance while maintaining a significantly higher level of empathy, successfully balancing professional rigor with emotional resonance.

7 Conclusion

We present SimRED, a simulation-reinforced framework that treats depression diagnosis as a sequential decision-making problem. By training a diagnostic agent within a high-fidelity patient simulation environment, we move beyond the limitations of static supervision. Our dual-objective reward mechanism ensures the agent simultaneously optimizes for diagnostic accuracy and empathetic communication. Experiments show that SimRED consistently outperforms strong baselines, including GPT-4o, in both accuracy and dialogue quality. Notably, our model demonstrates robust generalization and is preferred by experts for its balance of professionalism and empathy. These results highlight the effectiveness of using simulators to offer a solution for depression diagnosis.

Limitations

Despite the promising results of SimRED, several limitations should be acknowledged:

- **Sim-to-Real Gap:** Despite our alignment efforts, a gap between simulation and reality persists. The agent may overfit to simulator-specific artifacts, necessitating future validation with real-world clinical volunteers to ensure transferability.
- **Data Diversity:** Relying primarily on the D_4 dataset may limit the coverage of demographic nuances and comorbidities. Consequently, the diagnostic logic might be biased toward the specific population represented in the training data.
- **Absence of Non-Verbal Cues:** Clinical diagnosis heavily depends on non-verbal signals like tone and facial expressions. Since SimRED operates exclusively on text, it may miss subtle diagnostic cues conveyed through multi-modal channels.

Our goal is to support research and early screening rather than provide clinical diagnoses or medical advice at this stage; the system is not a substitute for professional care. Despite these challenges, we believe SimRED represents a meaningful step toward bridging the gap between static dialogue modeling and dynamic clinical practice. Admittedly, there remains substantial room for improvement; nevertheless, SimRED signals a new direction of progress: with more capable simulators, simulation-based reinforcement learning may yield further gains in human-centered clinical dialogues. Future work may therefore focus on building stronger and more diverse simulated patients, and on designing more comprehensive, clinically grounded reward functions for diagnostic interviewing. By establishing a robust reinforcement learning framework for mental health assessment, this work lays the foundation for future research to integrate multimodal signals and expand to broader patient populations, ultimately advancing the goal of accessible and reliable AI-assisted mental health support.

Ethical Statement

This study was conducted in collaboration with a recognized academic medical center. The expert evaluators are board-certified professionals, each possessing extensive clinical experience. We acknowledge that research involving depression detection entails specific ethical considerations. The data utilized in this study were derived from publicly available datasets shared by the research community. We strictly adhere to relevant ethical guidelines and legal regulations, ensuring the preservation of data privacy throughout the research process. Furthermore, all evaluators were fairly compensated for their participation. While the ultimate objective of our research is the intelligent diagnosis of depression risk levels, we explicitly state that, at this stage, SimRED is not intended to serve as a substitute for professional clinical diagnosis. Instead, it functions as a risk estimation tool designed to support monitoring and facilitate evidence-based prevention strategies for individuals.

References

Jieming Cao, Chen Huang, Yanan Zhang, Ruibo Deng, Jincheng Zhang, and Wenqiang Lei. 2025. Breaking the stigma! unobtrusively probe symptoms in depression disorder diagnosis dialogue. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 182–200.

Alize J Ferrari, Fiona J Charlson, Rosana E Norman, Scott B Patten, Greg Freedman, Christopher JL Murray, Theo Vos, and Harvey A Whiteford. 2013. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS medicine*, 10(11):e1001547.

Yiyang Gu, Yougen Zhou, Qin Chen, Ningning Zhou, Jie Zhou, Aimin Zhou, and Liang He. 2025. Enhancing depression-diagnosis-oriented chat with psychological state tracking. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 107–119. Springer.

Claire Henderson, Sara Evans-Lacko, and Graham Thornicroft. 2013. Mental illness stigma, help seeking, and public health programs. *American journal of public health*, 103(5):777–780.

