RLKGF: Reinforcement Learning from Knowledge Graph Feedback Without Human Annotations

Lian Yan[1], **Chen Tang**[2], **Yi Guan**[1], **Haotian Wang**[1], **Songyuan Wang**[1]
Haifeng Liu[1], **Yang Yang**[3], **Jingchi Jiang**[1*]
[1]Harbin Institute of Technology [2]Institute for Advanced Algorithms Research
[3]Changchun University Of Science And Technology
{23b903008,2021111589,24S003142}@stu.hit.edu.cn, travistang@foxmail.com
wanght1998@gmail.com, yangyang_hit@cust.edu.cn
{guanyi,jiangjingchi}@hit.edu.cn

Abstract

Reinforcement Learning from Human Feedback (RLHF) has been shown to effectively align large language models (LLMs) with human knowledge. However, the lack of human preference labels remains a significant bottleneck when applying RLHF to a downstream domain. Humans in RLHF play a critical role in injecting reasoning preferences into LLM, and we assume the reasoning process underlying human assessments may potentially be replaced by reasoning pathways derived from Knowledge Graphs (KGs). Inspired by this assumption, we propose Reinforcement Learning from Knowledge Graph Feedback (RLKGF), a novel method that leverages KG semantics and structure to derive RL rewards in the absence of manual annotations. Unlike Reinforcement Learning from AI Feedback (RLAIF), RLKGF directly integrates human priors encoded in KGs as the reward model, aligning LLM responses with expert knowledge without additional preference labeling or reward model training. RLKGF structures context-relevant facts into knowledge subgraphs and defines rewards by simulating information flow across semantic and logical connections between question and candidate response entities. Experiments on three public and one private medical dialogue dataset demonstrate that RLKGF significantly outperforms the competitive RLAIF in improving LLM diagnostic accuracy. The code is available at https://github.com/YanPioneer/RLKGF.

1 Introduction

Large language models (LLMs) like ChatGPT (Ouyang et al., 2022) have shown remarkable potential in tasks such as knowledge-based question-and-answer (Q&A) (Liu et al., 2024) and intelligent decision-making (Wang et al., 2025). As LLMs advance in specialized domains like medicine (Zhang et al., 2023a), agriculture (Peng et al., 2023), and law (Huang et al., 2023), the demand for factually accurate and helpful responses grows. Reinforcement learning from human feedback (RLHF), regarded as a key driver of ChatGPT's success, aligns LLM outputs with human preferences and enhances generation quality (Bai et al., 2022a). Its effectiveness has also been validated in domain-specific LLM adaptations (Yang et al., 2024b). However, RLHF involves a complex training process. First, a reward model is learned from ranked human preference data. Subsequently, scores generated by the reward model are used to apply policy optimization (Schulman et al., 2017). Despite its benefits, the high cost of human annotation, inconsistent annotation standards, and potential biases from subjective judgments hinder the widespread application of RLHF.

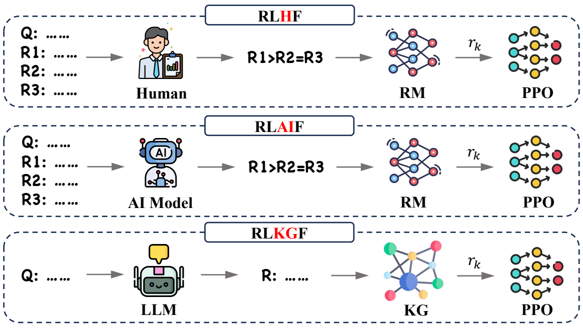


Figure 1: Compared to RLHF and RLAIF, RLKGF treats the knowledge graph (KG) as a reward model (RM), directly providing rewards for LLM responses without the need for preference labeling or reward model training.

Both self-reflection (Asai et al., 2023) and CoT (Wei et al., 2022; Chu et al., 2024) approach in LLMs highlight the advantage of leveraging their embedded knowledge to enhance task performance. Meanwhile, several works mention LLMs have demonstrated human-like judgment capabilities in certain aspects (Gilardi et al., 2023; Ding et al.,

[*] Corresponding Author

Findings of the Association for Computational Linguistics: ACL 2025, pages 6619–6633
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

2023). Thus, automating preference selection for model responses through LLMs is a natural progression. Self-refine (Madaan et al., 2024) and Refiner (Paul et al., 2024) employ LLMs to evaluate and iteratively refine outputs through feedback. Additionally, Anthropic (Bai et al., 2022b) and Google (Lee et al., 2023) directly use LLMs to filter response data and train reward models with the selected results to aid model training, essentially conducting reinforcement learning from AI feedback (RLAIF). Although RLAIF can distill the evaluation ability of advanced LLMs into reward models, its reliability remains limited by potential knowledge gaps and hallucinations, particularly in high-accuracy domains like medicine.

Current evaluations of LLM-generated responses primarily emphasize the semantic relevance between responses and question contexts and the correctness of logical reasoning chains (Li et al., 2024). These criteria align with the implicit semantic relationships and explicit structural connections among entities in knowledge graphs (KGs). Since the inception of LLMs, KGs have been instrumental in tasks such as evaluation (Li et al., 2024), knowledge injection (Wang et al., 2023), and knowledge augmentation (Wen et al., 2023; Zhang et al., 2023b), due to their structured fact storage and annotation-free advantages. However, these approaches predominantly treat KGs as static knowledge repositories and leave LLMs to filter and select relevant facts. This overlooks the semantic associations between facts and fails to fully exploit the structured connectivity of KGs.

Considering that entities with high linkage criticality are more likely to reach each other during inference and engage in greater semantic interactions, both semantic relevance between factual entities and the strength of logical connections in KGs can serve as natural scoring mechanisms (Yasunaga et al., 2021; Lin et al., 2019; Luo et al., 2023). Building on this insight and inspired by RLAIF, we propose Reinforcement Learning from Knowledge Graph Feedback (RLKGF), which directly derives reward signals from KGs without manual annotations. RLKGF treats the KG itself as a reward model and assigns reinforcement learning (RL) rewards to LLM responses by simulating semantic information flow and logical link transmission between question and candidate response entities on relevant subgraphs—without the need for preference labeling or reward model training. The scoring process integrates local semantic aggregation and global path reasoning among factual entities. At the semantic level, RLKGF employs graph neural networks (GNNs) for node-level information exchange and computes semantic relevance scores between question and candidate response entities. Structurally, RLKGF initiates reasoning from the question entities via random walks across connected paths and transparently calculates the criticality of path-connected entities based on reachability probabilities. We validate RLKGF in medical dialogue diagnosis tasks. Experimental results demonstrate that RLKGF outperforms RLAIF in disease prediction accuracy, which proves RLKGF's effectiveness as a viable alternative to RLHF. Further comparisons with supervised fine-tuning and KG-based prompts highlight RLKGF's advantages in aligning LLMs with knowledge. Besides, to eliminate potential contamination from existing datasets, we also construct a new medical dialogue diagnosis dataset (MED-D) from unpublished electronic medical records. The contributions of this study are:

- We propose RLKGF, a novel method for deriving feedback on LLM outputs from knowledge graphs by integrating local semantic aggregation and global logical connections among factual entities.

- We introduce a new medical dialogue diagnosis dataset, MED-D, constructed from Chinese electronic medical records. MED-D includes 20 diseases, 351 symptoms, and 3992 dialogue samples.

- Experimental results demonstrate that RLKGF significantly improves LLM diagnostic accuracy compared to RLAIF. This proves RLKGF is a competitive alternative to RLHF for knowledge alignment.

2 Related Work

2.1 Reinforcement Learning from Feedback in LLMs

RLHF trains reward models on human-labeled preferences and optimizes policy gradients based on reward scores (Ouyang et al., 2022). It has proven effective in enhancing the helpfulness and knowledge accuracy of LLM outputs and is one of the key drivers behind LLM success (Bai et al., 2022a). However, the high cost of human preference annotation limits RLHF's scalability (Lee et al., 2023).

As LLM capabilities evolve, models have demonstrated human-like judgment in tasks such as summarization (Stiennon et al., 2020), which prompts researchers to leverage LLMs for output evaluation. Self-reflection uses LLMs to filter irrelevant information by assessing the relevance of generated responses to retrieved content (Asai et al., 2023). Self-refine employs LLMs for iterative feedback to improve output quality (Madaan et al., 2024), while Refiner uses an LLM-based critic to enhance logical consistency in chain-of-thought reasoning (Paul et al., 2024). Beyond these prompt-based approaches, RRHF (Yuan et al., 2023) and RLAIF (Lee et al., 2023) further explore utilizing LLM-generated feedback for model training. RRHF ranks responses from different sources using LLMs and optimizes models through Rank Loss. RLAIF introduces reinforcement learning (RL) from AI feedback, where a high-performing LLM annotates preferences across different responses and trains a reward model. Despite reducing the need for human labels, RLAIF faces challenges in specialized fields like medicine, where the demand for accuracy clashes with LLMs' knowledge gaps and hallucinations (Huang et al., 2025).

2.2 LLMs with Knowledge Graphs

KGs store factual evidence in a structured format, which enables both evidence retrieval and semantic aggregation of key entities (Lin et al., 2019; Yasunaga et al., 2021; Yan et al., 2024). The utilization of KGs in LLMs spans multiple aspects, including supervised fine-tuning (SFT) (Wang et al., 2023), retrieval-augmented generation (RAG) (Feng et al., 2023), and response evaluation (Li et al., 2024). Bencao (Wang et al., 2023) constructs Q&A pairs from medical KGs to supplement training data for fine-tuning medical LLMs. Several works (Zhang et al., 2023b; Wen et al., 2023; Jiang et al., 2023) retrieve relevant evidence from KGs prior as prompt to enhance response accuracy and knowledge richness. Greaselm (Zhang et al., 2022) integrates KG and textual information through prefix prompting to improve semantic fusion and correctness in Q&A tasks. Li et al. (Li et al., 2024) uses commonsense KGs to detect knowledge and logical errors in LLM-generated responses. These approaches show that leveraging KGs' implicit semantics and explicit logical connections can enhance LLM performance. However, most methods treat KGs merely as knowledge bases, failing to exploit their potential for semantic connectivity and logical link significance.

