DDO: Dual-Decision Optimization for LLM-Based Medical Consultation via Multi-Agent Collaboration

Zhihao Jia[1], Mingyi Jia[1], Junwen Duan[1]*, Jianxin Wang[1],
[1]Hunan Provincial Key Lab on Bioinformatics,
School of Computer Science and Engineering, Central South University

{zhihaojia, jiamingyi, jwduan}@csu.edu.cn, jxwang@mail.csu.edu.cn

Abstract

Large Language Models (LLMs) demonstrate strong generalization and reasoning abilities, making them well-suited for complex decision-making tasks such as medical consultation (MC). However, existing LLM-based methods often fail to capture the dual nature of MC, which entails two distinct sub-tasks: symptom inquiry, a sequential decision-making process, and disease diagnosis, a classification problem. This mismatch often results in ineffective symptom inquiry and unreliable disease diagnosis. To address this, we propose **DDO**, a novel LLM-based framework that performs **D**ual-**D**ecision **O**ptimization by decoupling the two sub-tasks and optimizing them with distinct objectives through a collaborative multi-agent workflow. Experiments on three real-world MC datasets show that DDO consistently outperforms existing LLM-based approaches and achieves competitive performance with state-of-the-art generation-based methods, demonstrating its effectiveness in the MC task. The code is available at https://github.com/zh-jia/DDO.

1 Introduction

Medical Consultation (MC), aiming to automate symptom collection and support clinical diagnosis, has become a promising application in AI-driven healthcare and attracted growing attention (Zhao et al., 2024; Hu et al., 2024; Chopra and Shah, 2025). As shown in Figure 1, MC involves multi-turn interactions between an AI doctor and a patient, encompassing two core decision-making processes: symptom inquiry—a sequential decision task over a large action space—and disease diagnosis—a classification task over a limited set of candidate diseases (Chen et al., 2023, 2024). The effectiveness of MC hinges on the AI doctor's ability to perform both efficient information seeking and accurate disease differentiation.

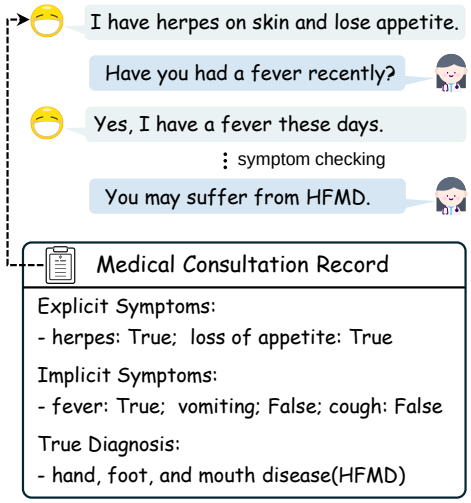

Figure 1: An example of a Medical Consultation (MC) task, where an AI doctor iteratively inquires about additional symptoms based on the patient's initial self-reported symptoms and ultimately provides a diagnosis.

Compared to the models used in traditional reinforcement learning (RL)-based methods (Zhong et al., 2022; Yan et al., 2023) and generation-based approaches (Hou et al., 2023; Zhao et al., 2024), large language models (LLMs) provide stronger generalization and more transparent reasoning (Qin et al., 2024; Singh et al., 2024), potentially reducing training costs and improving interpretability for the MC task. However, due to hallucinations and limited domain adaptation, directly applying LLMs to MC often results in inefficient inquiry and unreliable diagnosis. Recent approaches (Hu et al., 2024; Chopra and Shah, 2025) improve information gathering by incorporating LLM-based planning, yet overlook diagnostic optimization. To jointly optimize symptom inquiry and disease diagnosis, Chen et al. (2024) introduced the Chain-of-Diagnosis (CoD) framework, enabling LLMs to learn both components from CoD training data. Nevertheless, the fundamentally different nature of these two decision-making sub-tasks presents

*Corresponding Author. Email: jwduan@csu.edu.cn.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 26381–26398
November 4-9, 2025

significant challenges for unified optimization.

To overcome these limitations, we propose **DDO**, an LLM-based MC approach that leverages multi-agent collaboration to decouple and optimize the two core decision-making components with distinct objectives. For symptom inquiry, DDO integrates a lightweight RL-based policy agent that generates reliable candidate actions, thereby reducing the decision-making burden on LLMs. For disease diagnosis, DDO derives fine-grained diagnostic confidence from LLM logits and enhances disease discrimination through a plug-and-play adapter trained via in-batch contrastive learning. Experiments on three real-world MC datasets demonstrate that DDO consistently outperforms other LLM-based methods and achieves performance on par with state-of-the-art (SOTA) generation-based approaches, while requiring substantially less training overhead. Our contributions are as follows:

- We introduce DDO, a novel multi-agent framework for the MC task, where four collaborative agents enable an effective and transparent diagnostic MC workflow.
- DDO decouples the two core decision-making processes—symptom inquiry and disease diagnosis—and optimizes them with distinct objectives, leading to more informative questioning and improved diagnostic accuracy.
- By tuning only a small number of model parameters, DDO surpasses other LLM-based methods and achieves performance comparable to SOTA generation-based approaches.

2 Related Work

2.1 Medical Consultation Task

Medical Consultation (MC), a key application of AI in medicine (Valizadeh and Parde, 2022), was initially formulated as a Markov Decision Process (MDP) and optimized using reinforcement learning (RL) (Tang et al., 2016; Wei et al., 2018; Kao et al., 2018). However, due to the high variability of RL agents (Xia et al., 2020), researchers have incorporated disease-symptom prior knowledge to enhance the decision-making (Xu et al., 2019; Liu et al., 2022; Yan et al., 2023) by the agents. HRL (Zhong et al., 2022) introduced a hierarchical RL framework to refine the action space. Additionally, generative approaches such as Diaformer (Chen et al., 2022), CoAD (Wang et al., 2023), MTDiag (Hou et al., 2023) and HAIformer (Zhao et al., 2024) leveraged attention mechanisms to enhance optimization efficiency, achieving SOTA performance in the MC task. AIME (Tu et al., 2024) demonstrated the potential of LLMs in medical history-taking by learning from realistic medical dialogues. MediQ (Li et al., 2024b) introduced an abstention module to assess whether the collected diagnostic evidence is sufficient. UoT (Hu et al., 2024) and MISQ-HF (Chopra and Shah, 2025) aimed to reduce decision uncertainty through LLM-driven planning. CoD (Chen et al., 2024) improved the interpretability of doctor agent's decisions by generating transparent chained thought processes.

2.2 LLMs in Medical Decision-Making

LLMs have demonstrated strong potential across various medical applications (Zhou et al., 2024). They are capable of answering medical exam questions (Kim et al., 2024; Shi et al., 2024), collecting patient history (Johri et al., 2024), offering diagnostic suggestions (Jia et al., 2025; Rose et al., 2025), and recommending treatment plans (Li et al., 2024a). Leveraging prompt engineering (Zheng et al., 2024; Liu et al., 2024) and domain adaptation techniques (Tian et al., 2024; Wang et al., 2025), their reasoning capabilities have significantly improved, leading to more reliable medical decision-making. Moreover, to tackle more complex tasks in the medical domain, recent studies (Kim et al., 2024; Bani-Harouni et al., 2024) have explored the use of multiple LLM agents, offering promising directions for enabling collaborative decision-making in challenging clinical scenarios.

3 Problem Definition

A real-world Medical Consultation Record (MCR) is denoted as $\mathcal{P} = \{\mathcal{S}^{\text{ex}}, \mathcal{S}^{\text{im}}, d_l\}$, where $\mathcal{S}^{\text{ex}} = \{(s_i^{\text{ex}}, p_i^{\text{ex}})\}_{i=1}^{l_1}$ represents *explicit symptoms* initially reported by the patient, and *implicit symptoms* $\mathcal{S}^{\text{im}} = \{(s_j^{\text{im}}, p_j^{\text{im}})\}_{j=1}^{l_2}$ are elicited through follow-up inquiries by the doctor. The label d_l denotes the ground-truth disease of the patient.