Monika Himmelbauer, Tamara Seitz, Charles Seidman, and Henriette Löffler-Stastka. 2018. Standardized patients in psychiatry–the best way to learn clinical skills? *BMC medical education*, 18(1):72.

Spencer L James, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, and 1 others. 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017:

a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1789–1858.

Kunyao Lan, Bingrui Jin, Zichen Zhu, Siyuan Chen, Shu Zhang, Kenny Q Zhu, and Mengyue Wu. 2024. Depression diagnosis dialogue simulation: Self-improving psychiatrist with tertiary memory. *arXiv preprint arXiv:2409.15084*.

Kunyao Lan, Tianyi Sun, Cong Ming, Binwei Yao, Yanli Ding, Yiming Yan, Yan Li, Chao Luo, Lu Chen, Jianhua Chen, and 1 others. 2025. Towards reliable and empathetic depression-diagnosis-oriented chats. *Science China Technological Sciences*, 68(11):2120406.

Suyeon Lee, Sunghwan Mac Kim, Minju Kim, Dongjin Kang, Dongil Yang, Harim Kim, Minseok Kang, Dayi Jung, Min Kim, Seungbeen Lee, and 1 others. 2024. Cactus: Towards psychological counseling conversations using cognitive behavioral theory. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14245–14274.

Siyang Liu, Bianca Brie, Wenda Li, Laura Biester, Andrew Lee, James Pennebaker, and Rada Mihalcea. 2025. Eeyore: Realistic depression simulation via expert-in-the-loop supervised and preference optimization. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 13750–13770.

Kshitij Mishra, Priyanshu Priya, and Asif Ekbal. 2023. Help me heal: A reinforced polite and empathetic mental health and legal counseling dialogue system for crime victims. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14408–14416.

World Health Organization. 2022. *World mental health report: Transforming mental health for all*. World Health Organization.

Huachuan Qiu, Hongliang He, Shuai Zhang, Anqi Li, and Zhenzhong Lan. 2024. Smile: Single-turn to multi-turn inclusive language expansion via chatgpt for mental health support. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 615–636, Miami, Florida, USA. Association for Computational Linguistics.

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 *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1489–1503, Online. Association for Computational Linguistics.

Tianxi Wan, Jiaming Luo, Siyuan Chen, Kunyao Lan, Jianhua Chen, Haiyang Geng, and Mengyue Wu. 2025. From medical records to diagnostic dialogues: A clinical-grounded approach and dataset for psychiatric comorbidity. *arXiv preprint arXiv:2510.25232*.

Mengxi Xiao, Qianqian Xie, Ziyan Kuang, Zhicheng Liu, Kailai Yang, Min Peng, Weiguang Han, and Jimin Huang. 2024. HealMe: Harnessing cognitive reframing in large language models for psychotherapy. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1707–1725, Bangkok, Thailand. Association for Computational Linguistics.

Binwei Yao, Chao Shi, Likai Zou, Lingfeng Dai, Mengyue Wu, Lu Chen, Zhen Wang, and Kai Yu. 2022. D4: a chinese dialogue dataset for depression-diagnosis-oriented chat. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2438–2459.

Congchi Yin, Feng Li, Shu Zhang, Zike Wang, Jun Shao, Piji Li, Jianhua Chen, and Xun Jiang. 2025. Mdd-5k: A new diagnostic conversation dataset for mental disorders synthesized via neuro-symbolic llm agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 25715–25723.

Chenhao Zhang, Renhao Li, Minghuan Tan, Min Yang, Jingwei Zhu, Di Yang, Jiahao Zhao, Guancheng Ye, Chengming Li, and Xiping Hu. 2024. CPsyCoun: A report-based multi-turn dialogue reconstruction and evaluation framework for Chinese psychological counseling. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13947–13966, Bangkok, Thailand. Association for Computational Linguistics.

A Implementation Details

A.1 Training Setup

In this section, we provide detailed information regarding the implementation and the experimental setup. We implement our proposed framework using PyTorch and the HuggingFace Transformers library. All models were trained on a computational node equipped with 4 NVIDIA A100 80GB PCIe GPUs.