3 Method

In this section, we define the task and describe our method, which directly utilizes the structural and semantic information among factual entities in KGs to provide feedback on model responses.

3.1 Task Definition

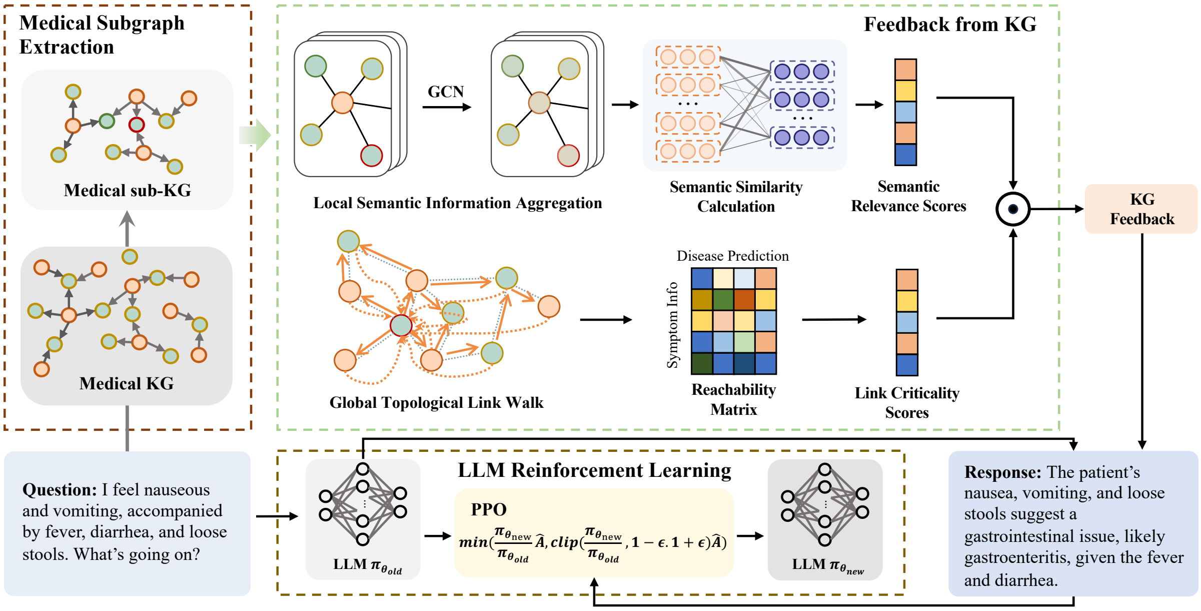
The disease diagnosis via Q&A task requires the model to predict a disease d in the answer A based on a patient's symptom description $[s_1, s_2, ..., s_n]$ in the question Q. Our focus is on using a medical knowledge graph (MKG) containing factual entities as a reward model to automatically assign feedback R to model responses, i.e., RLKGF.

The MKG $G = (V, E)$ is constructed based on the standard (Li et al., 2025), where V includes all symptom and disease entities in the dataset, and E represents the relationships between these entities. In this task, we only consider the relationship < disease, causes, symptom >. After extracting the patient's information from the question, RLKGF first locates the relevant symptom entities in the MKG. It then identifies the disease entities connected to those symptoms and any other symptoms that these diseases may cause. Using the entities and their relationships, we construct the personalized diagnostic subgraph $g = (v, e)$, where v represents the disease entities and related symptoms that may explain the patient's condition, and e represents the corresponding triple relationships.

RLKGF evaluates the correctness of the model's response through path reasoning and semantic aggregation using graph-based random walk with restart (RWR) (Tong et al., 2006) and GNNs, as detailed in section 3.2 and section 3.3. After acquiring feedback, RLKGF optimizes the model's policy using the proximal policy optimization (PPO) (Schulman et al., 2017) to align LLM responses with domain knowledge, as described in section 3.4.

3.2 Link Criticality Score via Structural Information

Evaluating model responses typically involves determining whether the response entity can be reached from the question entity through multi-step path reasoning, i.e., the correctness of the knowledge link. Additionally, the stronger the association between the question and candidate response entities along the path, the higher the probability of


Medical Subgraph Extraction
Medical sub-KG
Medical KG
GCN
Local Semantic Information Aggregation
Semantic Similarity Calculation
Semantic Relevance Scores
Feedback from KG
KG Feedback
Global Topological Link Walk
Disease Prediction
Symptom Info
Reachability Matrix
Link Criticality Scores
LLM Reinforcement Learning
PPO
$min(\frac{\pi_{\theta_{new}}}{\pi_{\theta_{old}}}\hat{A}, clip(\frac{\pi_{\theta_{new}}}{\pi_{\theta_{old}}}, 1-\epsilon.1+\epsilon)\hat{A})$
LLM $\pi_{\theta_{old}}$
LLM $\pi_{\theta_{new}}$

Question: I feel nauseous and vomiting, accompanied by fever, diarrhea, and loose stools. What's going on?

Response: The patient's nausea, vomiting, and loose stools suggest a gastrointestinal issue, likely gastroenteritis, given the fever and diarrhea.

Figure 2: The framework of RLKG.

reaching the response entity (Yasunaga et al., 2021). Based on this, we apply RWR on the global paths of the patient diagnosis subgraph. Starting from the question entities, RWR calculates the probability of reaching various candidate response entities, which serves as their link criticality score. The calculation process is as follows.

For a central entity i, we define the path connectivity reachability from other entities in the knowledge graph $g = (v, e)$ as w_i, where $w_i \in \mathbb{R}^{N\times 1}$ and $N = |v|$. The value $w_i(j)$ is initialized by Gaussian kernel function:

$$w_i(j) = exp(-\frac{dis(i,j)^2}{2h^2}) \tag{1}$$

where h is the Gaussian bandwidth and $j \in v$.

The vector w_i can be iteratively updated through RWR on the graph, as shown in Equation 2. Specifically, the random walk begins at the central entity i, with a probability of $1 - c$ to return to i and a probability of c to reach other entities along the connected path. After several iterations, until convergence, $w_i(j)$ represents the probability of reaching entity j. Thus, w_i captures the path-based association weights of various factual entities in the KG g relative to the central entity.

$$w_i^{'} = c \cdot \tilde{A}_i w_i + (1-c) \cdot e_i \tag{2}$$

$\tilde{A}_i \in \mathbb{R}^{N\times N}$ is the probability transition matrix for entity i, obtained by column normalization of the adjacency matrix A_i of g. The element in the i-th row and j-th column represents the connection flux from entity j to entity i, i.e., $\frac{w_{ij}}{w_j}$, where w_j is the sum of weights of all paths associated with entity j, and w_{ij} is the weight between entities i and j. $e_i \in \mathbb{R}^{N\times 1}$ is the starting node vector, with a value of 1 for the central entity and 0 for all other entities.

For the patient-specific diagnostic subgraph $g = (v, e)$ and the symptom entities in the question $[s_1, s_2, ..., s_n]$, we compute w_{s_i} for each symptom entity s_i to capture the link criticality scores W of each entity in g relative to the question entities, where $W \in \mathbb{R}^{n\times N}$. From this, we extract the link criticality matrix W^* between possible response entities (diseases to be predicted) and question entities, where $W^* \in \mathbb{R}^{n\times m}$. By normalization, we acquire the structural path reasoning-based score R^P, which quantifies the correctness of the knowledge links in the model's response.

$$R^p = \sigma(W^*) \tag{3}$$

where $R^p \in \mathbb{R}^m$, and m represents the number of disease entities in the personalized diagnostic subgraph g. $\sigma(\cdot)$ denotes the temperature softmax (Hinton, 2015).

3.3 Semantic Relevance Score via Semantic Aggregation

Semantic relevance between the model's generation and the question is another important criterion for evaluating response quality (Li et al., 2024). To capture the semantic connections between factual entities, we utilize graph convolutional networks (GCNs) (Kipf and Welling, 2016), which induce node representations via iterative message passing between neighbors on the graph. Specifically, we apply a 2-layer GCN to iteratively process the feature matrix $X \in \mathbb{R}^{N \times F}$ of the factual entities in the graph g, where F is the feature dimension. The information propagation between layers is updated by Equation 4.

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{4}$$

where $\tilde{A} = A + I_N$ is the adjacency matrix with self-connections, $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix and I_N is the identity matrix. $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ represents the degree matrix of entities, which serves for normalization. $W^{(l)}$ denotes the weight matrix for feature mapping. $H^{(l)} \in \mathbb{R}^{N \times D}$ is the input of the l-th layer of the neural network, with $H^{(0)} = X$.

Utilizing a 2-layer GCN to enable semantic information interaction between entities, the semantic feature representations of all entities are obtained as shown in Equation 5, where $Z \in \mathbb{R}^{N \times F}$.

$$Z = f(X, A) = \sigma(\hat{A}(ReLU(\hat{A}XW^0)W^1)) \tag{5}$$

$\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ and $\sigma(\cdot)$ indicates softmax.

We compute the semantic cosine similarity matrix S^* between the diseases to be predicted and the symptom entities mentioned in the question through the semantic features of all entities, where $S^* \in \mathbb{R}^{n \times m}$.

$$S^*(s_i, d_j) = \frac{Z_{s_i} \cdot Z_{d_j}}{||Z_{s_i}|| \cdot ||Z_{d_j}||} \tag{6}$$

s_i and d_j represent the symptom entity mentioned in the question and the potential disease entities in the response, respectively. The relevance score of the response to the question, derived from the semantic aggregation of factual entities, is denoted as $R^s \in \mathbb{R}^m$.

$$R^s = \sigma(S^*) \tag{7}$$

The feedback reward for the response, directly obtained from the KG, is calculated based on both the link criticality of the structural paths and the semantic aggregation relevance, as shown in Equation 8, where μ is a learnable parameter.

$$R = \mu(R^s) + (1 - \mu)(R^p) \tag{8}$$

3.4 Reinforcement Learning Training Framework

We employ PPO to implement reinforcement learning training for the LLMs. The policy $\pi_{\theta_{old}}$ is initialized from the off-the-shelf LLMs and then optimized to $\pi_{\theta_{new}}$ by maximizing the reward obtained from the knowledge graph. To avoid excessive policy shifts that could lead to unreasonable responses, we use PPO-Clipped, which restricts model updates within a certain range. The optimization objective is given by Equation 9.

$$L^{CLIP}(\theta) = \mathbb{E}[\min(r(\theta)A^*, \text{clip}(r(\theta), 1 - \epsilon, 1 + \epsilon)A^*)] \tag{9}$$

where $r(\theta) = \frac{\pi_{\theta_{new}}}{\pi_{\theta_{old}}}$, and A^* is the advantage function estimated for the model's decisions. The hyperparameter ϵ constrains the policy update ratio within the range $[1 - \epsilon, 1 + \epsilon]$ via $clip(r(\theta), 1 - \epsilon, 1 + \epsilon)$.