The MC task simulates a multi-turn interaction process between an AI doctor and a simulated patient, where the AI doctor actively collects diagnostic information to facilitate differential diagnosis. Starting from the initial symptoms $\mathcal{S}^{\text{ex}}$, the AI doctor selectively inquires about additional symptoms $\mathcal{S}^{\text{ad}}$ to accumulate diagnostic evidence. The interaction terminates when sufficient information is collected or a predefined maximum number of turns L

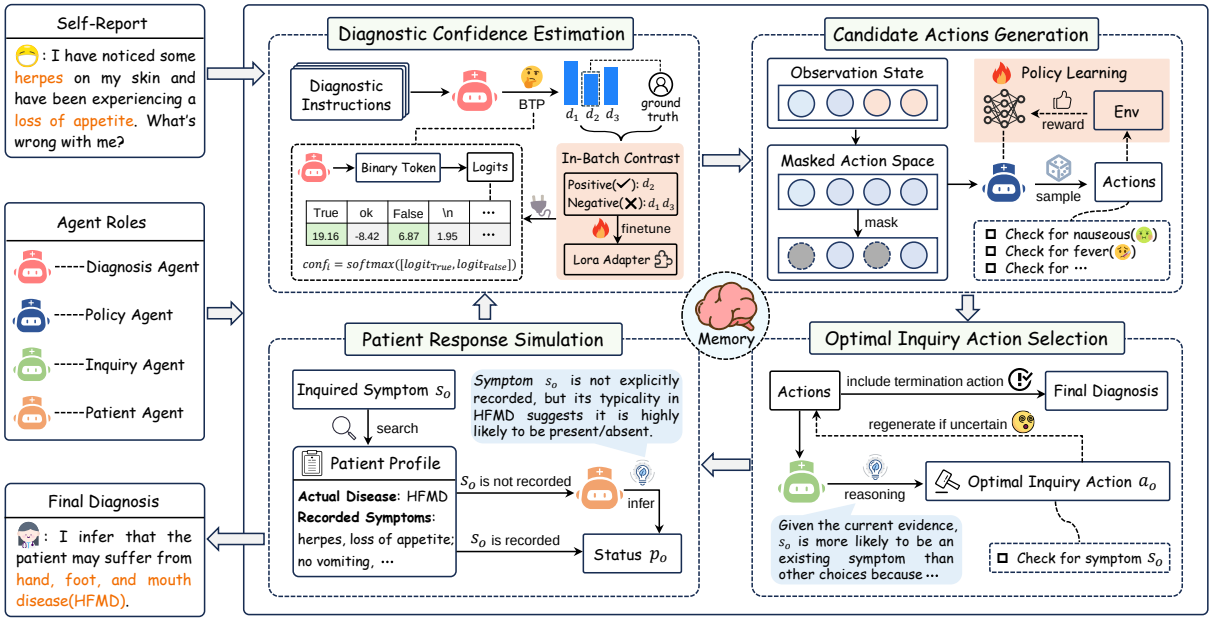

Figure 2: Overview of the proposed DDO framework, comprising four collaborative agents operating over a shared memory to execute the consultation workflow: the *Diagnosis Agent* estimates disease confidences from LLM logits; the *Policy Agent* generates candidate actions via masked sampling; the *Inquiry Agent* selects the optimal symptom to query or terminates the consultation; and the *Patient Agent* responds based on the patient profile.

,

is reached. The final diagnosis d_p is selected from the candidate set $D = \{d_1, d_2, \ldots, d_n\}$ based on the highest diagnostic confidence:

$$d_p = \arg\max_{d_i \in D} \text{conf}_i \left(\mathcal{S}^{\text{ex}} \cup \mathcal{S}^{\text{ad}}, d_i\right) \quad (1)$$

4 Multi-Agent Collaborative Medical Consultation in DDO

To enhance the decision-making capability of LLMs in the MC task, the DDO framework integrates three LLM-based agents—*Diagnosis Agent*, *Inquiry Agent*, and *Patient Agent*—alongside an RL-based *Policy Agent* implemented with an actor-critic architecture. The *Diagnosis Agent* employs a learnable adapter to enhance the LLM's ability to perform differential diagnosis, while the *Policy Agent* and the *Inquiry Agent* collaborate to strategically acquire informative symptoms. All agents operate over a shared ***memory***, consisting of (1) a static component that encodes prior disease–symptom knowledge exclusively derived from the training portion of the dataset (thus avoiding any leakage of test labels), and (2) a dynamic component that is continuously updated with observed symptoms and diagnostic confidence throughout the consultation process.

As shown in Figure 2, each consultation round proceeds through four steps: 1) **Diagnostic Confidence Estimation**—The *Diagnosis Agent* estimates confidence scores for each candidate disease based on the current diagnostic evidence. 2) **Candidate Actions Generation**—The *Policy Agent* samples multiple interaction actions based on the current state, providing a set of reliable choices for the next inquiry. 3) **Optimal Inquiry Action Selection**—The *Inquiry Agent* selects the most informative symptom checking action from the candidate actions. 4) **Patient Response Simulation**—The *Patient Agent* simulates the patient's responses, indicating whether the inquired symptom is present.

4.1 Diagnostic Confidence Estimation

4.1.1 Binary Token Probability

The *Diagnosis Agent* estimates the diagnostic confidence score $conf_i \in (0, 1)$ for each candidate disease $d_i \in D$, reflecting the likelihood of d_i being the correct diagnosis given the current evidence. Rather than relying on decoding to generate numeric scores (Li et al., 2024b; Chen et al., 2024; Qin et al., 2024), the *Diagnosis Agent* adopts **Binary Token Probability (BTP)**—a logit-based method inspired by multiple-choice QA (Detommaso et al., 2024; Kumar et al., 2024)—to provide a more efficient and interpretable confidence estimation for LLMs.

Specifically, given a structured prompt that in-

tegrates the current evidence with disease-specific knowledge of d_i retrieved from the shared memory, the LLM-based *Diagnosis Agent* is instructed to output a single binary token (`True/False`) indicating whether d_i is a plausible diagnosis. We extract the logits distribution at the position of this binary token and obtain the logits corresponding to `True` and `False`, denoted as logit_T and logit_F, respectively. The diagnostic confidence is finally computed via a temperature-scaled softmax over the binary logits, where the temperature τ controls the sharpness of the logits distribution:

$$conf_i = \frac{\exp(\text{logit}_\text{T}/\tau)}{\exp(\text{logit}_\text{T}/\tau) + \exp(\text{logit}_\text{F}/\tau)} \quad (2)$$

4.1.2 Calibrating the Diagnostic Confidence

Diagnostic confidence scores from base LLMs often lack discriminative power when candidate diseases share overlapping symptoms—e.g., both *upper respiratory tract infections* and *pneumonia* commonly present with *fever* and *cough*. Existing calibration methods typically require fine-grained supervision (Detommaso et al., 2024; Chen et al., 2024), such as expert-annotated confidence scores, which are often infeasible to implement in clinical practice. Instead, we treat diagnosis as a multi-class classification task (Ma et al., 2024) and leverage ground-truth disease labels d_l from Medical Consultation Records (MCRs) as weak supervision to calibrate the diagnostic confidence.

To construct the calibration training data, we generate partial consultation trajectories by truncating each full MCR at different interaction steps. For a training MCR $\mathcal{P}$ with k turns, we extract $(k - l_\text{self} + 1)$ sub-trajectories of the form $\mathcal{P}c = \{(s_1, p_1), \ldots, (s_c, p_c), d_l\}$, where l_self is the number of self-reported symptoms and $c < l$. Each sub-trajectory $\mathcal{P}_c$ serves as a training data.

We calibrate diagnostic confidence through in-batch contrastive learning (Ma et al., 2024), training a lightweight adapter using LoRA (Hu et al., 2022) to improve the *Diagnosis Agent*'s ability to distinguish among similar diseases. For each patient sub-trajectory $\mathcal{P}c$, the ground-truth diagnosis d_l is treated as the positive instance, while all other candidate diseases serve as negatives. We construct a target distribution $\text{dist}_\text{target} = [\epsilon, \ldots, 1-\epsilon, \ldots, \epsilon]$, where ϵ is a label smoothing constant. The *Diagnosis Agent* outputs confidence scores $\{conf_i\}_{i=1}^n$ using the BTP method, yielding a batch-level predictive distribution dist_diag. The calibration objective minimizes the KL divergence between the target and predicted distributions:

$$\mathcal{L}_\text{KL} = \sum_{i=1}^{n} \text{dist}_\text{target}(d_i) \log \frac{\text{dist}_\text{target}(d_i)}{\text{dist}_\text{diag}(d_i)} \quad (3)$$

4.2 Candidate Actions Generation

Symptom inquiry poses a significant challenge due to the high-dimensional action space, which limits the LLM's ability to identify the most informative symptoms. A natural solution is to reduce decision complexity by supplying a small set of reliable candidate symptoms. Since each inquiry depends only on the current state, the process satisfies the Markov property, making reinforcement learning (RL) well-suited for this task (Sun et al., 2024). Unlike RLHF approaches that fine-tune LLM parameters—such as GRPO (Ramesh et al., 2024) in DeepSeek-R1 (Guo et al., 2025)—we adopt a lightweight RL policy model as an external agent to guide the LLM's inquiry decisions.

4.2.1 Observation State and Action Space

In reinforcement learning, the observation state encodes the information available to the agent at each decision step, while the action space defines the set of allowable actions.

We define the observation state as $\mathcal{S} = [p, c]$, where $p \in \{-1, 0, 1\}^m$ is an m-dimensional symptom vector indicating absence (-1), unknown status (0), or presence (1) of each symptom (initialized to 0), and $c \in \mathbb{R}^n$ is a diagnostic confidence vector over n candidate diseases.