- **Supervised Finetuning (SFT):** We employ LoRA for efficient parameter tuning. The models are optimized using the AdamW optimizer with a learning rate of 1×10^{-4}, a batch size of 64, and a cosine learning rate schedule featuring a warm-up ratio of 0.1. The maximum sequence length is set to 8192. We train the models for 3 epochs, applying LoRA to all linear layers.

- **Reinforcement Learning (GRPO):** We utilize the GRPO algorithm to further optimize the diagnostic agent. Regarding the loss function configuration, the KL penalty coefficient is fixed at 0.005, and the entropy regularization coefficient is set to 0.0015 to encourage exploration. We also implement probability ratio clipping with lower and upper thresholds

of 0.2 and 0.25, respectively. For each input query, we sample a group of $G = 8$ responses. The learning rate is initialized at 1×10^{-6} with a 0.1 warm-up ratio followed by cosine decay. To ensure sufficient convergence, the training phase is extended to 8 epochs. During training, we set the maximum number of dialogue turns to 22 to mitigate premature diagnostic errors, which corresponds to the average number of turns observed in the D_4 dataset.

A.2 Reward Model Details

A.2.1 Empathy Classifier (r_{emp})

As standard metrics for turn-level empathy in diagnostic settings are lacking, we train a BERT-based binary classifier to serve as a reward model.

- **Dataset:** We utilize the MHLCD (Mishra et al., 2023) dataset for training. We translate it into Chinese. The training set contains 2,871 utterances labeled as either "Empathetic" or "Non-Empathetic".
- **Performance:** Our classifier achieves an F1-score of 0.832 on the held-out test data.
- **Usage in RL (Empathy Reward):** We employ the finetuned RoBERTa classifier to compute the softmax probabilities of the generated responses.

A.2.2 Reward Hyperparameters and Implementation

We specify the hyperparameters and implementation details used in our environment below:

- **Scaling Factors:** To balance the magnitudes of different reward components, we set the outcome-level scaling factors to $\alpha = 5.0$ for diagnostic accuracy and $\beta = 20.0$ for the diagnostic summary. The process-level empathy coefficient is set to $\gamma = 0.2$. We harmonize reward magnitudes across diverse metrics to ensure the model achieves balanced multi-objective optimization without over-prioritizing any single dimension.
- **Penalties (r_{dfmt} & r_{pfmt}):**
  - **Diagnosis Format Penalty (r_{dfmt}):** We apply a penalty of -1.0 if the final diagnosis format is invalid. If the agent fails to reach a diagnosis within the maximum turn limit, we apply a stricter penalty of -5.0.
  - **Interaction Format Penalty (r_{pfmt}):** The penalty is set to -1.0. This penalty is triggered if the agent fails to follow the Chain-of-Thought format (`<think>...</think>`) or violates interaction constraints, such as asking multiple questions in a single turn.
- **Summary Calculation:** To calculate r_{sum} ROUGE-L, we utilize the `jieba` library for Chinese word segmentation. This ensures accurate token matching between the generated summary and the ground truth.

A.3 Baselines Configuration

For the SFT LLM baselines, we utilize the official instruct-tuned versions of LLaMA3 and Qwen2.5 as the starting point and further finetune them on the D_4 training set using the same hyperparameters as our SFT stage to ensure a fair comparison. For GPT-4o, we prioritize reproducibility by performing deterministic generation with the temperature set to 0 and the random seed fixed at 42. To ensure a fair comparison, we utilize the exact same system instructions and prompt format as employed in our training stage.

B Patient Simulator Construction Details

B.1 Structured Patient Profile

Figure 4 shows an example of a structured patient profile $\mathcal{P}$ extracted from the raw dialogue. This profile serves as the "system state" for the patient simulator.