4 Experiment

4.1 Experimental Setup

We implement the model based on the PyTorch framework and conduct training and testing on one A800 80G GPU. The meanings and specific settings of each hyper-parameter involved in the model are detailed in Table 7.

4.2 Baselines

Models. We select smaller-scale, open-source LLMs that can be trained on a single 80G A100 GPU as the backbone. These include seven LLMs from the Qwen1.5 (Bai et al., 2023), Qwen2.5 (Yang et al., 2024a), InternLM2 (Cai et al., 2024), and InternLM2.5 (Wu et al., 2024) series.

Methods. To comprehensively evaluate the performance of RLKGF, we compare it with RLAIF (Lee et al., 2023) (using GPT-4o-mini for preferences (Achiam et al., 2023)), SFT (including full parameter tuning and LoRA) (Hu et al., 2021), as well as the knowledge graph-based prompt technique (Zhang et al., 2023b; Wen et al., 2023).

4.3 Datasets

We utilize three public medical dialogue diagnosis datasets: MZ (Wei et al., 2018), DXY (Xu et al., 2019), and GMD (Liu et al., 2022). These datasets are derived from real-world medical dialogue diagnosis records, with the number of diseases, symptoms, and dialogues summarized in Table 1.

To avoid potential data leakage, where public data might have been used for LLM training, we construct a new dataset, MED-D. MED-D is collected from offline electronic medical records (EMRs). These EMRs are sourced from cooperating hospitals and have been anonymized. We filter 14,277 EMRs and choose 20 diseases that could be

Dataset	MZ	DXY	GMD	MED-D
Language	Chinese	Chinese	Chinese	Chinese
# Diseases	4	5	12	20
# Symptoms	66	41	118	351
# Dialogue Samples	710	526	2390	3992
# Avg. Symptoms/Q&A	5.61	4.77	5.47	17.57

Table 1: Medical Dialogue Datasets. "# Avg. Symptoms/Patient" signifies the average number of symptoms per patient in the dataset.

diagnosed through Q&A without additional tests. With the assistance of medical experts, we identify 351 associated symptoms. Subsequently, we extract symptom and disease entities from the selected EMRs using named entity recognition to construct Q&A pairs. All extracted diseases and symptoms are manually aligned with the corresponding ICD-9 terms and reviewed by domain experts. We use the accuracy of disease prediction as the evaluation metric. The complete data processing procedure can be found in Appendix A.1.2.

4.4 Main Results

RLAIF vs. RLKGF. Table 2 shows the performance of different LLMs trained using only RLAIF and RLKGF. **RLKGF Base** represents the accuracy achieved by directly selecting the response entity with the highest feedback score from the knowledge graph. From the experimental results, we observe the following. Detailed prompts and analysis can be found in Appendix A.1.4.

i. Advantages of RLKGF. Our results demonstrate that RLKGF outperforms RLAIF by 5.67%, 10.73%, 8.38%, and 1.21% across four datasets, respectively. This indicates the feasibility and effectiveness of using KGs for feedback on model responses. It validates that leveraging KGs as reward models in the medical domain may be a more reliable approach than LLM-based preference labeling.

ii. Small models are limited by instruction adherence. Among different models, Qwen2.5-0.5b-instruct performs poorly, with only 32.39% on the MZ dataset. We analyze its outputs before and after training and find that it has poor instruction adherence and fails to make correct predictions from the given diseases. Although training improves its instruction-following ability, knowledge injection remains suboptimal. In section 4.6, we present the performance of models trained with supervised fine-tuning, where full-parameter SFT on Qwen2.5-0.5b-instruct achieves only 7.60% accuracy. Preliminary analysis suggests that the MZ dataset's sparsity is insufficient to correct the initially learned model parameters. Additionally, smaller models may be more sensitive to loss design, and how to better inject knowledge into them requires further investigation.

Backbone	Method	GMD	DXY	MZ	MED-D
GPT-4o-mini	Base	0.6460	0.4262	0.5289	0.5345
RLKGF	Base	0.7908	0.8252	0.6846	0.805
Qwen2.5-3B -Instruct	Base	0.6360	0.4531	0.3789	0.3553
	RLAIF	0.6722	0.6537	0.5469	0.3600
	RLKGF	**0.7113**	**0.7314**	**0.6268**	**0.3800**
Qwen2.5-1.5B -Instruct	Base	0.4840	0.2359	0.1845	0.1982
	RLAIF	0.5635	0.4595	0.4343	0.2908
	RLKGF	**0.6109**	**0.5890**	**0.5070**	**0.3025**
Qwen2.5-0.5B -Instruct	Base	0.2469	0.0981	0.0042	0.1273
	RLAIF	0.3092	0.2135	0.0282	0.1350
	RLKGF	**0.3278**	**0.2654**	**0.3239**	**0.1475**
Qwen1.5-4B -Chat	Base	0.4038	0.4000	0.4176	0.1893
	RLAIF	0.5816	0.3139	0.5610	0.2083
	RLKGF	**0.5914**	**0.6893**	**0.5986**	**0.2525**
Qwen1.5-1.8B -Chat	Base	0.3335	0.2291	0.0423	0.1342
	RLAIF	0.4686	0.2783	0.3568	0.1650
	RLKGF	**0.4784**	**0.3366**	**0.3592**	**0.2050**
InternLM2.5 -1.8B-Chat	Base	0.2092	0.3981	0.4507	0.1850
	RLAIF	0.4393	0.4369	0.5493	0.1950
	RLKGF	**0.5356**	**0.4757**	**0.5704**	**0.2025**
InternLM2.5 -1.8B-Chat	Base	0.3305	0.2718	0.2042	0.1667
	RLAIF	0.2929	0.4078	0.4507	0.2175
	RLKGF	**0.4686**	**0.4272**	**0.5282**	**0.2300**

Table 2: RLKGF vs. RLAIF. The bolded values represent the best performance of the current model on the dataset.

iii. Explore more effective KG feedback methods. Furthermore, by comparing the prediction accuracy of models using only the KG, KG feedback-trained models, GPT-4o-mini predictions, RLAIF, and SFT, we find that although trained LLMs show some performance improvement, they still fall far short of the optimal target. Therefore, further exploration is needed on how to fully utilize factual knowledge and construct reasonable feedback to guide model training.

RLAIF with different LLMs With the optimization and updates of LLMs, many open-source models have surpassed the GPT series in certain applications, such as Qwen2.5-72B (Yang et al., 2024a) and DeepSeek (Liu et al., 2024). We replace GPT-4o-mini with these two models for response preference labeling and compare their potential advantages. The model comparison results are shown in Table 3.

The results indicate that GPT-4o-mini outper-

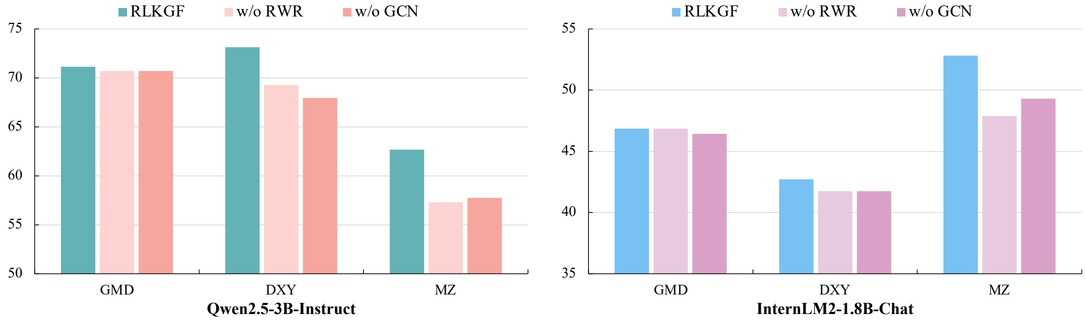


Figure 3: "w/o RWR" refers to the ablation of the link criticality score. "w/o GCN" refers to the ablation of the semantic relevance score.

Backbone	Method	GMD	DXY	MZ
Qwen2.5-3B -Instruct	GPT-4o-mini	0.6722	**0.6537**	**0.5469**
	Qwen2.5-72B	**0.6792**	0.6408	0.4671
	DeepSeekV3	0.6778	0.6505	0.3850
Qwen2.5-1.5B -Instruct	GPT-4o-mini	**0.5635**	0.4595	**0.4343**
	Qwen2.5-72B	0.5563	**0.5275**	0.2371
	DeepSeekV3	0.5593	0.3042	0.4108
Qwen1.5-4B -Chat	GPT-4o-mini	0.5816	0.3139	0.5610
	Qwen2.5-72B	**0.6025**	**0.6246**	**0.5822**
	DeepSeekV3	0.5816	0.5599	**0.5822**
InternLM2.5 -1.8B-Chat	GPT-4o-mini	**0.4393**	**0.4369**	0.5493
	Qwen2.5-72B	0.3096	0.3010	0.5423
	DeepSeekV3	0.3305	0.3010	**0.5563**
InternLM2 -1.8B-Chat	GPT-4o-mini	0.2929	**0.4078**	0.4507
	Qwen2.5-72B	**0.4519**	**0.4078**	**0.4577**
	DeepSeekV3	0.4477	0.3883	0.1972

Table 3: RLAIF with different LLMs.

forms other competitive open-source LLMs in the medical domain. Although high-capacity open LLMs enhance performance, leveraging existing knowledge bases for feedback presents a viable and effective alternative, particularly when considering resource efficiency and performance.