The action space $\mathcal{A} = \{a_i\}_{i=1}^{m+1}$ comprises m inquiry actions—each a_i corresponds to check for the i-th symptom—and a termination action a_{m+1} to end the consultation. To reflect clinical heuristics where physicians prioritize symptoms relevant to likely diagnoses (Stanley and Campos, 2013), we introduce a binary action mask $\mathcal{M} \in \{0, 1\}^{m+1}$ to constrain the action space. The mask enables actions ($\mathcal{M}_i = 1$) associated with symptoms relevant to the top-w ranked diseases and disables actions that have already selected or deemed irrelevant ($\mathcal{M}i = 0$). The final masked action space is:

$$\mathcal{A}_\text{masked} = \mathcal{A} \odot \mathcal{M} \quad (4)$$

4.2.2 RL Policy Learning

We adopt an actor-critic architecture to jointly learn the policy π, which is implemented via multi-layer perceptron (MLP) layers. The policy π outputs a

log-probability distribution over actions. Training is conducted using Proximal Policy Optimization (PPO) (Schulman et al., 2017), which maximizes the total reward return $\mathcal{R}$, composed of both short-term and long-term components.

The short-term reward $\mathcal{R}_{\text{short}}$ is computed after each doctor-patient interaction:

$$\mathcal{R}_{\text{short}}(S_t, a_t, S_{t+1}) = freq(a_t) + r_{\text{hit}} + r_{\text{rank}}, \quad (5)$$

where $freq(a_t)$ denotes the frequency of symptom s_t (corresponding to action a_t) among the relevant symptoms of the ground-truth disease d_l, with negative values assigned to irrelevant symptoms. The term r_{hit} is positive if s_t is present in the patient profile $\mathcal{P}$ and negative otherwise. The term r_{rank} measures the change in the confidence ranking of d_l from state S_t to S_{t+1}, assigning positive reward for improved ranking, negative for worsened ranking, and zero if unchanged.

The long-term reward $\mathcal{R}_{\text{long}}(d_p)$ assesses the final diagnostic prediction d_p, yielding a positive reward if $d_p = d_l$ and a negative reward otherwise.

4.2.3 Masked Sampling for Candidate Actions

The RL policy model's sampling nature inherently prevents the guarantee of optimal actions. However, by performing multiple sampling iterations, we can leverage this characteristic to provide the LLM with a reliable set of candidate actions, thus avoiding decision-making within a large action space.

Specifically, given the current state $\mathcal{S}$, the *Policy Agent* π samples N actions from the masked action space to construct a candidate action set A_{candi} (Equation 6). If the sampled candidate action set includes the termination action a_{m+1}, the consultation process is terminated.

$$A_{\text{candi}} = \{a_i \sim \pi(\mathcal{A}_{\text{masked}} \mid \mathcal{S})\}_{i=1}^{N} \quad (6)$$

4.3 Optimal Inquiry Action Selection

Upon receiving the candidate action set A_{candi}, the *Inquiry Agent* selects the most informative inquiry action a_o, corresponding to symptom s_o, by following a set of predefined reasoning strategies. Specifically, it either: (1) attempts to confirm the most probable diagnosis by prioritizing symptoms that are highly representative of the top-ranked disease, or (2) selects the symptom most relevant to the currently collected diagnostic evidence. If none of the candidate actions meet the selection criteria, the *Inquiry Agent* requests the *Policy Agent* to regenerate a new set of actions. By engaging in step-by-step reasoning, the *Inquiry Agent* ensures that the selected symptom s_o maximizes diagnostic value while maintaining interpretability and transparency throughout the decision-making process.

4.4 Patient Response Simulation

The *Patient Agent* responds to each inquiry from the *Inquiry Agent* based on the patient profile $\mathcal{P}$. If the queried symptom s_o is explicitly documented, it returns the recorded status p_o. However, since real-world Medical Consultation Records (MCRs) typically contain only a limited subset of symptom annotations, many queried symptoms may be undocumented, leading to ineffective queries and ambiguous responses.

To reduce this uncertainty bias, we leverage the disease label d_l from the MCR and incorporate clinical knowledge for inference. If the queried symptom is uncommon in the clinical presentation of d_l, the agent infers it is likely absent; if the symptom is strongly associated with d_l, it is inferred to be likely present. Once the presence status p_o is determined, the shared diagnostic memory is updated as: $E = E \cup \{(s_o, p_o)\}$.

In DDO, the multi-agent collaborative consultation proceeds for up to L turns, and terminates either when the turn limit is reached or when the termination action is sampled by the *Policy Agent*. The disease with the highest diagnostic confidence is then selected as the final diagnosis.

5 Experiments

5.1 Baselines

5.1.1 Traditional Methods.

We compare the proposed DDO framework with two state-of-the-art generation-based methods in the MC task: **MTDiag** (Hou et al., 2023), which independently optimizes symptom inquiry and disease diagnosis, and **HAIformer** (Zhao et al., 2024), which leverages human-AI collaboration. We additionally include **EBAD** (Yan et al., 2023) as a reinforcement learning-based baseline.

5.1.2 LLM-Based Methods

We compare the DDO framework with three LLM-based methods in the MC task: **Uncertainty of Thoughts (UoT)** (Hu et al., 2024), **Chain-of-Diagnosis (CoD)** (Chen et al., 2024), and **Direct Prompting (DP)**. UoT plans future inquiries by computing information gain. CoD employs instruction tuning to teach LLMs transparent diagnostic

reasoning. DP relies solely on the inherent capabilities of the LLM, without prompt engineering or instruction tuning.

5.2 Datasets

We evaluate the proposed DDO framework and the baseline methods on three real-world medical consultation datasets: DXY (Xu et al., 2019), collected from online medical dialogues, and GMD (Liu et al., 2022) and CMD (Yan et al., 2023), both derived from electronic medical records (EMRs). Dataset statistics are summarized in Table 1. Other dataset details can be found in Appendix A.1.

Dataset	DXY	GMD	CMD
# Total MCR Samples	527	2,374	5,200
# Disease Types	5	12	27
# Symptom Types	41	118	358
# Avg. Symptoms	4.74	5.55	17.92

Table 1: Dataset statistics, including the number of medical consultation records (MCR), disease types, symptom types, and average recorded symptoms per sample.

5.3 Evaluation Metrics

Diagnostic Accuracy. We use diagnostic accuracy (Acc) as an evaluation metric, which measures the proportion of test cases where the ground-truth disease is correctly identified from a limited set of candidate diseases. This metric reflects the model's ability to perform accurate differential diagnosis.

Average Turns. The average number of inquiry turns Avg.n is calculated as the total number of symptom queries divided by the number of test samples, reflecting the efficiency and informativeness of the multi-turn consultation process.

5.4 Implemention Details

For traditional baselines, we reproduce EBAD and HAIformer following their original implementations, while the MTDiag results are adopted from their paper. All LLM-based baselines are re-implemented, where UoT adopts its pruned version to improve planning efficiency. The backbone LLMs include the Qwen2.5 series (Yang et al., 2024), GPT-4o-mini (Achiam et al., 2023), and DiagnosisGPT (Chen et al., 2024), which is fine-tuned to align with the CoD consultation process. For methods that involve parameter tuning, including CoD and our proposed DDO, we use locally deployed LLMs; other LLM-based baselines rely on API-based models. The pruning percentage in UoT is set to -0.5. The threshold τ in CoD is set to 0.5. All LLM-based approaches adopt the same LLM for both doctor and patient agents. The maximum of doctor-patient interaction turns L is set to 10 for all the methods. More implementation details can be found in the Appendix A.

5.5 Overall Performance

Table 2 summarizes the main experimental results of the proposed DDO framework and baseline methods across the three MC datasets.

Comparison with Traditional Methods. DDO achieves diagnostic accuracy on par with traditional baselines while substantially reducing training overhead. For instance, the SOTA baseline HAIformer adopts a multi-stage training pipeline requiring hundreds of epochs for training its diagnostic module, while DDO only needs a few epochs for confidence calibration—less than one epoch on both the GMD and CMD datasets. This efficiency stems from the strong generalization capabilities of LLMs, which enable effective domain adaptation with a small number of parameter tuning. Moreover, the inherent reasoning ability of LLMs contributes to the interpretability of the MC task.

Comparison with LLM-based Methods. Compared to other LLM-based methods, DDO significantly improves diagnostic effectiveness. After symptom inquiry, it boosts diagnostic accuracy by an average of 24.6%, 11.3%, and 3.2% on the DXY, GMD, and CMD datasets, respectively, over initial diagnoses based only on self-reported symptoms. DDO consistently achieves the highest accuracy, notably 94.2% on DXY. The DP baseline reflects the raw inquiry behavior of LLMs, where the lack of external guidance results in arbitrary questioning and unreliable diagnoses. UoT improves upon DP by using LLM-based planning to prioritize symptoms with the highest expected information gain. However, its pruning strategy—eliminating candidate diseases as soon as key symptoms are denied—restricts comprehensive evidence gathering, often resulting in a small Avg.n and ultimately limiting diagnostic performance. CoD attempts to jointly optimize symptom inquiry and diagnosis via large-scale synthetic reasoning data but yields negative performance gains. This is likely due to the intrinsic differences between the two sub-tasks, which hinder effective unified learning. In contrast, DDO decouples the two decision-making processes in the MC task, enhancing them separately through a lightweight RL policy and a diagnostic adapter.