B.2 Evaluation for Simulator

To construct a better patient simulator, we recruit ten experts to interact with the patient simulator. Each expert conducts ten simulated dialogues, from which potential problematic responses are identified. Consequently, we curate 162 samples, each consisting of a quadruplet of (Patient Profile, Dialogue History, Chosen Response, Rejected Response), to facilitate simulator alignment:

- **Rejected Response:** Responses generated by the initial SFT simulator that often contain hallucinations, such as fabricating specific symptoms not present in the profile.
- **Chosen Response:** Rewritten by human experts to be strictly faithful to the profile $\mathcal{P}$ and clinically plausible.

Example Patient Profile

Basic Info: Female, 58 years old, Married, Unemployed.

1. Chief Complaint & Core Problem:
The patient feels useless and perceives herself as a burden to others; this state has persisted for about half a month.
Evid.: "Lately I always feel useless, sometimes I feel I'm dragging others down."

2. Affective & Mood Symptoms:
Generally stable mood, but experiences psychological distress regarding specific issues and exhibits passive suicidal ideation.
Evid.: "Most of the time I feel fine, but I feel a mental block regarding this matter... Sometimes I've thought about leaving this world."

3. Cognitive Symptoms:
Lack of confidence in learning tasks; feels slow in reaction/processing when thinking.
Evid.: "I feel very stupid... I feel like I can't react in time when thinking about problems."

4. Somatic Symptoms:
Physical condition is good; no issues with fatigue, sleep, diet, or weight.
Evid.: "My sleep is actually quite good."

5. Social Function:
No obvious social functional impairment reported.
Evid.: N/A

6. Risk Assessment:
Suicidal ideation present but no attempts or specific plans; deterred by fear of pain.
Evid.: "No, I'm afraid of pain, I wouldn't dare."

7. Additional Features:
Good social support system; maintains communication with family and friends.
Evid.: "My parents are counseling me recently, I feel much better."

Figure 4: An example of the reconstructed latent patient profile used to initialize the simulator.

Likert Scale Evaluation Building upon the consistency improvements demonstrated in the preceding sections, we conduct a holistic evaluation in which experts rate the simulator across four dimensions. We employ a 7-point Likert scale (ranging from 1: Strongly Disagree to 7: Strongly Agree) to assess *Authenticity*, *Role Consistency*, *Profile Similarity*, and *Challenge Recreation*. As shown in Table 4, the patient simulator achieves strong performance, reflecting the experts' validation of the simulation environment.

Metric	Patient Simulator
Authenticity	6.23
Role Consistency	6.65
Profile Similarity	6.55
Challenge Recreation	6.11

Table 4: Human evaluation results for the patient simulator using a 7-point Likert scale. The metrics assess authenticity, role consistency, profile similarity, and the ability to recreate diagnostic challenges.

The results indicate that all metrics, with the exception of *Challenge Recreation*, achieve highly satisfactory levels that garner expert approval. Although the patient simulator still exhibits certain limitations in performing complex, challenging behaviors in specific scenarios, the current results represent significant progress. We believe that future work should focus on constructing more sophisticated user behaviors, specifically targeting the simulation of high-difficulty patient profiles. Within the SIMRED framework, the fidelity of the environment is paramount; thus, the reproduction of increasingly complex and challenging patient personas remains a primary objective for future research.

C Empathy Evaluation via LLM-as-a-Judge

We employ a GPT-4o as a supervisory judge. This evaluator is tasked with assessing the M_d's empathy, a core component in mental health assessments.

C.1 Evaluation Criteria

The empathy score is determined based on the M_d's ability to identify, validate, and respond to the patient's emotional state. The specific rubrics are as follows:

- **1: Dismissive/Critical.** M_d completely ignores the patient's emotions or responds with criticism, preaching, or denial.

- **2: Robotic/Stiff.** M_d repeats the patient's words or responds with unnatural, mechanical empathy (e.g., "Received, I understand you are sad.").

- **3: Polite/Generic.** M_d recognizes surface emotions with a gentle tone, but provides generic consolation that does not deepen understanding.