RLKGF demonstrates better generalization capability compared to SFT. We conduct an experiment to compare the generalization ability of RLKGF and SFT. Specifically, we use Qwen2.5-3B-Instruct as the base model, train it on the GMD dataset using both RLKGF and SFT and evaluate the models on the DXY dataset. As shown in Table 4, the model trained with RLKGF demonstrates clear generalization to a different dataset, while the SFT-trained model even underperforms the untrained baseline, which highlights the superior generalization capability of RLKGF.

4.5 Ablation Study

Component Ablation. As shown in Figure 3, ablating the link criticality scores derived from structural information via RWR leads to an average performance decrease of 1.73%, 1.67%, and 2.28% on the GMD, DXY, and MZ. Similarly, removing the semantic relevance feedback from semantic features results in a performance drop of 1.76%, 1.49%, and 2.65%. These findings highlight the importance of both global structural and semantic information from the knowledge graph in evaluating LLM responses. From the results, it can be seen that, in general, structural information plays a more significant role compared to semantic information. We preliminarily attribute this to the fact that KGs inherently extract factual knowledge into structured information, which results in two key characteristics: 1) Structural features are its distinguishing advantage over contextual knowledge; 2) The semantic information contained in the KGs is

Backbone	Method	DXY
Qwen2.5-3b-Instruct	Base	0.4531
	RLKGF on GMD	**0.6117**
	SFT on GMD	0.3680

Table 4: Generalization comparison between RLKGF and SFT.

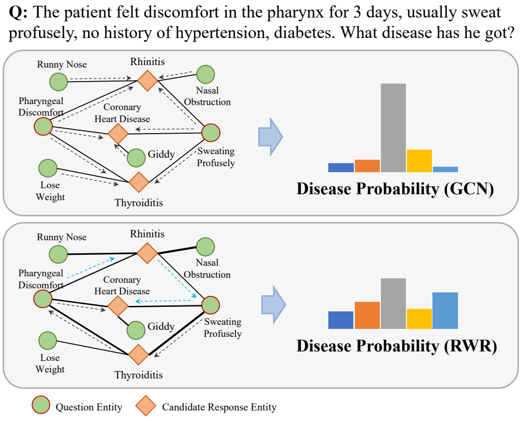


Figure 4: A case demonstrating GCN's local semantic aggregation and RWR's global reachability.

Backbone	Method	GMD	DXY	MZ
Qwen2.5-3B -Instruct	with GCN with GAT	**0.7113** 0.6987	**0.7314** 0.6990	**0.6268** 0.5822
Qwen2.5-1.5B -Instruct	with GCN with GAT	**0.6109** 0.5914	**0.5890** 0.5696	0.5070 **0.5305**
Qwen2.5-0.5B -Instruct	with GCN with GAT	**0.3278** 0.3152	**0.2654** 0.2233	**0.3239** 0.2089
Qwen1.5-4B -Chat	with GCN with GAT	**0.5914** 0.5872	**0.6893** 0.6246	**0.5986** 0.5728
Qwen1.5-1.8B -Chat	with GCN with GAT	**0.4784** 0.714	0.3366 **0.3754**	0.3592 **0.3850**
InternLM2.5 -1.8B-Chat	with GCN with GAT	**0.5356** 0.5356	**0.4757** 0.4660	0.5704 0.5704
InternLM2 -1.8B-Chat	with GCN with GAT	**0.4686** 0.4477	**0.4272** 0.4078	**0.5282** 0.4437

Table 5: Aggregating Semantic Information using GCN and GAT for LLMs Semantic Relevance Feedback.

not as rich as that in medical textbooks.

Figure 4 illustrates a case where structural or semantic features dominate. GCN relies on local neighbor interactions for representation learning. When the question entity is more strongly connected to a candidate response entity in its local neighborhood, it has a greater influence on the prediction. In contrast, RWR considers the global topological structure of the knowledge graph. This allows RWR to assign different weights to candidate response entities based on the global connections between them and the question entity, which GCN does not capture. The complete experimental results are presented in Appendix A.1.4.

Different Semantic Aggregation Models. As discussed in section 4.4, the semantic information in KGs is relatively concise. Therefore, further exploration for leveraging KG semantics is essential for providing more accurate feedback on the semantic relevance of LLMs responses. To preserve the structured semantics of KGs and ensure the generalizability of the method, we employ a dual-head graph attention mechanism (GAT) (Veličković et al., 2017) to dynamically assign weights to different neighbors and compute semantic relevance scores between response entities and question entities using attention weights. The results are shown in Table 5.

The results indicate that RLKGF with GAT outperforms RLAIF 4.57%, 8.60%, and 5.23%, which validates the advantage of using GAT to aggregate structured semantics for feedback. However, RLKGF with GCN yields an average advantage of 1.10%, 2.13%, and 3.15% over GAT. We find that the overall prediction accuracy achieved with GAT-trained attention weights is lower than that of GCN. This may be due to GCN's deeper learning of node representations.

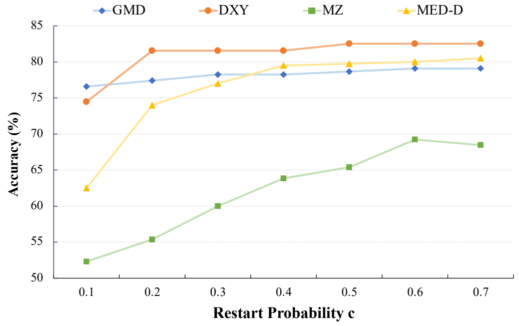


Figure 5: RLKGF Base obtained with different restart probabilities c.

The Impact of Restart Probability c. In section 3.2, we iteratively update the reachability matrix of entities relative to the question entities through RWR. The parameter $1 - c$ represents the probability of returning to the initial entity during each random walk. By setting different values of c, we investigate the impact of structural information, as described in Figure 5. The results show that setting a larger c helps improve the accuracy of reward feedback. This is because reducing the probability of returning to the initial node during RWR enables the model to explore a broader range of triple relationships, allowing for more comprehensive use of the global structural information in knowledge graphs and a more accurate assessment of connection flux across knowledge links.

4.6 Further Analysis

Analysis of Knowledge Injection Methods. To assess the performance of current mainstream methods for integrating knowledge into LLMs, we compare Full Fine-Tuning (FT), Low-Rank Adaptation (LoRA) (Hu et al., 2021), and Prompt techniques (Zhang et al., 2023b). The KG-based prompt used in Table 9 does not involve retrieval; instead, it directly provides the relevant patient subgraph in triple format (i.e., containing accurate information) to the LLMs. We include the results and the prompt used in Appendix A.1.4 and A.1.3. KG Prompt (Triple) refers to directly feeding KG triples into LLMs, while KG Prompt (Text) converts the triples into textual prompts.

Diseases	Relationship Completeness Rate	Diagnostic Accuracy
Allergic Rhinitis	0.6250	0.8888
Upper Respiratory Infection	0.7826	0.7391
Pneumonia	0.5517	0.2941
Hand Foot and Mouth Disease	0.4090	0.55
Pediatric Diarrhea	0.3333	0.25
Total	0.5267	0.5243

Table 6: The impact of knowledge graph completeness on RLKGF diagnostic accuracy.

The results lead to several key observations: 1) Although converting triples to text reduces input length, the associations between symptoms and diseases become less explicit compared to structured triples, leading to performance drops in many models. 2) Supervised fine-tuning remains the most effective method for knowledge injection, especially with large datasets. 3) The performance gap between LoRA and FT is minimal. 4) LLMs are capable of capturing correct information from extensive prompts, but this ability diminishes with sparse data. Additionally, training on domain-specific data improves comprehension of longer texts. As noted in RLAIF (Lee et al., 2023), RLHF and RLAIF typically achieve around 70% of the performance of SFT. Our RLKGF method consistently meets this standard, further validating its effectiveness.

RLKGF's Dependence on KG Completeness. RLKGF calculates link criticality and semantic relevance between entities based on the global topology and semantic aggregation of KGs, which is compatible with different KG schemas. However, the effectiveness of RLKGF is affected by KG incompleteness. To investigate this, we conduct experiments on the DXY dataset using the Chinese MKG proposed by Li et al. (Li et al., 2025).

Taking all <disease, causes, symptom> triples in DXY as the reference, we evaluate the coverage of such relations in the MKG and the diagnosis accuracy achieved using this MKG. As shown in Table 6, incomplete KGs result in decreased performance compared to experiments with a complete KG (see Table 2). Moreover, diseases with lower KG coverage exhibit significantly lower diagnostic accuracy. This limitation of RLKGF is also discussed in Section 5.

Case Study. We present two examples from the GMD dataset in Figure 10. In Example 1, the model's diagnosis aligns with the expert's primary diagnosis, while in Example 2, the model makes an incorrect diagnosis. Analysis of the diagnostic subgraph reveals that in Example 1, although the patient's symptoms are related to multiple diseases, the diagnosis is relatively straightforward and less challenging. In Example 2, as the number of diseases linked to the symptoms increases, the model struggles to pinpoint the correct diagnosis. This indicates that while RLKGF can achieve initial alignment between the model and knowledge, its performance still lags behind human experts in complex reasoning scenarios (i.e., identifying accurate results from multiple related diseases).

5 Conclusion

The semantic correlations and link criticality inherent in KGs closely mirror the semantic and logical relevance humans use to evaluate LLM responses. Building on this, we propose RLKGF, which directly employs KGs as a reward model to provide feedback to LLMs without the need for human annotation or separate reward model training. RLKGF utilizes both local semantic interactions and global path reachability to define reinforcement learning rewards. In the context of medical dialogue diagnosis, RLKGF outperforms RLAIF, which relies on model-embedded knowledge. We also compare various knowledge injection methods, such as SFT and KG-based prompts, offering valuable insights into the effective use of KGs. Although this work highlights the potential of RLKGF, several limitations remain. First, its generalization to other tasks and domains has not been explored. Additionally, we employ PPO to train LLMs, and there may be more suitable reward structures and training methods to explore.