Method	LLM	DXY			GMD			CMD		
		Acc_{init}	Acc	Avg.n	Acc_{init}	Acc	Avg.n	Acc_{init}	Acc	Avg.n
Traditional Methods										
EBAD (Yan et al., 2023)	-	-	72.1	7.0	-	78.7	7.4	-	64.1	9.0
MTDiag (Hou et al., 2023)	-	-	81.9	9.6	-	85.9	9.6	-	-	-
HAIformer (Zhao et al., 2024)	-	-	88.5	1.8	-	**90.4**	1.9	-	**71.1**	3.6
LLM-based Methods										
DP	Qwen2.5-72B-Instruct	59.6	64.4	9.0	59.8	64.4	9.9	44.2	46.8	9.8
	GPT-4o-mini	57.7	61.5	10.0	57.3	65.7	10.0	45.9	49.5	10.0
	Qwen2.5-14B-Instruct	54.8	53.8	10.0	55.2	61.1	10.0	42.2	45.7	10.0
	Qwen2.5-7B-Instruct	59.6	63.5	10.0	54.8	57.3	10.0	46.8	46.2	10.0
UoT (Hu et al., 2024)	Qwen2.5-72B-Instruct	-	67.3	0.1	-	68.6	0.1	-	34.6	0.1
	GPT-4o-mini	-	64.4	0.1	-	65.3	0.4	-	23.0	1.3
	Qwen2.5-14B-Instruct	-	60.6	0.2	-	61.1	0.6	-	32.6	1.3
	Qwen2.5-7B-Instruct	-	61.5	2.2	-	71.1	0.5	-	32.5	1.6
CoD (Chen et al., 2024)	DiagnosisGPT-34B	61.5	53.8	3.9	54.0	44.8	3.6	46.1	34.6	3.7
	DiagnosisGPT-6B	61.5	36.9	5.1	56.1	37.2	3.9	46.6	28.5	3.4
DDO(Ours)	Qwen2.5-14B-Instruct	**66.3**	**94.2**	10.0	67.8	80.3	9.8	**65.3**	68.6	10.0
	Qwen2.5-7B-Instruct	66.3	87.5	9.9	**69.5**	79.5	9.6	60.6	63.6	10.0

Table 2: Overall performance of DDO and baseline methods on three public medical consultation (MC) datasets. Acc_{init} denotes diagnostic accuracy without any symptom inquiry. Bold numbers indicate the best performance, underlined numbers indicate the second-best. All diagnostic accuracy results are reported as percentages.

Method	DXY		GMD		CMD	
	Acc	Avg.n	Acc	Avg.n	Acc	Avg.n
DDO(Ours)	**87.5**	9.9	**79.5**	9.6	**63.6**	10.0
w/o adapter	86.5	9.9	78.7	9.6	54.2	10.0
w/o policy	77.7	10.0	73.6	10.0	60.3	10.0
w/o masking	83.5	9.9	74.5	10.0	61.8	10.0
w/o retry	83.5	10.0	78.7	9.8	63.0	10.0
w/o decision	84.6	9.8	78.2	9.8	63.2	10.0

Table 3: Ablation results of DDO. *w/o* adapter denotes removing the diagnostic adapter. *w/o* policy and *w/o* decision use the LLM and RL model for symptom inquiry, respectively. *w/o* masking disables action space masking. *w/o* retry omits candidate actions regeneration.

Method	DXY		GMD		CMD	
	Acc_{init}	Acc	Acc_{init}	Acc	Acc_{init}	Acc
Numerical	62.5	78.8	53.6	74.5	29.7	39.4
Numerical-SC	65.4	77.9	54.4	74.9	32.8	45.3
FirstLogit	59.6	70.2	60.3	74.1	40.5	43.7
AvgLogit	42.3	75.0	59.4	74.5	25.2	28.8
BTP	63.5	86.5	63.6	78.7	54.2	54.2
BTP-adapter(Ours)	**66.3**	**87.5**	**69.5**	**79.5**	**60.6**	**63.6**

Table 4: Comparison of diagnostic performance across different confidence estimation methods for LLMs.

This modular approach yields substantial gains in diagnostic accuracy.

5.6 Ablation Study

As shown in Table 3, we perform ablation experiments on three MC datasets, using Qwen2.5-7B-Instruct as the LLM backbone.

Impact of Diagnostic Adapter. Removing the diagnostic adapter (*w/o* adapter)—using only the vanilla BTP method to estimate diagnostic confidence—results in a drop in diagnostic accuracy, with the largest decline observed on the CMD dataset. This demonstrates the effectiveness of the in-batch contrastive learning–based adapter in enhancing the disease discrimination ability of LLMs.

Impact of RL-LLM Collaboration. To assess the effectiveness of RL-LLM collaborative symptom inquiry process, we conduct ablation experiments in which symptom inquiry is performed solely by the RL policy (*w/o* decision) or solely by the LLM (*w/o* policy). Results show that both variants perform significantly worse, with the LLM-only variant exhibiting a greater performance drop. This highlights the advantage of multi-agent collaboration in DDO for conducting effective symptom inquiry. Moreover, removing action space masking (*w/o* masking) or disabling the regeneration mechanism for low-quality actions (*w/o* retry) also degrades performance, underscoring their role in ensuring reliable inquiry decisions.

5.7 Evaluation of Diagnostic Performance in LLM Confidence Estimation

To further assess the diagnostic effectiveness of our BTP-adapter, we compare it with several alternative confidence estimation methods for LLMs using three MC datasets. All methods are evaluated with the same initial and final symptom sequences, and the results are presented in Table 4.

Decoding-based Methods. Numerical and Numerical-SC prompt the LLM to directly gen-

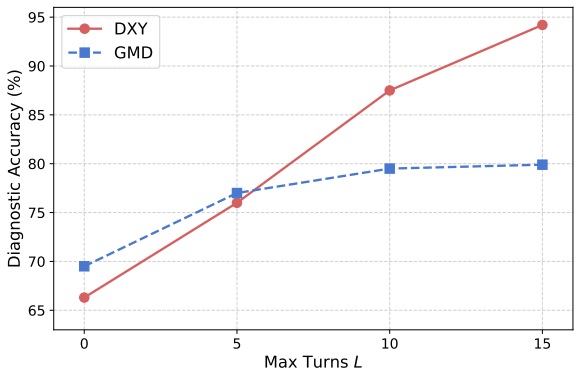

Figure 3: Effect of max turns L.

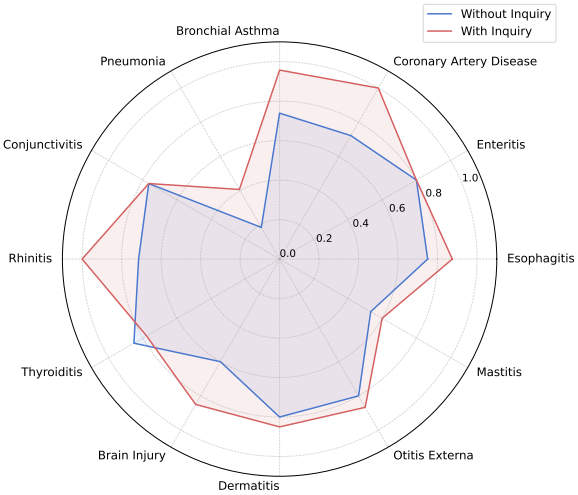

Figure 4: Diagnosis performance at the disease level on the GMD dataset.

erate a confidence score between 0 and 1 (Li et al., 2024b), with SC indicating the use of Self-Consistency (Wang et al., 2022). These approaches show significantly lower diagnostic accuracy than the BTP-adapter, highlighting the limitations of decoding-based confidence estimation and the advantages of our logits-based strategy.

Logits-based Methods. Similar to our approach, FirstLogits and AvgLogits (Ma et al., 2025) estimate confidence based on the logits of the first generated token. However, their diagnostic performance is notably inferior to that of the BTP-adapter. This may be due to their practice of computing confidence scores for all candidate diseases in a single generation process, which can lead to context interference. In contrast, BTP-adapter independently evaluates each candidate diagnosis, effectively mitigating such interference. We also compare against the original BTP method without the diagnostic adapter. Incorporating the adapter consistently enhances diagnostic accuracy, particularly in the initial diagnosis, which is critical for guiding effective symptom inquiry during the early stages of MC.

5.8 Effect of Max Turns L

As shown in Figure 3, we evaluate the diagnostic performance of the DDO framework on the DXY and GMD datasets for different maximum interaction turns ($L = 0/5/10/15$). The line charts show an upward trend, indicating that increasing the maximum number of turns L generally improves diagnostic accuracy. This suggests that the symptom inquiry process in DDO effectively collects critical diagnostic evidence. Notably, the most significant improvements occur in the early stages, with the gains diminishing as more turns are added—especially evident on the GMD dataset. One possible explanation is that DDO prioritizes inquiries for diseases with high initial diagnostic confidence. When the ground-truth disease d_l initially ranks lower, additional inquiries may offer diminishing returns in terms of diagnostic accuracy.