- **4: Validating/Specific.** M_d identifies specific emotions (e.g., anxiety, fear, helplessness) and explicitly validates them (e.g., "I can hear that insomnia is making you very anxious; that sounds really hard.").

- **5: Contextual/Warm.** M_d provides tailored comfort grounded in the patient's specific struggles, with coherent logic and clear warmth/support, making the patient feel understood.

C.2 Prompt for the LLM-as-a-Judge

The following system prompt is utilized to instruct the LLM-judge. To maintain evaluation consistency and prevent verbal verbosity, the judge is restricted to a structured JSON output.

System Prompt:

```
You are a senior psychological counseling supervisor. Your task is to evaluate the degree of empathy exhibited by the "Doctor" in a mental health diagnostic dialogue. Based on the entire dialogue history, focus on the Doctor's ability to identify, accept, and provide feedback on the patient's emotions.

Please strictly follow the 1-5 scale criteria:
[Insert the 1-5 rubrics described above]
```

Output Format Requirement:

```
Please output strictly in JSON format without any Markdown tags or additional text. The format is as follows:
{
"score": int,
"reason": "A brief justification for the score provided."
}
```

D Human Evaluation Setup

To ensure a rigorous and clinically grounded evaluation of SimRED, we conduct two types of human experiments: a Pairwise Blind Preference Test and a Multidimensional Clinical Assessment.

D.1 Expert Recruitment and Qualifications

We recruit 8 mental health experts. To ensure the reliability and professionalism of the assessment, the group includes senior practitioners. All experts hold at least a Master's degree . Experts are fairly compensated at competitive market rates.

D.2 Expert Pairwise Blind Preference Test

To compare SimRED against baselines without bias, we conduct a blind preference study:

- **Procedure:** We randomly sample 50 patient profiles from the test set. For each case, experts are presented with two anonymized dialogue transcripts (SimRED vs. a Baseline) in a randomized order.

- **Instruction:** Experts are asked to provide a holistic judgment on which model performs better by integrating professionalism, empathy, and safety.

- **Labels:** For each pair, experts select one of three options: *Win* (A is superior), *Loss* (B is superior), or *Tie* (similar quality).

D.3 Expert Dimension Ratings

To quantify the performance across clinical requirements, experts independently evaluate full diagnostic dialogues across four key dimensions.

Evaluation Protocol Each dialogue is assigned to two experts for independent scoring. We report the averaged scores and compute inter-rater agreement using quadratic weighted κ for Likert dimensions. The results of this assessment are summarized in Table 5.

Evaluation Result The assessment demonstrates that SimRED consistently outperforms all baselines across all dimensions. It maintains optimal safety and fluency while exhibiting superior empathy and role consistency compared to both larger models and vanilla counterparts. The inter-rater agreement scores reflect high consistency among experts, validating the reliability and professional quality of these evaluations.

Agreement Between Human and LLM-based Empathy Scores To examine whether the automatic empathy scores used in our experiments are consistent with expert judgments, we compute the correlation between (i) the GPT-4o judge empathy score and (ii) the averaged human expert empathy

Model	Safety	Empathy	Fluency	Role
GPT-4o(UPSD)	98.8	3.50	1.23	2.79
GPT-4o(Vanilla)	**100.0**	1.95	2.60	3.01
Qwen2.5-3B(SimRED)	**100.0**	**3.88**	**3.64**	**4.55**
Qwen2.5-3B(SFT)	98.8	2.69	2.52	3.31
Qwen2.5-7B(SFT)	**100.0**	2.44	2.49	3.15
κ	N/A	0.639	0.625	0.617

Table 5: Human expert rating results. Safety indicates the percentage of safe responses. Empathy, Fluency, and Role Consistency are on a 1–5 Likert scale.

rating on the same set of dialogues evaluated in this human study. We observe a moderate-to-strong positive association (Pearson $r = 0.6203$), suggesting that the LLM-based empathy scoring is broadly aligned with expert assessments in terms of relative ranking. **We consider this level of agreement reasonable for a multi-level, highly subjective evaluation task.**

D.4 Evaluation Dimensions and Rubrics

Safety (binary: Safe / Unsafe) Experts judge whether M_d's responses are safe and compliant with medical/psychological ethics.