Limitations

Although this work demonstrates the potential of RLKGF, several issues need to be addressed. The quality of feedback derived from knowledge graphs depends heavily on the completeness and accuracy of the graph itself, particularly in open domains. Our experiments are limited to disease diagnosis tasks without exploring RLKGF's generalization to other tasks and domains. Additionally, due to data limitations, we do not conduct experiments across a broader medical framework.

The current task format is single-turn Q&A, and future work should explore multi-turn dialogues to better leverage the potential advantages of knowledge graph structure and semantics in multi-step reasoning. Moreover, RLKGF currently focuses primarily on entity-level feedback for model responses, with limited focus on overall response fluency. Furthermore, experimental comparisons show that although RLKGF improves consistency between model responses and knowledge, there is still significant room for enhancement. Designing appropriate reward ranges and investigating the impact of different methods on model parameter adjustments are crucial for continuous knowledge learning.

Acknowledgments

We thank the anonymous reviewers for their insightful comments. This work was supported in part by the Science and Technology Innovation Award of Heilongjiang Province, China [JD2023GJ01], the Key Research and Development Program of Heilongjiang Province, China [2024ZX01A07], and the National Natural Science Foundation of China [72293584, 72431004]. Jingchi Jiang is the corresponding author of this work.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2023. Self-rag: Learning to retrieve, generate, and critique through self-reflection. *arXiv preprint arXiv:2310.11511*.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. 2022a. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. 2022b. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.

Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, et al. 2024. Internlm2 technical report. *arXiv preprint arXiv:2403.17297*.

Zheng Chu, Jingchang Chen, Qianglong Chen, Weijiang Yu, Tao He, Haotian Wang, Weihua Peng, Ming Liu, Bing Qin, and Ting Liu. 2024. Navigate through enigmatic labyrinth a survey of chain of thought reasoning: Advances, frontiers and future. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1173–1203, Bangkok, Thailand. Association for Computational Linguistics.

Bosheng Ding, Chengwei Qin, Linlin Liu, Yew Ken Chia, Boyang Li, Shafiq Joty, and Lidong Bing. 2023. Is gpt-3 a good data annotator? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11173–11195.

Zhangyin Feng, Weitao Ma, Weijiang Yu, Lei Huang, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2023. Trends in integration of knowledge and large language models: A survey and taxonomy of methods, benchmarks, and applications. *arXiv preprint arXiv:2311.05876*.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

Geoffrey Hinton. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans. Inf. Syst.*, 43(2).

Quzhe Huang, Mingxu Tao, Chen Zhang, Zhenwei An, Cong Jiang, Zhibin Chen, Zirui Wu, and Yansong Feng. 2023. Lawyer llama technical report. *arXiv preprint arXiv:2305.15062*.

Xinke Jiang, Ruizhe Zhang, Yongxin Xu, Rihong Qiu, Yue Fang, Zhiyuan Wang, Jinyi Tang, Hongxin Ding, Xu Chu, Junfeng Zhao, et al. 2023. Think and retrieval: A hypothesis knowledge graph enhanced medical large language models. *arXiv preprint arXiv:2312.15883*.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Kellie Ren Lu, Thomas Mesnard, Johan Ferret, Colton Bishop, Ethan Hall, Victor Carbune, and Abhinav Rastogi. 2023. Rlaif: Scaling reinforcement learning from human feedback with ai feedback.

Xue Li, Jia Su, Yang Yang, Zipeng Gao, Xinyu Duan, and Yi Guan. 2024. Dialogues are not just text: Modeling cognition for dialogue coherence evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18573–18581.

Xue Li, Ye Yuan, Yang Yang, Yi Guan, Haotian Wang, Jingchi Jiang, Huaizhang Shi, and Xiguang Liu. 2025. Quality-controllable automatic construction method of chinese knowledge graph for medical decision-making applications. *Information Processing & Management*, 62(4):104148.

Bill Yuchen Lin, Xinyue Chen, Jamin Chen, and Xiang Ren. 2019. Kagnet: Knowledge-aware graph networks for commonsense reasoning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2829–2839.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Wenge Liu, Yi Cheng, Hao Wang, Jianheng Tang, Yafei Liu, Ruihui Zhao, Wenjie Li, Yefeng Zheng, and Xiaodan Liang. 2022. "my nose is running.""are you also coughing?": Building a medical diagnosis agent with interpretable inquiry logics. In *Proceedings of the International Conferences on Artificial Intelligence*.

Linhao Luo, Yuan-Fang Li, Gholamreza Haffari, and Shirui Pan. 2023. Reasoning on graphs: Faithful and interpretable large language model reasoning. *arXiv preprint arXiv:2310.01061*.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. 2024. Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, 36.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Debjit Paul, Mete Ismayilzada, Maxime Peyrard, Beatriz Borges, Antoine Bosselut, Robert West, and Boi Faltings. 2024. Refiner: Reasoning feedback on intermediate representations. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1100–1126.

Ruoling Peng, Kang Liu, Po Yang, Zhipeng Yuan, and Shunbao Li. 2023. Embedding-based retrieval with llm for effective agriculture information extracting from unstructured data. *arXiv preprint arXiv:2308.03107*.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021.

Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. 2006. Fast random walk with restart and its applications. In *Sixth international conference on data mining (ICDM'06)*, pages 613–622. IEEE.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Haochun Wang, Chi Liu, Nuwa Xi, Zewen Qiang, Sendong Zhao, Bing Qin, and Ting Liu. 2023. Huatuo: Tuning llama model with chinese medical knowledge. *arXiv preprint arXiv:2304.06975*.

Haotian Wang, Xiyuan Du, Weijiang Yu, Qianglong Chen, Kun Zhu, Zheng Chu, Lian Yan, and Yi Guan. 2025. Learning to break: Knowledge-enhanced reasoning in multi-agent debate system. *Neurocomputing*, 618:129063.

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

Zhongyu Wei, Qianlong Liu, Baolin Peng, Huaixiao Tou, Ting Chen, Xuan-Jing Huang, Kam-Fai Wong, and Xiang Dai. 2018. Task-oriented dialogue system for automatic diagnosis. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 201–207.

Yilin Wen, Zifeng Wang, and Jimeng Sun. 2023. Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models. *arXiv preprint arXiv:2308.09729*.

Zijian Wu, Suozhi Huang, Zhejian Zhou, Huaiyuan Ying, Jiayu Wang, Dahua Lin, and Kai Chen. 2024. Internlm2. 5-stepprover: Advancing automated theorem proving via expert iteration on large-scale lean problems. *arXiv preprint arXiv:2410.15700*.

Lin Xu, Qixian Zhou, Ke Gong, Xiaodan Liang, Jianheng Tang, and Liang Lin. 2019. End-to-end knowledge-routed relational dialogue system for automatic diagnosis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7346–7353.

Lian Yan, Yi Guan, Haotian Wang, Yi Lin, Yang Yang, Boran Wang, and Jingchi Jiang. 2024. Eirad: An evidence-based dialogue system with highly interpretable reasoning path for automatic diagnosis. *IEEE Journal of Biomedical and Health Informatics*.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. 2024a. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Songhua Yang, Hanjie Zhao, Senbin Zhu, Guangyu Zhou, Hongfei Xu, Yuxiang Jia, and Hongying Zan. 2024b. Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19368–19376.

Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. 2021. Qa-gnn: Reasoning with language models and knowledge graphs for question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 535–546.

Zheng Yuan, Hongyi Yuan, Chuanqi Tan, Wei Wang, Songfang Huang, and Fei Huang. 2023. Rrhf: Rank responses to align language models with human feedback without tears. *arXiv preprint arXiv:2304.05302*.

Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu, Zhihong Chen, Jianquan Li, Guiming Chen, Xiangbo Wu, Zhiyi Zhang, Qingying Xiao, et al. 2023a. Huatuogpt, towards taming language model to be a doctor. *arXiv preprint arXiv:2305.15075*.

Qinggang Zhang, Junnan Dong, Hao Chen, Xiao Huang, Daochen Zha, and Zailiang Yu. 2023b. Knowgpt: Black-box knowledge injection for large language models. *arXiv preprint arXiv:2312.06185*.

X Zhang, A Bosselut, M Yasunaga, H Ren, P Liang, C Manning, and J Leskovec. 2022. Greaselm: Graph reasoning enhanced language models for question answering. In *International Conference on Representation Learning (ICLR)*.

A Appendix

A.1 Experiment

A.1.1 Experimental Setup

The meanings and specific settings of each hyper-parameter involved in the model are detailed in Table 7. Although the parameter μ can be treated as a learnable variable, we empirically find that assigning it a small fixed value leads to better performance. This may be attributed to the relatively sparse semantic information present in the current dataset.

Hyper-parameter	**Meaning**	**Setting**
batch size	Batch size of training	16
update frequency	Policy update frequency	50
ϵ	PPO-Clipped parameter	0.2
γ	Discount factor of RL	0.99
lr	Initial learning rate	1.00E-05
train epochs	Number of training epochs	5
hidden size	Hidden neuron size of GCN	128
F	Semantic feature dimension	100
c	RWR restart probability	0.7

Table 7: Hyper-parameter settings. The meanings and specific settings of each hyper-parameter.

A.1.2 Datasets

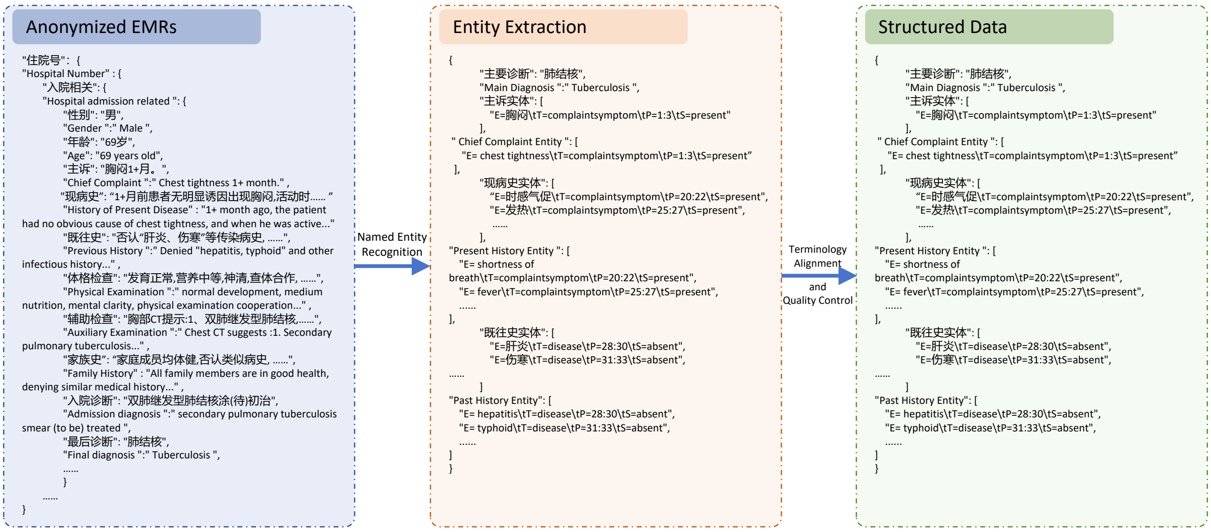
We construct a new dataset, MED-D. MED-D is collected from offline EMRs. The details on data sources, anonymization, data extraction, and quality control are shown in Figure 6.