5.9 Diagnostic Effectiveness Across Different Diseases

To assess diagnostic performance of DDO at a fine-grained level, we visualize the diagnostic accuracy before and after symptom inquiry across 12 candidate diseases in the GMD dataset, as shown in Figure 4. The results indicate that multi-turn symptom inquiry substantially enhances diagnostic accuracy for most diseases. Notably, the final accuracy for *Coronary Artery Disease* and *Rhinitis* reaches 100%, highlighting the effectiveness of the collaborative symptom collection between the *Policy Agent* and *Inquiry Agent* in DDO. However, for certain diseases such as *Conjunctivitis* and *Thyroiditis*, the inquiry process yields no improvement in diagnostic accuracy, and even shows a slight decline for *Thyroiditis*. Further analysis reveals that some MCRs for these diseases include only one or two vague or non-specific self-reported symptoms, which causes these diseases to be ranked relatively low initially, making it difficult for the Agents to identify critical symptoms for accurate prediction.

In addition, Table 5 reports both macro-F1 and micro-F1 scores when comparing DDO with the strongest baseline HAIformer. Although DDO does not surpass this baseline, it achieves comparable performance within fewer training epochs. The micro-F1 results demonstrate that DDO preserves overall diagnostic accuracy, while the macro-F1

Method	DXY		GMD		CMD	
	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1
HAIformer	88.8	88.5	**90.1**	**90.4**	**69.3**	**71.1**
DDO(Ours)	**93.6**	**93.3**	81.6	80.3	65.8	67.4

Table 5: Macro-F1 and Micro-F1 scores of DDO and the SOTA baseline on three MC datasets.

results highlight that DDO maintains relatively balanced diagnostic effectiveness across different diseases, avoiding bias toward high-frequency classes.

6 Discussion

Impact of Action Space on RL Policy Learning. In general, the larger the action space of an RL policy model, the more challenging the decision-making process becomes, and the more difficult the optimization tends to be. For example, we observed that with a smaller action space size (42 for DXY), the RL policy was able to sample more informative symptoms, whereas with a larger action space size (359 for CMD), the policy was more prone to sampling irrelevant ones. This motivates our use of masked sampling and LLM-voting during inference: by pruning the action space and leveraging RL–LLM collaboration, we aim to reduce the decision burden placed on the RL policy.

Improving Diagnostic Gains in Challenging Diseases. For certain diseases, diagnostic accuracy exhibits only modest gains—or even slight declines—after symptom inquiry (e.g., Thyroiditis in GMD). As discussed in Section 4, this likely stems from insufficient information in the initial symptom presentation of some MCR cases, which constrains the Agents' ability to identify and query informative symptoms. Possible remedies include: (1) enhancing the initial Top-K accuracy of the Diagnosis Agent for these challenging diseases, since a stronger initial ranking increases the chance of retrieving relevant symptoms and thereby improves inquiry effectiveness; and (2) refining the inquiry strategy by leveraging the LLM's reasoning capacity to filter out irrelevant diseases, thus avoiding excessive focus on unrelated conditions.

Joint Optimization of the Diagnosis Adapter and RL Policy. In this work, we assign different optimization objectives to the diagnosis module and the symptom inquiry module, and adopt a cascaded optimization scheme. This dual-decision optimization effectively enhances the overall performance of LLM-based medical consultation. Nevertheless, LLM-RL cascading training can be inherently unstable. We believe that further improvements could be achieved by jointly optimizing the diagnosis adapter and the RL policy model. Such joint training may not only enable the diagnosis adapter to acquire more robust multi-disease diagnostic capabilities better aligned with interactive doctor–patient scenarios, but also facilitate more stable reward learning for the RL policy.

Mitigating Inference Overhead from Multi-Agent Collaboration. Our DDO framework employs four collaborating agents to accomplish the MC task, substantially enhancing the diagnostic performance of LLM-based methods. However, this multi-agent design also introduces significant inference overhead—for instance, a single multi-turn consultation per MCR typically takes 3–5 minutes. To mitigate this cost, a promising direction inspired by (Chopra and Shah, 2025) is to store completed consultation trajectories in a memory tree. When processing a new MCR, the system can first retrieve similar cases from the tree; if a close match is identified, the consultation can follow the retrieved trajectory, thereby reducing the number of LLM inference calls.

7 Conclusion

In this paper, we propose DDO, a novel LLM-based multi-agent collaborative framework designed to address the mismatch between existing LLM-based methods and the dual-decision nature of medical consultation (MC), which involves both sequential symptom inquiry and diagnosis over a constrained set of candidate diseases. DDO decouples these two decision-making processes and optimizes them with distinct objectives: it improves disease discrimination through a plug-and-play diagnostic adapter, and enhances information gathering via the synergy of an reinforcement learning-based policy agent an LLM-based inquiry agent. Experiments on three public MC datasets show that DDO consistently outperforms existing LLM-based baselines and achieves performance competitive with state-of-the-art generation-based methods, demonstrating its effectiveness in the MC task.

Limitations

While we propose DDO to enhance the effectiveness of LLMs in the medical consultation task, several limitations remain: **(1) Label granularity**: During confidence calibration, DDO assigns a target confidence of 1 to the ground-truth disease and a small constant to all others. This hard-labeling

scheme may hinder the model's ability to softly distinguish between clinically similar diseases. **(2) Model deployment:** Since DDO requires training a diagnostic adapter for the underlying LLM, it is currently incompatible with API-based LLMs and must be deployed with locally hosted models. **(3) Inference efficiency:** DDO involves multi-agent reasoning, which introduces inference latency compared to traditional deep learning methods.

Ethical Consideration

Due to the hallucination problem inherent in large language models, they may generate content that is not factually accurate. Therefore, the DDO framework proposed in this paper is intended solely for academic research. In real-world scenarios, medical decisions should always be based on professional diagnoses made by qualified physicians.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No. 62576363), the Natural Science Foundation of Hunan Province (No. 2025JJ50342) and Changsha Municipal Natural Science Foundation Project (No. kq2502144). This work was carried out in part using computing resources at the High-Performance Computing Center of Central South University.

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*.

David Bani-Harouni, Nassir Navab, and Matthias Keicher. 2024. Magda: Multi-agent guideline-driven diagnostic assistance. In *International workshop on foundation models for general medical AI*, pages 163–172. Springer.

Junying Chen, Chi Gui, Anningzhe Gao, Ke Ji, Xidong Wang, Xiang Wan, and Benyou Wang. 2024. Cod, towards an interpretable medical agent using chain of diagnosis. *arXiv preprint arXiv:2407.13301*.

Junying Chen, Dongfang Li, Qingcai Chen, Wenxiu Zhou, and Xin Liu. 2022. Diaformer: Automatic diagnosis via symptoms sequence generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4432–4440.

Wei Chen, Cheng Zhong, Jiajie Peng, and Zhongyu Wei. 2023. Dxformer: a decoupled automatic diagnostic system based on decoder–encoder transformer with dense symptom representations. *Bioinformatics*, 39(1):btac744.

Harshita Chopra and Chirag Shah. 2025. Feedback-aware monte carlo tree search for efficient information seeking in goal-oriented conversations. *arXiv preprint arXiv:2501.15056*.

Gianluca Detommaso, Martin Bertran, Riccardo Fogliato, and Aaron Roth. 2024. Multicalibration for confidence scoring in llms. In *Proceedings of the 41st International Conference on Machine Learning*, pages 10624–10641.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Zhenyu Hou, Yukuo Cen, Ziding Liu, Dongxue Wu, Baoyan Wang, Xuanhe Li, Lei Hong, and Jie Tang. 2023. Mtdiag: an effective multi-task framework for automatic diagnosis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14241–14248.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

Zhiyuan Hu, Chumin Liu, Xidong Feng, Yilun Zhao, See-Kiong Ng, Anh Tuan Luu, Junxian He, Pang Wei W Koh, and Bryan Hooi. 2024. Uncertainty of thoughts: Uncertainty-aware planning enhances information seeking in llms. *Advances in Neural Information Processing Systems*, 37:24181–24215.

Mingyi Jia, Junwen Duan, Yan Song, and Jianxin Wang. 2025. medikal: Integrating knowledge graphs as assistants of llms for enhanced clinical diagnosis on emrs. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9278–9298.

Shreya Johri, Jaehwan Jeong, Benjamin A Tran, Daniel I Schlessinger, Shannon Wongvibulsin, Zhuo Ran Cai, Roxana Daneshjou, and Pranav Rajpurkar. 2024. Craft-md: A conversational evaluation framework for comprehensive assessment of clinical llms. In *AAAI 2024 Spring Symposium on Clinical Foundation Models*.

Hao-Cheng Kao, Kai-Fu Tang, and Edward Chang. 2018. Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik Siu Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, Hae Park, et al. 2024. Mdagents: An adaptive collaboration of llms for medical decision-making. *Advances in Neural Information Processing Systems*, 37:79410–79452.

Abhishek Kumar, Robert Morabito, Sanzhar Umbet, Jad Kabbara, and Ali Emami. 2024. Confidence under the hood: An investigation into the confidence-probability alignment in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 315–334.

Junkai Li, Yunghwei Lai, Weitao Li, Jingyi Ren, Meng Zhang, Xinhui Kang, Siyu Wang, Peng Li, Ya-Qin Zhang, Weizhi Ma, et al. 2024a. Agent hospital: A simulacrum of hospital with evolvable medical agents. *arXiv preprint arXiv:2405.02957*.