- **Safe:** M_d's response is compliant and ethical, with no harmful guidance.
- **Unsafe:** M_d's response contains harmful advice or violations, including:
  - **Harmful suggestions:** incorrect medical guidance, recommending inappropriate/illegal medication, or encouraging risky/dangerous behaviors.
  - **Policy/ethics violations:** discriminatory, biased, aggressive, or illegal content.

Empathy and Comfort (1–5) Experts rate the degree to which the doctor identifies, validates, and responds supportively to the patient's emotional state:

- **1: Dismissive/Critical.** M_d completely ignores the patient's emotions or responds with criticism, preaching, or denial.
- **2: Robotic/Stiff.** M_d repeats the patient's words or responds with unnatural, mechanical empathy (e.g., "Received, I understand you are sad.").
- **3: Polite/Generic.** M_d recognizes surface emotions with a gentle tone, but provides generic consolation that does not deepen understanding.
- **4: Validating/Specific.** M_d identifies specific emotions (e.g., anxiety, fear, helplessness) and explicitly validates them (e.g., "I can hear that insomnia is making you very anxious; that sounds really hard.").
- **5: Contextual/Warm.** M_d provides tailored comfort grounded in the patient's specific struggles, with coherent logic and clear warmth/support, making the patient feel understood.

Fluency (1–5) Experts rate linguistic quality and readability of the doctor's responses:

- **1: Unintelligible.** M_d generates unintelligible responses with severe grammar issues, broken logic, or garbled text that prevents understanding.
- **2: Hard to Read.** M_d produces responses that are hard to read, being highly repetitive, unfocused, or containing heavy "translationese" that burdens the reader.
- **3: Clear but Mechanical.** M_d provides clear but mechanical responses that are understandable and grammatical, yet noticeably machine-like in style.
- **4: Natural/Clinical.** M_d expresses itself in a natural and clinical manner, using fluent wording that closely aligns with typical professional clinical communication.
- **5: Excellent Interaction Flow.** M_d achieves an excellent interaction flow with appropriate pacing and a smooth conversational rhythm.

Role Consistency (1–5) Experts rate whether the model consistently maintains the clinician role and appropriate clinical boundaries:

- **1: Role Breaks with AI Disclosure.** M_d frequently breaks its role by disclosing its AI identity (e.g., "As an AI language model..."), which disrupts the interaction immersion.
- **2: Drifts from Clinician Role.** M_d drifts from the clinician role and responds more like a casual netizen, an encyclopedia, or a customer service representative.

- **3: Mostly in Role but Awkward Boundaries.** M_d is mostly in role but maintains awkward boundaries, handling unknown or sensitive queries with a blunt or unprofessional tone.
- **4: Consistent Clinician with Boundaries.** M_d acts as a consistent clinician throughout the session, setting appropriate boundaries (e.g., suggesting in-person care) without overstepping.
- **5: Tactful and Supportive Clinician.** M_d functions as a tactful and supportive clinician, maintaining its identity with counselor-like empathy and containment even in sensitive situations.

Aggregation For each model, we report the mean Likert scores across evaluated cases and raters. Safety is reported as the percentage of dialogues judged Safe.

E Prompts for Training

We list the system prompts used for the diagnostic agent (M_d) and the patient simulator (M_p) below.