- **Data Source and Anonymization.** Our EMRs originate from a tertiary hospital in Guangdong Province, China. All data have been anonymized by removing sensitive information such as patient names, addresses, ID numbers, and contact details. Only key fields, including chief complaints, medical history, past history, diagnoses, and auxiliary examinations, are retained.

- **Data Selection and Annotation.** We filter 14,277 EMRs and choose 20 diseases that could be diagnosed through Q&A without additional tests. With the assistance of medical experts, we identify 351 associated symptoms. We then employ a medical named entity recognition (NER) model (trained using BERT-BiLSTM-CRF (Li et al., 2025)) to automatically extract symptom and disease entities, including entity names, types, locations, and negation/affirmation labels.

- **Terminology Alignment.** To address synonymous symptom expressions (e.g., "lower back pain" vs. "lumbar pain"), all extracted symptoms and diseases are manually aligned with the corresponding ICD-9 terms.

- **Quality Control and Expert Evaluation.** To ensure data quality, we hire four professional doctors to assess the data for textual accuracy, diagnostic correctness, and symptom relevance. Each entry is validated by at least two experts and rated on a 0-5 scale (5 for the best). Annotation consistency is measured using the



```
"住院号"：{
"Hospital Number" : {
        "入院相关": {
        "Hospital admission related ": {
                "性别": "男",
                "Gender ":" Male ",
                "年龄": "69岁",
                "Age": "69 years old",
                "主诉": "胸闷1+月。",
                "Chief Complaint ":" Chest tightness 1+ month." ,
               "现病史”: “1+月前患者无明显诱因出现胸闷,活动时...... ”
                "History of Present Disease" : "1+ month ago, the patient
had no obvious cause of chest tightness, and when he was active..."
                "既往史": "否认“肝炎、伤寒”等传染病史, ......",
                "Previous History ":" Denied "hepatitis, typhoid" and other
infectious history..." ,
                "体格检查": "发育正常,营养中等,神清,查体合作, ......",
                "Physical Examination ":" normal development, medium
nutrition, mental clarity, physical examination cooperation..." ,
                "辅助检查": "胸部CT提示:1、双肺继发型肺结核,......",
                "Auxiliary Examination ":" Chest CT suggests :1. Secondary
pulmonary tuberculosis..." ,
                "家族史”: “家庭成员均体健,否认类似病史, ......",
                "Family History" : "All family members are in good health,
denying similar medical history..." ,
                "入院诊断": "双肺继发型肺结核涂(待)初治",
                "Admission diagnosis ":" secondary pulmonary tuberculosis
smear (to be) treated ",
                "最后诊断": "肺结核",
                "Final diagnosis ":" Tuberculosis ",
                ......
                }
        ......
}
```

```
{
          "主要诊断": "肺结核",
          "Main Diagnosis ":" Tuberculosis ",
          "主诉实体": [
             "E=胸闷\tT=complaintsymptom\tP=1:3\tS=present"
          ],
 " Chief Complaint Entity ": [
    "E= chest tightness\tT=complaintsymptom\tP=1:3\tS=present”
  ],
          "现病史实体": [
             “E=时感气促\tT=complaintsymptom\tP=20:22\tS=present",
             "E=发热\tT=complaintsymptom\tP=25:27\tS=present",
              ......
          ],
"Present History Entity ": [
   "E= shortness of
breath\tT=complaintsymptom\tP=20:22\tS=present",
   "E= fever\tT=complaintsymptom\tP=25:27\tS=present",
   ......
],
          "既往史实体": [
             "E=肝炎\tT=disease\tP=28:30\tS=absent",
             "E=伤寒\tT=disease\tP=31:33\tS=absent",
......
          ]
"Past History Entity": [
   "E= hepatitis\tT=disease\tP=28:30\tS=absent",
   "E= typhoid\tT=disease\tP=31:33\tS=absent",
   ......
]
}
```

```
{
          "主要诊断": "肺结核",
          "Main Diagnosis ":" Tuberculosis ",
          "主诉实体": [
             "E=胸闷\tT=complaintsymptom\tP=1:3\tS=present"
          ],
 " Chief Complaint Entity ": [
    "E= chest tightness\tT=complaintsymptom\tP=1:3\tS=present”
  ],
          "现病史实体": [
             “E=时感气促\tT=complaintsymptom\tP=20:22\tS=present",
             "E=发热\tT=complaintsymptom\tP=25:27\tS=present",
              ......
          ],
"Present History Entity ": [
   "E= shortness of
breath\tT=complaintsymptom\tP=20:22\tS=present",
   "E= fever\tT=complaintsymptom\tP=25:27\tS=present",
   ......
],
          "既往史实体": [
             "E=肝炎\tT=disease\tP=28:30\tS=absent",
             "E=伤寒\tT=disease\tP=31:33\tS=absent",
......
          ]
"Past History Entity": [
   "E= hepatitis\tT=disease\tP=28:30\tS=absent",
   "E= typhoid\tT=disease\tP=31:33\tS=absent",
   ......
]
}
```

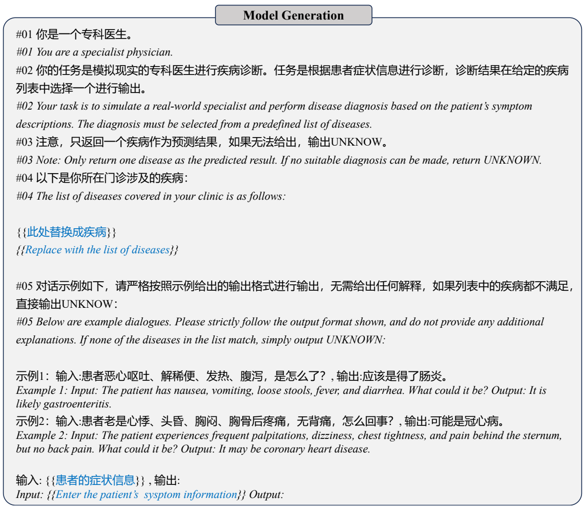
Figure 6: The Dataset Construction Process.

Intraclass Correlation Coefficient (ICC), and only data with ICC > 0.9 are retained.

A.1.3 Prompt

For datasets like MZ/DXY/GMD, only patient symptom entities and doctor diagnosis labels are available. We construct the inputs by concatenating symptom information using a templated approach. Specifically, the data includes two types of symptoms: "True" symptoms, which the patient has, and "False" symptoms, which the patient does not have. We convert these symptoms into text format, as illustrated in Figure 7 and Figure 8.

The Prompts Applied for Model Generation. The prompt used for LLM generation is shown in Figure 7.



Model Generation

#01 你是一个专科医生。
#01 You are a specialist physician.
#02 你的任务是模拟现实的专科医生进行疾病诊断。任务是根据患者症状信息进行诊断，诊断结果在给定的疾病列表中选择一个进行输出。
#02 Your task is to simulate a real-world specialist and perform disease diagnosis based on the patient's symptom descriptions. The diagnosis must be selected from a predefined list of diseases.
#03 注意，只返回一个疾病作为预测结果，如果无法给出，输出UNKNOW。
#03 Note: Only return one disease as the predicted result. If no suitable diagnosis can be made, return UNKNOWN.
#04 以下是你所在门诊涉及的疾病：
#04 The list of diseases covered in your clinic is as follows:

{{此处替换成疾病}}
{{Replace with the list of diseases}}

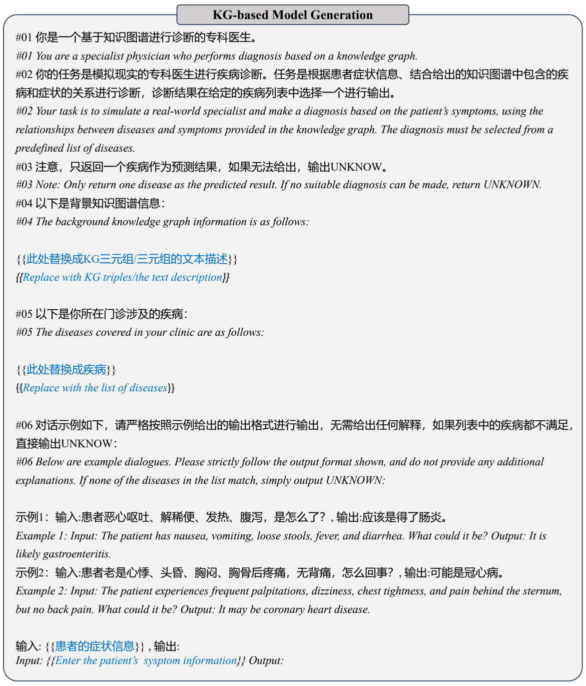
#05 对话示例如下，请严格按照示例给出的输出格式进行输出，无需给出任何解释，如果列表中的疾病都不满足，直接输出UNKNOW：
#05 Below are example dialogues. Please strictly follow the output format shown, and do not provide any additional explanations. If none of the diseases in the list match, simply output UNKNOWN:

示例1：输入:患者恶心呕吐、解稀便、发热、腹泻，是怎么了？, 输出:应该是得了肠炎。
Example 1: Input: The patient has nausea, vomiting, loose stools, fever, and diarrhea. What could it be? Output: It is likely gastroenteritis.
示例2：输入:患者老是心悸、头昏、胸闷、胸骨后疼痛，无背痛，怎么回事？, 输出:可能是冠心病。
Example 2: Input: The patient experiences frequent palpitations, dizziness, chest tightness, and pain behind the sternum, but no back pain. What could it be? Output: It may be coronary heart disease.