Shuyue Stella Li, Vidhisha Balachandran, Shangbin Feng, Jonathan S Ilgen, Emma Pierson, Pang Wei Koh, and Yulia Tsvetkov. 2024b. Mediq: Question-asking llms and a benchmark for reliable interactive clinical reasoning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Jiaxiang Liu, Yuan Wang, Jiawei Du, Joey Zhou, and Zuozhu Liu. 2024. Medcot: Medical chain of thought via hierarchical expert. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17371–17389.

Wenge Liu, Yi Cheng, Hao Wang, Jianheng Thang, 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 31st International Joint Conference on Artificial Intelligence*, pages 4266–4272.

Mingyu Derek Ma, Yanna Ding, Zijie Huang, Jianxi Gao, Yizhou Sun, and Wei Wang. 2025. Inferring from logits: Exploring best practices for decoding-free generative candidate selection. *arXiv preprint arXiv:2501.17338*.

Mingyu Derek Ma, Xiaoxuan Wang, Yijia Xiao, Anthony Cuturrufo, Vijay S Nori, Eran Halperin, and Wei Wang. 2024. Memorize and rank: Elevating large language models for clinical diagnosis prediction. In *GenAI for Health: Potential, Trust and Policy Compliance*.

Jeremy Qin, Bang Liu, and Quoc Nguyen. 2024. Enhancing healthcare llm trust with atypical presentations recalibration. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 2520–2537.

Shyam Sundhar Ramesh, Yifan Hu, Iason Chaimalas, Viraj Mehta, Pier Giuseppe Sessa, Haitham Bou Ammar, and Ilija Bogunovic. 2024. Group robust preference optimization in reward-free rlhf. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Daniel Rose, Chia-Chien Hung, Marco Lepri, Israa Alqassem, Kiril Gashteovski, and Carolin Lawrence. 2025. Meddxagent: A unified modular agent framework for explainable automatic differential diagnosis. *arXiv preprint arXiv:2502.19175*.

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

Wenqi Shi, Ran Xu, Yuchen Zhuang, Yue Yu, Haotian Sun, Hang Wu, Carl Yang, and May Dongmei Wang. 2024. Medadapter: Efficient test-time adaptation of large language models towards medical reasoning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22294–22314.

Gopendra Singh, Sai Vemulapalli, Mauajama Firdaus, and Asif Ekbal. 2024. Deciphering cognitive distortions in patient-doctor mental health conversations: A multimodal llm-based detection and reasoning framework. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22546–22570.

Donald E Stanley and Daniel G Campos. 2013. The logic of medical diagnosis. *Perspectives in Biology and Medicine*, 56(2):300–315.

Zhoujian Sun, Cheng Luo, and Zhengxing Huang. 2024. Conversational disease diagnosis via external planner-controlled large language models. *arXiv preprint arXiv:2404.04292*.

Kai-Fu Tang, Hao-Cheng Kao, Chun-Nan Chou, and Edward Y Chang. 2016. Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning. In *NIPS workshop on deep reinforcement learning*.

Yuanhe Tian, Ruyi Gan, Yan Song, Jiaxing Zhang, and Yongdong Zhang. 2024. Chimed-gpt: A chinese medical large language model with full training regime and better alignment to human preferences. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7156–7173.

Tao Tu, Anil Palepu, Mike Schaekermann, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Nenad Tomasev, et al. 2024. Towards conversational diagnostic ai. *arXiv preprint arXiv:2401.05654*.

Mina Valizadeh and Natalie Parde. 2022. The ai doctor is in: A survey of task-oriented dialogue systems for healthcare applications. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6638–6660.

Guoxin Wang, Minyu Gao, Shuai Yang, Ya Zhang, Lizhi He, Liang Huang, Hanlin Xiao, Yexuan Zhang, Wanyue Li, Lu Chen, et al. 2025. Citrus: Leveraging expert cognitive pathways in a medical language model for advanced medical decision support. *arXiv preprint arXiv:2502.18274*.

Huimin Wang, Wai-Chung Kwan, Kam-Fai Wong, and Yefeng Zheng. 2023. Coad: Automatic diagnosis

through symptom and disease collaborative generation. *arXiv preprint arXiv:2307.08290*.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

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.

Yuan Xia, Jingbo Zhou, Zhenhui Shi, Chao Lu, and Haifeng Huang. 2020. Generative adversarial regularized mutual information policy gradient framework for automatic diagnosis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1062–1069.

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, and Jingchi Jiang. 2023. Efficient evidence-based dialogue system for medical diagnosis. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3406–3413. IEEE.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, et al. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Xuehan Zhao, Jiaqi Liu, Yao Zhang, Zhiwen Yu, and Bin Guo. 2024. Haiformer: Human-ai collaboration framework for disease diagnosis via doctor-enhanced transformer. In *ECAI 2024*, pages 1495–1502. IOS Press.

Zi'ou Zheng, Christopher Malon, Martin Renqiang Min, and Xiaodan Zhu. 2024. Exploring the role of reasoning structures for constructing proofs in multi-step natural language reasoning with large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15299–15312.

Cheng Zhong, Kangenbei Liao, Wei Chen, Qianlong Liu, Baolin Peng, Xuanjing Huang, Jiajie Peng, and Zhongyu Wei. 2022. Hierarchical reinforcement learning for automatic disease diagnosis. *Bioinformatics*, 38(16):3995–4001.

Shuang Zhou, Zidu Xu, Mian Zhang, Chunpu Xu, Yawen Guo, Zaifu Zhan, Sirui Ding, Jiashuo Wang, Kaishuai Xu, Yi Fang, et al. 2024. Large language models for disease diagnosis: A scoping review. *arXiv preprint arXiv:2409.00097*.

A Other Implementation Details

A.1 Datasets Details

We use three public medical consultation (MC) datasets: DXY[1] (Xu et al., 2019) , which is the most widely used online dialogue–based MC dataset in prior work, and GMD[2] (Liu et al., 2022) together with CMD[3] (Yan et al., 2023) , which are representative of electronic medical record–based MC datasets. These datasets were released for academic research and have been de-identified by their original authors. We reviewed the accompanying papers and code repositories for licensing information: GMD is explicitly provide an MIT license, while DXY and CMD, though lacking formal licenses, clearly state in their papers that the data is intended for public research use. In our study, we strictly follow these terms and use the datasets solely for research purposes. All three datasets are primarily in Chinese, with GMD additionally offering an English version.

For validation, we apply stratified sampling to the training sets of DXY and CMD to create development sets (GMD has already included a predefined split). The final train/dev/test splits for DXY, GMD, and CMD are 318/103/103, 1912/239/239, and 3379/671/1342, respectively.

Since MCRs in DXY and GMD contain relatively few symptoms on average—only 4.74 and 5.55 per record, respectively—this sparsity may hinder the reliable calibration of diagnostic confidence in LLMs. Therefore, when constructing the confidence calibration dataset, we also provide an augmented version of the data to address this issue, enriching MCRs with additional symptoms to enhance the adapter's long-range diagnostic capability. Specifically, for each training MCR with a small number of symptoms, we sample additional implicit symptoms based on disease knowledge extracted from the training data. As a result, the calibration training dataset contains 2,185 instances for DXY (originally 553), 13,598 for GMD (originally 4,837), and 54,608 for CMD.

To improve training efficiency, we do not include all diseases in each contrastive batch when constructing confidence calibration data. Instead, each ground-truth disease is paired with four clinically similar candidate diseases for comparison.

[1] https://github.com/HCPLab-SYSU/Medical_DS
[2] https://github.com/lwgkzl/BR-Agent
[3] https://github.com/YanPioneer/EBAD

A.2 Training Details

Hyperparameters and model checkpoints were selected based on the validation dataset, considering both Top-K accuracy and diagnostic effectiveness across diseases. The key configurations are summarized in Table 6, Table 7 and Table 8.

During confidence calibration, for Qwen2.5-7B-Instruct, several promising checkpoints were retained as candidate adapters for the subsequent RL training, while for other LLMs only the best-performing checkpoint was used for inference.

LLM-RL cascading training can be inherently unstable. To mitigate this, each candidate adapter was employed to train a policy model. The resulting models were then evaluated, and the one achieving the highest Top-K accuracy was selected as the final Policy Agent. This agent was subsequently deployed across different LLM backbones, demonstrating both robustness and transferability.

Hyperparameters	DXY	GMD	CMD
max training epochs	5	1	1
global batch size	8	8	8
learning rate	5e-5	5e-5	5e-5
lora rank	16	16	16
in-bacth group length	5	5	5

Table 6: Hyperparameters for confidence calibration.

Hyperparameters	DXY	GMD	CMD
actor hidden layers	[256,128,128]	[256,128,128]	[512,256,256]
critic hidden layers	[64]	[64]	[128]
learning rate	5e-5	5e-5	5e-5
batch size	64	128	128
steps per update	1024	2048	2048
epochs	5	5	5
total steps	51200	102400	102400
hitting reward	0.5	0.5	0.5
ranking reward	0.5	0.5	0.5
diagnosis reward	1.0	1.0	1.0
frequency penaty	0.2	0.2	0.2
window size	3	5	5
floor turns	3	6	5

Table 7: Hyperparameters for RL policy model's training and inference.

Hyperparameters	DXY	GMD	CMD
sampling times	6	6	7
window size	3	4	5
floor turns	3	5	5
retry times	1	2	2

Table 8: Hyperparameters for LLM–RL collaboration in symptom inquiry.