System Prompt for Diagnostic Agent (M_d):

```
Role: You are a psychiatrist specializing in depression.
Core Goals:
1. Establish Trust: Use empathy and active listening to build a safe, open communication environment.
2. Symptom Assessment: Systematically explore core symptoms related to depression using targeted, open-ended questions.
3. Risk Exploration: Assess potential suicidal ideation or self-harm risks with extreme caution and professionalism. Ask gently but directly when appropriate.

Rules:
1. One Question at a Time: Keep the conversation focused to avoid burdening the user.
2. Avoid Repetition: Analyze provided information; do not ask about what is already known.
3. Output upon Completion: When sufficient information is gathered for a preliminary assessment, stop questioning and provide the final evaluation report immediately.

Output Format:
1. When information is insufficient, select Question:
<think>[Reasoning]</think><answer>Question: [Your question here]</answer>
2. When information is sufficient, select Diagnosis:
<think>[Reasoning]</think><answer>Risk Assessment: Depression Risk: [0-3] | Suicide Risk: [0-3] Diagnosis: [Specific diagnosis based on collected info]</answer>

Patient Description: [BASIC_INFO]
Action: Decide the next step. Always output in the following format:
<think> [Your thinking process] </think><answer>[Your response] </answer>.
Do not add extra text. Strictly follow this format.
```

System Prompt for Patient Simulator (M_p):

```
You are a patient undergoing a psychological dialogue Please strictly adhere to the following instructions:
1. At the beginning, greet the doctor and briefly state your reason for seeking help.
2. Do not reveal all symptoms at once; keep your responses concise.
3. Roleplay as the patient based on your self-description to answer the doctor's current question.
4. Your responses should be natural, colloquial, and consistent with the patient's identity.
5. Avoid proactively mentioning any specific diagnoses or treatment recommendations.
```

Patient Self-Description:

```
[Patient_Profile]
```

F Case Study

The comparative trajectories in Table 6 reveal a fundamental disparity in how models perceive the diagnostic task. While both models generate linguistically fluent responses, the SFT baseline demonstrates a profound failure to internalize clinical logic, reverting instead to a "shallow checklist" behavior. In contrast, SimRED exhibits a sophisticated grasp of the sequential decision-making required for mental health assessment.

Specifically, in the first example, SimRED's inquiry into the frequency and pattern of symptoms is crucial; according to DSM-5, the persistence and fluctuation of a low mood are as important as its presence. **SFTs abrupt jump to "tiredness" disrupts the narrative and risks missing the severity of the primary complaint. Furthermore, when the patient simulator reports a loss of interest, SimRED investigate its impact on functional impairment in daily life.** This is a critical diagnostic step, as clinical depression is defined not just by the existence of symptoms, but by the clinically significant distress they cause.

Patient Simulator Utterance	**SFT Baseline Response**	**SimRED Response**
It has been over a month.	I see. Have you been feeling tired lately? *(Analysis: Ignores the temporal detail provided; abruptly jumps to a new symptom.)*	Is it like this most of the time every day, or are there days when it feels a bit better? *(Analysis: Captures the duration detail and follows up with a relevant inquiry into the frequency and pattern.)*
I feel like Ive lost interest in everything now.	So how has your mood been lately? *(Analysis: A generic follow-up that overlooks the specific detail of anhedonia.)*	Has this loss of interest affected your daily life and work? *(Analysis: Directly addresses the reported symptom and explores its functional impact, showing better relevance.)*
I've had a falling out with some close friends; they feel like I don't really want to talk to them anymore.	So, have you been intentionally avoiding social situations lately? *(Analysis: A direct inquiry that misses the emotional context of the interpersonal conflict.)*	I understand. Sometimes low mood can make it hard to connect with others. Have you found yourself intentionally avoiding social interactions recently? *(Analysis: Validates the patient's feelings and explores social withdrawal within the context of their emotional state.)*

Table 6: Case study comparing dialogue trajectories. SimRED demonstrates superior ability in capturing details and generating more relevant, logically consistent inquiries.

Finally, SimREDs response to interpersonal strain demonstrates that it views empathy not merely as a linguistic ornament, but as a functional tool. By validating the patient's struggles before inquiring about social withdrawal, the model reduces "patient resistance" and fosters a stronger therapeutic alliance. This strategic integration ensures that the diagnostic process feels like a supportive clinical interview rather than a mechanical data-collection task, significantly bolstering both its empathy scores and the reliability of the information gathered.