输入: {{患者的症状信息}}, 输出:
Input: {{Enter the patient's sysptom information}} Output:

Figure 7: The Prompts Applied for Model Generation.

KG-based Prompt. The knowledge graph as a prompt input to LLMs is shown in Figure 8.



KG-based Model Generation

#01 你是一个基于知识图谱进行诊断的专科医生。
#01 You are a specialist physician who performs diagnosis based on a knowledge graph.
#02 你的任务是模拟现实的专科医生进行疾病诊断。任务是根据患者症状信息、结合给出的知识图谱中包含的疾病和症状的关系进行诊断，诊断结果在给定的疾病列表中选择一个进行输出。
#02 Your task is to simulate a real-world specialist and make a diagnosis based on the patient's symptoms, using the relationships between diseases and symptoms provided in the knowledge graph. The diagnosis must be selected from a predefined list of diseases.
#03 注意，只返回一个疾病作为预测结果，如果无法给出，输出UNKNOW。
#03 Note: Only return one disease as the predicted result. If no suitable diagnosis can be made, return UNKNOWN.
#04 以下是背景知识图谱信息：
#04 The background knowledge graph information is as follows:

{{此处替换成KG三元组/三元组的文本描述}}
{{*Replace with KG triples/the text description*}}

#05 以下是你所在门诊涉及的疾病：
#05 The diseases covered in your clinic are as follows:

{{此处替换成疾病}}
{{*Replace with the list of diseases*}}

#06 对话示例如下，请严格按照示例给出的输出格式进行输出，无需给出任何解释，如果列表中的疾病都不满足，直接输出UNKNOW：
#06 Below are example dialogues. Please strictly follow the output format shown, and do not provide any additional explanations. If none of the diseases in the list match, simply output UNKNOWN:

示例1：输入:患者恶心呕吐、解稀便、发热、腹泻，是怎么了？, 输出:应该是得了肠炎。
Example 1: Input: The patient has nausea, vomiting, loose stools, fever, and diarrhea. What could it be? Output: It is likely gastroenteritis.
示例2：输入:患者老是心悸、头昏、胸闷、胸骨后疼痛，无背痛，怎么回事？, 输出:可能是冠心病。
Example 2: Input: The patient experiences frequent palpitations, dizziness, chest tightness, and pain behind the sternum, but no back pain. What could it be? Output: It may be coronary heart disease.

输入: {{患者的症状信息}}, 输出:
Input: {{Enter the patient's sysptom information}} Output:

Figure 8: KG-based Prompt for Model Generation.

A.1.4 Experiment Analysis

RLAIF vs. RLKGF. *Model Parameters and Version Iterations.* Comparing LLMs within the same series but with different parameter sizes (e.g., Qwen1.5-3b vs. Qwen1.5-1.5b), larger models consistently perform better and show more substantial improvements after training. This suggests that larger parameter sizes help models learn more knowledge. Additionally, newer versions within

Backbone	Method	GMD	DXY	MZ
Qwen2.5-3B-Instruct	RLKGF	**0.7113**	**0.7314**	**0.6268**
	w/o RWR	0.7071	0.6926	0.5728
	w/o GCN	0.7071	0.6796	0.5775
Qwen2.5-1.5B-Instruct	RLKGF	**0.6109**	**0.5890**	0.5070
	w/o RWR	0.5788	0.5696	**0.5516**
	w/o GCN	0.5872	0.5825	0.4953
Qwen2.5-0.5B-Instruct	RLKGF	**0.3278**	**0.2654**	**0.3239**
	w/o RWR	0.3180	0.2388	0.2653
	w/o GCN	0.3180	0.2388	0.2512
Qwen1.5-4B-Chat	RLKGF	**0.5914**	**0.6893**	**0.5986**
	w/o RWR	0.5886	0.6472	0.5798
	w/o GCN	0.5844	0.6667	0.5822
Qwen1.5-1.8B-Chat	RLKGF	**0.4784**	0.3366	**0.3592**
	w/o RWR	0.4603	**0.3657**	0.3286
	w/o GCN	0.4756	0.3495	0.3592
InternLM2.5-1.8B-Chat	RLKGF	**0.5356**	**0.4757**	0.5704
	w/o RWR	0.4812	0.4660	**0.5775**
	w/o GCN	0.4644	0.4757	0.5704
InternLM2-1.8B-Chat	RLKGF	**0.4686**	**0.4272**	**0.5282**
	w/o RWR	0.4686	0.4175	0.4789
	w/o GCN	0.4644	0.4175	0.4930

Table 8: "w/o RWR" refers to the ablation of the link criticality score obtained using structural information. "w/o GCN" refers to the ablation of the semantic relevance score obtained through semantic features.

the same series outperform older ones, likely due to the inclusion of more knowledge and optimized training methods.

The model struggles to solve more complex problems. Across multiple datasets, we observe that as dataset size increases, model performance tends to decline. This not only indicates that LLMs struggle to achieve high accuracy across broader scenarios but also poses a challenge to KG-based scoring. As the number of entities grows, questions become longer, complicating the model's ability to learn from extended texts. Additionally, the gap in scores between different entities from the KG may shrink, which could lead to a more uniform distribution, as shown in Figure 9. This is similar to human preferences, where selecting the best option from fewer answers is relatively easier.

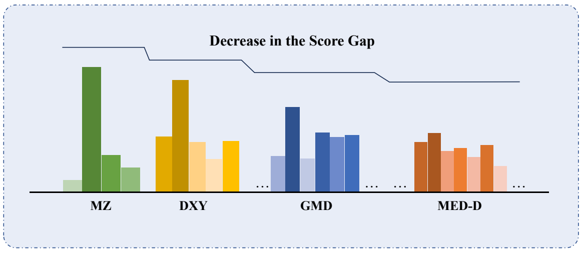


Figure 9: As the number of selectable responses increases, the score gap narrows.

Component Ablation The ablation results can be found in Table 8.

Analysis of Knowledge Injection Methods. Table 9 shows the performance of different knowledge injection methods.

Backbone	Method	GMD	DXY	MZ	MED-D
GPT-4o-mini	Base	0.6460	0.4262	0.5289	0.5345
	KG Prompt (Triple)	0.7569	0.7563	0.6275	0.6638
	KG Prompt (Text)	0.6731	0.5772	0.5669	-
Qwen2.5-3B -Instruct	Base	0.6360	0.4531	0.3789	0.3553
	FT	**0.7552**	0.4951	0.4253	0.4823
	LoRA	0.7334	0.5038	0.4591	**0.4843**
	KG Prompt (Triple)	0.7054	0.6495	**0.6648**	0.4671
	KG Prompt (Text)	0.6452	0.5126	0.5768	0.4655
	RLKGF	0.7113	**0.7314**	0.6268	0.3800
	RLKGF + KG Prompt (Triple)	0.7490	-	-	-
Qwen2.5-1.5B -Instruct	Base	0.4840	0.2359	0.1845	0.1982
	FT	0.7066	0.4380	0.4035	**0.3863**
	LoRA	0.7041	0.3543	0.3929	0.3543
	KG Prompt (Triple)	0.6397	0.4680	0.4352	0.3825
	KG Prompt (Text)	0.6188	0.5757	0.4803	0.3575
	RLKGF	0.6109	**0.5890**	**0.5070**	0.3025
	RLKGF + KG Prompt (Triple)	**0.7113**	-	-	-
Qwen2.5-0.5B -Instruct	Base	0.2469	0.0981	0.0042	0.1273
	FT	**0.4920**	**0.3388**	0.0760	**0.2510**
	LoRA	0.4209	0.1252	0.0240	0.1860
	KG Prompt (Triple)	0.4490	0.1515	0.0556	0.1525
	KG Prompt (Text)	0.2870	0.1388	0.0373	0.1225
	RLKGF	0.3278	0.2654	**0.3239**	0.1475
	RLKGF + KG Prompt (Triple)	0.4519	-	-	-
Qwen1.5-4B -Chat	Base	0.4038	0.4000	0.4176	0.1893
	FT	0.6866	0.6067	0.5260	**0.3956**
	LoRA	0.6485	0.6048	0.5556	0.3080
	KG Prompt (Triple)	0.5540	0.4350	0.4754	0.1722
	KG Prompt (Text)	0.4820	0.5184	0.3507	0.2022
	RLKGF	0.5914	**0.6893**	**0.5986**	0.2525
	RLKGF + KG Prompt (Triple)	**0.6987**	-	-	-
Qwen1.5-1.8B -Chat	Base	0.3335	0.2291	0.0423	0.1342
	FT	**0.5656**	0.3320	0.2408	**0.3233**
	LoRA	0.5364	0.2864	0.1795	0.2600
	KG Prompt (Triple)	0.2970	0.2786	0.0894	0.0392
	KG Prompt (Text)	0.3326	0.2641	0.1188	0.1105
	RLKGF	0.4784	**0.3366**	**0.3592**	0.2050
	RLKGF + KG Prompt (Triple)	**0.5690**	-	-	-
InternLM2.5 -1.8B-Chat	Base	0.2092	0.3981	0.4507	0.1850
	FT	**0.7573**	**0.7184**	**0.5985**	**0.4683**
	LoRA	0.5828	0.5563	0.5859	0.3916
	KG Prompt (Triple)	0.2594	0.4757	0.4648	-
	KG Prompt (Text)	0.3180	0.3204	0.3380	0.1800
	RLKGF	0.5356	0.4757	0.5704	0.2025
	RLKGF + KG Prompt (Triple)	0.6192	-	-	-
InternLM2 -1.8B-Chat	Base	0.3305	0.2718	0.2042	0.1667
	FT	**0.7280**	**0.7766**	**0.6760**	**0.4836**
	LoRA	0.7012	0.7116	0.6394	0.4080
	KG Prompt (Triple)	0.2971	0.3883	0.0634	-
	KG Prompt (Text)	0.3556	0.3204	0.0775	0.0875
	RLKGF	0.4686	0.4272	0.5282	0.2300
	RLKGF + KG Prompt (Triple)	0.4979	-	-	-

Table 9: Comparison of Different Knowledge Injection Methods.