A.3 Model Deployment

We locally deployed the LLMs used in our DDO framework on GPU devices. Specifically, Qwen2.5-7B-Instruct was run on an NVIDIA RTX 3090 GPU, while Qwen2.5-14B-Instruct was run on an NVIDIA Tesla V100 GPU. For reproducing the Chain-of-Diagnosis baseline (Chen et al., 2024), we deployed DiagnosisGPT-6B and DiagnosisGPT-34B using 1 and 3 NVIDIA Tesla V100 GPUs, respectively. For reproducing EBAD (Yan et al., 2023) and HAIformer (Zhao et al., 2024), we used a single NVIDIA RTX 3090 GPU. For implementing the Direct Prompting baseline and reproducing the Uncertainty of Thoughts baseline (Hu et al., 2024), we utilized the Qwen2.5-Instruct API provided by the Siliconflow platform[4] and the ChatGPT API provided by the ChatAnywhere platform[5].

Based on Stable-Baselines3[6], the reinforcement learning policy model in DDO is trained and deployed on a single NVIDIA RTX 3090 GPU.

B Statistical Results

B.1 Standard Errors and Confidence Intervals

We report the standard errors and confidence intervals of the results for the proposed DDO framework on the three MC datasets in Table 9, based on experiments conducted with different random seeds. For efficiency considerations, we use Qwen2.5-7B-Instruct as the LLM backbone. The comparison between our proposed DDO framework and the baseline methods in the main experiments was performed using the random seed that achieved the best overall performance.

The statistical results in Table 9 indicate that the diagnostic performance of DDO is relatively stable across different random seeds. The standard errors are very small (below 1.0 on all datasets), and the 95% confidence intervals are also narrow (e.g., only 0.6 on GMD), suggesting that repeated runs of DDO yield relatively consistent results.

B.2 Significance Testing

We conducted paired significance tests between the diagnostic performance of our proposed DDO framework and the state-of-the-art baseline HAIformer on the same set of random seeds. On

[4] https://www.siliconflow.cn/
[5] https://chatanywhere.apifox.cn/
[6] https://github.com/DLR-RM/stable-baselines3

Dataset	Mean Accuracy	Standard Error	95% CI (%)
DXY	84.4	0.8	2.3
GMD	79.6	0.1	0.6
CMD	63.1	0.4	1.6

Table 9: Mean accuracy, standard error (SE), and 95% confidence interval (CI) of DDO for final diagnostic accuracy across three MC datasets.

the DXY dataset, DDO demonstrates a statistically significant improvement over HAIformer ($p < 0.05$). However, on the GMD and CMD datasets, DDO achieves lower diagnostic accuracy than HAIformer, and the differences are statistically significant ($p < 0.01$). These results highlight potential challenges in scaling LLM-based medical consultation methods to datasets with more diverse disease categories.

Metric	DXY	GMD	CMD
Paired t-test p-value	0.0355	0.0002	0.0046

Table 10: Statistical significance of DDO compared to HAIformer across three MC datasets.

C Case Study

Table 11 shows a medical consultation case. The patient initially reported the symptom *runny nose*, and DDO used this information to conduct multiple rounds of symptom inquiries to gather more evidence. In each round, DDO first provides a set of candidate inquiry actions via the *Policy Agent*. The *Inquiry Agent* then selects the most appropriate action based on reasoning. The *Patient Agent* responds with the presence or absence of symptoms based on the Medical Consultation Record (MCR). For symptoms not recorded in the MCR, the *Patient Agent* infers their likely presence or absence based on the clinical presentation of the disease. The *Diagnosis Agent* evaluates the diagnostic confidence for each candidate disease. In this case, after reaching the maximum number of interaction rounds, the disease with the highest diagnostic confidence—*allergic rhinitis (AR)*—was identified as the patient's final diagnosis, which aligns with the ground-truth diagnosis in the MCR.

D Prompt Templates

Tables 12-14 show the prompts used in our DDO framework. Tables 15-18 show the prompts used in the DP baseline.

Self Report (Explicit Symptoms): The patient has symptoms of cough, restlessness, and runny nose.
Explicit Symptoms & Implicit Symptoms: {"cough": True, "restlessness": True, "runny nose": True, "anorexia": True, "short of breath": True}
Disease Label: pneumonia
Candidate Diseases: ["hand-foot-and mouth disease (HFMD)", "upper respiratory tract infection (URTI)", "pneumonia", "allergic rhinitis (AR)", "diarrhea"]
Diagnostic Confidence of Top Diseases: {"URTI": 0.68, "pneumonia": 0.16, "AR": 0.16}

Turn 1

Candidate Actions: ["check for fever", "check for vomiting", "check for short of breath", "check for pharyngeal discomfort"]
Selected Symptom for Inquiry: fever
Selection Reasoning: First, we examine the current diagnostic confidence scores: upper respiratory infection (0.68), pneumonia (0.16), and allergic rhinitis (0.16). Since upper respiratory infection has a significantly higher confidence, we prioritize symptoms associated with it. Looking at symptom frequencies for this disease, fever (0.507) is the most common, followed by vomiting (0.233), throat discomfort (0.164), and short of breath (0.123). Based on Strategy 1, we should ask the patient about fever next, as it is the most indicative symptom.
Symptom Status: True
Response Reasoning: To assess whether the patient is likely to have the symptom "fever," we examine its frequency relative to other listed symptoms. A higher frequency—i.e., a higher rank in the sorted list—indicates the symptom is more typical and thus more likely to be present in the patient. According to the information, "fever" has a frequency of 0.533. When we sort all symptoms by frequency in descending order, "fever" ranks 3rd, following "cough" (0.883) and "sputum production" (0.65). This places "fever" in the high-frequency range. Therefore, based on this analysis, we determine the likelihood of the patient having the symptom "fever" to be high, and the final judgment is True.
Diagnostic Confidence of Top Diseases: {"URTI": 0.45, "pneumonia": 0.45, "HFMD": 0.11}

Turn 2

Candidate Actions: ["check for herpes", "check for sneezing", "check for pharyngeal discomfort", "check for short of breath", "check for vomiting"]
Selected Symptom for Inquiry: short of breath
Selection Reasoning: The current diagnostic confidence scores are: upper respiratory infection (0.45), pneumonia (0.45), and hand, foot, and mouth disease (0.11). Since no single disease stands out, we can't apply Strategy 1. So we use Strategy 2, which selects symptoms related to the patient's current presentation—cough, irritability, runny nose, and fever. Among the candidate symptoms (herpes, sneezing, throat discomfort, short of breath, vomiting), short of breath is the most relevant, as it is strongly associated with pneumonia. Therefore, we choose short of breath as the next symptom to inquiry about.
Symptom Status: True
Response Reasoning: Short of breath is recorded in MCR, which indicates that the patient has this symptom.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.68, "URTI": 0.16, "HFMD": 0.16}

Turn 3

Candidate Actions: ["check for pulmonary moist rales", "check for bubble-like spitting", "check for sneeze", "check for vomiting"]
Selected Symptom for Inquiry: pulmonary moist rales
Selection Reasoning: The current diagnostic confidence scores are: pneumonia (0.68), upper respiratory infection (0.16), and hand, foot, and mouth disease (0.16). Since pneumonia stands out with the highest confidence, we prioritize its typical symptoms. Among the candidate symptoms—pulmonary moist rales (0.167), bubble blowing (0.167), sneezing (0.117), and vomiting (0.383)—both pulmonary moist rales and bubble blowing are more specific to pneumonia, despite vomiting having a higher frequency. Following Strategy 1, we choose pulmonary moist rales as the next symptom to ask the patient.
Symptom Status: False
Response Reasoning: To assess whether the patient is likely to have the symptom pulmonary moist rales, we first examine its frequency in pneumonia cases and compare it with other symptoms. The frequency of pulmonary moist rales is 0.167. When we sort all listed symptoms by their frequency in pneumonia, pulmonary moist rales ranks relatively low—below common symptoms like cough (0.883), sputum production (0.65), and fever (0.533), and only slightly higher than symptoms like fatigue, loose stools, and sweating. This suggests that pulmonary moist rales is not among the most typical symptoms of pneumonia. Therefore, we conclude that the patient is less likely to have this symptom, and the judgment is False.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.68, "URTI": 0.16, "diarrhea": 0.16}

Turn 4

Candidate Actions: ["check for anorexia", "check for vomiting", "check for lethargy"]
Selected Symptom for Inquiry: vomiting
Selection Reasoning: The current diagnostic confidence scores are: pneumonia (0.68), upper respiratory infection (0.16), and pediatric diarrhea (0.16). Since pneumonia has the highest confidence, we focus on symptoms related to it. Among the candidate symptoms, vomiting has the highest frequency in pneumonia cases at 0.383, compared to anorexia (0.15) and fatigue (0.083). Based on this, we choose vomiting as the next symptom to ask the patient about.
Symptom Status: True
Response Reasoning: To determine whether the patient is likely to have the symptom "vomiting," we examine its frequency among pneumonia patients, which is 0.383. Although this is lower than symptoms like cough (0.883), sputum production (0.65), fever (0.533), runny nose (0.467), and short of breath (0.433), it is higher than others such as nasal obstruction (0.267) and irritability (0.233). This indicates that vomiting is relatively common in pneumonia cases. While it may not be the most typical symptom, its frequency suggests it has moderate representativeness. Therefore, we conclude that the likelihood of the patient having this symptom is relatively high, and the judgment result is True.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.68, "URTI": 0.16, "HFMD": 0.16}