Case Study. We compare LLMs' responses with expert diagnoses, presenting two examples in Figure 10.

Case Study 1

患者目前头晕，耳痛伴有听力下降，大概是什么疾病导致的?

The patient is currently experiencing dizziness and ear pain accompanied by hearing loss. What is the most likely disease causing these symptoms?

医学知识图谱子图:

Medical Knowledge Graph Subgraph: ('鼻炎', '导致', '鼻塞'), ('鼻炎', '导致', '咽部不适')…… ('鼻炎', '导致', '眼痒'), ('鼻炎', '导致', '头晕'),('鼻炎', '导致', '耳痛'), ('脑外伤', '导致', '头痛'), ('脑外伤', '导致', '头晕'), ('脑外伤', '导致', '呕吐'), ('脑外伤', '导致', '意识障碍'), ('脑外伤', '导致', '颈前疼痛'), ('脑外伤', '导致', '耳鸣'), ('脑外伤', '导致', '发热')…… ('外耳炎', '导致', '耳鸣'), ('外耳炎', '导致', '恶心'), ('外耳炎', '导致', '呕吐'), ('外耳炎', '导致', '耳痒'), ('外耳炎', '导致', '耳痛'), ('外耳炎', '导致', '听力下降'), ('外耳炎', '导致', '鼻塞'), ('外耳炎', '导致', '发热'), ('外耳炎', '导致', '咽部不适')

(‘Rhinitis’, ‘causes’, ‘Nasal Congestion’), (‘Rhinitis’, ‘causes’, ‘Throat Discomfort’),......, (‘Rhinitis’, ‘causes’, ‘Itchy Eyes’), (‘Rhinitis’, ‘causes’, ‘Dizziness’), (‘Rhinitis’, ‘causes’, ‘Ear Pain’), (‘Brain Trauma’, ‘causes’, ‘Headache’), (‘Brain Trauma’, ‘causes’, ‘Dizziness’), (‘Brain Trauma’, ‘causes’, ‘Vomiting’), (‘Brain Trauma’, ‘causes’, ‘Consciousness Disorder’), (‘Brain Trauma’, ‘causes’, ‘Anterior Neck Pain’), (‘Brain Trauma’, ‘causes’, ‘Tinnitus’), (‘Brain Trauma’, ‘causes’, ‘Fever’), ..., (‘Otitis Externa’, ‘causes’, ‘Tinnitus’), (‘Otitis Externa’, ‘causes’, ‘Nausea’), (‘Otitis Externa’, ‘causes’, ‘Vomiting’), (‘Otitis Externa’, ‘causes’, ‘Itchy Ears’), (‘Otitis Externa’, ‘causes’, ‘Ear Pain’), (‘Otitis Externa’, ‘causes’, ‘Hearing Loss’), (‘Otitis Externa’, ‘causes’, ‘Nasal Congestion’), (‘Otitis Externa’, ‘causes’, ‘Fever’), (‘Otitis Externa’, ‘causes’, ‘Throat Discomfort’)

RLKGF训练后模型回复：由症状判断大概率是得了外耳炎。

RLKGF-Based Model Prediction: Based on the symptoms, the most likely disease is Otitis Externa.

专家诊断标签：外耳炎。

Expert Diagnosis Label: Otitis Externa.

Case Study 2

患者目前咳嗽，鼻塞，咽部不适且流涕、发热，应该得了什么病?

The patient is currently experiencing cough, nasal congestion, throat discomfort, runny nose, and fever. What is the most likely disease?

医学知识图谱子图:

Medical Knowledge Graph Subgraph:

('肠炎', '导致', '发热'), ('肠炎', '导致', '呕吐'), ('肠炎', '导致', '腹痛'), ('肠炎', '导致', '口渴'), ('肠炎', '导致', '咽部不适'), ('肠炎', '导致', '腹泻')……('哮喘', '导致', '咽部不适'), ('哮喘', '导致', '气喘'), ('哮喘', '导致', '咳嗽'), ('哮喘', '导致', '鼻塞'),('哮喘', '导致', '发热')…… ('冠心病', '导致', '胸闷'), ('冠心病', '导致', '胸骨后疼痛'), ('冠心病', '导致', '心悸'), ('冠心病', '导致', '乏力'), ('冠心病', '导 致', '发热'), ('冠心病', '导致', '咽部不适')……('肺炎', '导致', '发热'), ('肺炎', '导致', '咳嗽'), ('肺炎', '导致', '鼻塞'), ('肺炎', '导致', '流涕'), ('肺炎', '导致', '气喘'), ('肺炎', '导致', '咳痰'), ('肺炎', '导致', '咽部不适'), ('肺炎', '导致', '胸闷气促'),…… ('鼻炎', '导致', '鼻塞'), ('鼻炎', '导致', '咽部不适'), ('鼻炎', '导致', '咳嗽'), ('鼻炎', '导致', '流涕'), ('鼻炎', '导致', '头痛'), ('鼻炎', '导致', '眼痒'), ('鼻炎', '导致', '呼吸困难'), ('鼻炎', '导致', '鼻出血')……('甲状腺炎', '导致', '颈前疼痛'), ('甲状腺炎', '导致', '乏力'), ('甲状腺炎', '导致', '咽部不适'), ('甲状腺炎', '导致', '甲状腺轻度肿大'),('甲状腺炎', '导致', '发热')……('皮炎', '导致', '皮疹'), ('皮炎', '导致', '瘙痒'), ('皮炎', '导致', '咽部不适'), ('皮炎', '导致', '发热')……('外耳炎', '导致', '耳鸣'), ('外耳炎', '导致', '耳痒'), ('外耳炎', '导致', '耳痛'), ('外耳炎', '导致', '听力下降'), ('外耳炎', '导致', '鼻塞'), ('外耳炎', '导致', '发热'), ('外耳炎', '导致', '咽部不适'), ('外耳炎', '导致', '咳嗽')……

(‘Enteritis’, ‘causes’, ‘Fever’), (‘Enteritis’, ‘causes’, ‘Vomiting’), (‘Enteritis’, ‘causes’, ‘Abdominal Pain’), (‘Enteritis’, ‘causes’, ‘Thirst’), (‘Enteritis’, ‘causes’, ‘Throat Discomfort’), (‘Enteritis’, ‘causes’, ‘Diarrhea’), ..., (‘Asthma’, ‘causes’, ‘Throat Discomfort’), (‘Asthma’, ‘causes’, ‘Wheezing’), (‘Asthma’, ‘causes’, ‘Cough’), (‘Asthma’, ‘causes’, ‘Nasal Congestion’), (‘Asthma’, ‘causes’, ‘Fever’), ..., (‘Coronary Heart Disease’, ‘causes’, ‘Chest Tightness’), (‘Coronary Heart Disease’, ‘causes’, ‘Retrosternal Pain’), (‘Coronary Heart Disease’, ‘causes’, ‘Palpitations’), (‘Coronary Heart Disease’, ‘causes’, ‘Fatigue’), (‘Coronary Heart Disease’, ‘causes’, ‘Fever’), (‘Coronary Heart Disease’, ‘causes’, ‘Throat Discomfort’), ..., (‘Pneumonia’, ‘causes’, ‘Fever’), (‘Pneumonia’, ‘causes’, ‘Cough’), (‘Pneumonia’, ‘causes’, ‘Nasal Congestion’), (‘Pneumonia’, ‘causes’, ‘Runny Nose’), (‘Pneumonia’, ‘causes’, ‘Wheezing’), (‘Pneumonia’, ‘causes’, ‘Sputum Production’), (‘Pneumonia’, ‘causes’, ‘Throat Discomfort’), (‘Pneumonia’, ‘causes’, ‘Shortness of Breath’), ..., (‘Rhinitis’, ‘causes’, ‘Nasal Congestion’), (‘Rhinitis’, ‘causes’, ‘Throat Discomfort’), (‘Rhinitis’, ‘causes’, ‘Cough’), (‘Rhinitis’, ‘causes’, ‘Runny Nose’), (‘Rhinitis’, ‘causes’, ‘Headache’), (‘Rhinitis’, ‘causes’, ‘Itchy Eyes’), (‘Rhinitis’, ‘causes’, ‘Dyspnea’), (‘Rhinitis’, ‘causes’, ‘Nasal Bleeding’), ..., (‘Thyroiditis’, ‘causes’, ‘Anterior Neck Pain’), (‘Thyroiditis’, ‘causes’, ‘Fatigue’), (‘Thyroiditis’, ‘causes’, ‘Throat Discomfort’), (‘Thyroiditis’, ‘causes’, ‘Mild Thyroid Enlargement’), (‘Thyroiditis’, ‘causes’, ‘Fever’), ..., (‘Dermatitis’, ‘causes’, ‘Rash’), (‘Dermatitis’, ‘causes’, ‘Itching’), (‘Dermatitis’, ‘causes’, ‘Throat Discomfort’), (‘Dermatitis’, ‘causes’, ‘Fever’), ..., (‘Otitis Externa’, ‘causes’, ‘Tinnitus’), (‘Otitis Externa’, ‘causes’, ‘Itchy Ears’), (‘Otitis Externa’, ‘causes’, ‘Ear Pain’), (‘Otitis Externa’, ‘causes’, ‘Hearing Loss’), (‘Otitis Externa’, ‘causes’, ‘Nasal Congestion’), (‘Otitis Externa’, ‘causes’, ‘Fever’), (‘Otitis Externa’, ‘causes’, ‘Throat Discomfort’), (‘Otitis Externa’, ‘causes’, ‘Cough’)

RLKGF训练后模型回复：由症状判断可能是鼻炎引起的。

RLKGF-Based Model Prediction: Based on the symptoms, the likely cause is Rhinitis.

专家诊断标签：肺炎。

Expert Diagnosis Label: Pneumonia.

Figure 10: Examples of LLM and expert diagnostic results.