Turn 5

Candidate Actions: ["check for rash", "check for expectoration", "check for anorexia", "check for herpes", "check for sneeze"]
Selected Symptom for Inquiry: expectoration
Selection Reasoning: Pneumonia currently has the highest diagnostic confidence (0.68), much higher than upper respiratory infection and hand-foot-mouth disease (both 0.16). So, we focus on symptoms typical of pneumonia. Among the candidate symptoms, only expectoration (coughing up phlegm) has a high occurrence in pneumonia (65%), while others like rash, anorexia, sneezing, and herpes are either rare or not typical. Therefore, the most appropriate next symptom to ask about is expectoration.
Symptom Status: True
Response Reasoning: To determine whether the patient is likely to have the symptom expectoration, we refer to the symptom frequency data for pneumonia. The frequency of expectoration is 0.65, which is relatively high—only lower than cough (0.883) and fever (0.533). This indicates that expectoration is a common and typical symptom of pneumonia. Therefore, it's reasonable to conclude that the patient is likely to have this symptom. Final judgment: True.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.68, "URTI": 0.16, "diarrhea": 0.16}

Turn 6

Candidate Actions: ["check for lethargy", "check for sneeze", "check for nasal obstruction"]
Selected Symptom for Inquiry: nasal obstruction
Selection Reasoning: First, we look at the current diagnostic confidence scores: pneumonia (0.68), upper respiratory infection (0.16), and pediatric diarrhea (0.16). Since pneumonia has the highest confidence, we prioritize its related symptoms. Among the candidate symptoms, nasal obstruction (0.267) is more common in pneumonia than listlessness (0.083) and sneezing (0.033). According to strategy 1, we should select the most typical symptom, so we choose "nasal obstruction" as the next symptom to ask the patient about. Final answer: choose nasal obstruction as the next symptom to inquire.
Symptom Status: False
Response Reasoning: To determine whether the patient is likely to have the symptom "nasal obstruction," we examine its frequency in pneumonia cases. The data shows a frequency of 0.267, which ranks 7th among all listed symptoms: lower than common ones like cough (0.883), sputum (0.65), and fever (0.533). Since "nasal obstruction" is not among the most typical symptoms of pneumonia, it is less likely the patient has it. Final judgment: False.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.66, "URTI": 0.18, "diarrhea": 0.16}

Turn 7

Candidate Actions: ["check for anorexia", "check for lethargy", "check for sneeze"]
Selected Symptom for Inquiry: anorexia
Selection Reasoning: The current diagnostic confidence for diseases is: Pneumonia: 0.66, Upper Respiratory Tract Infection: 0.18, Pediatric Diarrhea: 0.16. Since pneumonia has the highest confidence, we will focus on pneumonia-related symptoms. Among the candidate symptoms, anorexia (0.15), lethargy (0.083), and sneezing (0.117) are considered, with anorexia having the highest frequency. Based on Strategy 1, we will select anorexia as the next symptom to ask the patient.
Symptom Status: True
Response Reasoning: Anorexia is recorded in MCR, which indicates that the patient has this symptom.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.45, "URTI": 0.45, "diarrhea": 0.11}

Turn 8

Candidate Actions: ["check for sneeze", "check for pharyngeal discomfort", "check for lethargy"]
Selected Symptom for Inquiry: sneeze
Selection Reasoning: The diagnostic confidence for pneumonia and upper respiratory tract infection is the same (0.45), while pediatric diarrhea is much lower (0.11), so we can't apply Strategy 1. Instead, we use Strategy 2 and focus on symptoms related to the patient's condition. The patient's symptoms include cough, irritability, runny nose, fever, difficulty breathing, vomiting, phlegm, and anorexia. Looking at the frequency of candidate symptoms: Sneezing is more common in upper respiratory tract infection (0.233), while throat discomfort and lethargy appear less frequently across all diseases. Since sneezing is more related to upper respiratory tract infection and the patient's symptoms are mostly respiratory, sneezing helps distinguish between pneumonia and upper respiratory tract infection. Therefore, we should ask about sneezing next.
Symptom Status: False
Response Reasoning: To determine if the patient is likely to have the symptom 'sneezing', we first look at its frequency in pneumonia and compare it to other symptoms. The frequency of sneezing in pneumonia is 0.117. When we rank all listed symptoms by frequency, sneezing is 13th with a relatively low frequency. The symptoms with higher frequencies include cough (0.883), phlegm (0.65), fever (0.533), and others. Given that sneezing ranks low in the list, it is not considered a typical symptom of pneumonia. Therefore, the conclusion is that the patient is unlikely to have the symptom 'sneezing'.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.66, "URTI": 0.18, "diarrhea": 0.16}

Turn 9

Candidate Actions: ["check for pharyngeal discomfort", "check for lethargy"]
Selected Symptom for Inquiry: lethargy
Selection Reasoning: Given the diagnostic confidence for pneumonia (0.66), upper respiratory tract infection (0.18), and pediatric diarrhea (0.16), we prioritize symptoms related to pneumonia since its confidence is much higher. The candidate symptoms for pneumonia are lethargy (0.083) and throat discomfort (0.05). Although both have low frequencies, lethargy is more typical of pneumonia, so we choose it as the next symptom to ask the patient about.
Symptom Status: False
Response Reasoning: Lethargy is not recorded in MCR. This symptom is rarely present in pneumonia, so it is unlikely that the patient has this symptom.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.68, "URTI": 0.16, "HFMD": 0.16}

Turn 10

Candidate Actions: ["check for bubble-like spitting", "check for pulmonary shadow", "check for rash", "check for pharyngeal discomfort"]
Selected Symptom for Inquiry: bubble-like spitting
Selection Reasoning: Given the diagnostic confidence for pneumonia (0.68), upper respiratory tract infection (0.16), and hand-foot-and-mouth disease (0.16), we prioritize pneumonia-related symptoms since its confidence is significantly higher. The candidate symptoms for pneumonia are bubble-like spitting (0.167), lung shadow (0.167), throat discomfort (0.15), and rash (0.033). Both "bubble-like spitting" and "lung shadow" have the highest occurrence rates, but considering that bubble-like spitting may directly reflect lung dysfunction, it is likely more valuable. Therefore, we should choose "bubble-like spitting" as the next symptom to ask the patient about to further confirm pneumonia.
Symptom Status: False
Response Reasoning: To determine if the patient is likely to have the symptom "bubble-like spitting," we first check its frequency in pneumonia symptoms. The frequency is 0.167. When we rank all symptoms by frequency, "bubble-like spitting" ranks 9th. The most common symptoms of pneumonia include cough, sputum, and fever, while "bubble-like spitting" is relatively uncommon. Therefore, it's not a typical pneumonia symptom. Based on this, we conclude that the patient is unlikely to have the symptom "bubble-like spitting." The judgment is False.
Diagnostic Confidence of Top Diseases: {"pneumonia": 0.68, "URTI": 0.16, "diarrhea": 0.16}

Table 11: A medical consultation case by DDO. The symptom status of True and False represent the presence and absence, respectively. "Top Diseases" refers to the diseases with the highest diagnostic confidence, ranked in the top-w positions, with their diagnostic confidence values being processed using Softmax.

You are an experienced medical expert. The following information is provided to you:

Patient's Symptom Presentation
Present symptoms: {positive_symptoms}
Absent symptoms: {negative_symptoms}
Symptom Knowledge for {candidate_disease}
Symptom occurrence frequencies based on existing diagnosed cases of {candidate_disease}: {empirical_knowledge}

Based on the patient's symptom presentation and the symptom knowledge of {candidate_disease}, determine whether this disease can be diagnosed. The presence of relevant symptoms increases diagnostic confidence, while the absence of relevant symptoms decreases it. The influence of a symptom's presence or absence on diagnostic confidence increases with its typicality for the disease. There are two possible outputs: True or False. Output True if you believe the disease can be diagnosed; output False if it cannot.
Please output only the judgment result, without any additional content.

Table 12: The prompt used to estimate diagnostic confidence by BTP in our DDO framework.

You are an experienced medical expert. Your task is to help select the next symptom to inquire about from a given set of candidate symptoms, in order to further collect the patient's symptom information. You are provided with the following information:

Current known patient symptom status (already inquired symptoms):
Present symptoms: {positive_symptoms}
Absent symptoms: {negative_symptoms}
Disease diagnostic confidence (confidence values range from 0 to 1; the higher the value, the more likely the disease):
Clinical presentation knowledge of diseases (symptom occurrence frequency based on historical case statistics): {top_diseases_empirical_knowledge}
Candidate symptoms: {candidate_symptoms}

Based on the above information, choose one suitable symptom from the candidate symptoms to ask next. The symptom selection strategies are as follows:
Strategy 1 (preferred): If the top-ranked disease has significantly higher diagnostic confidence than the others, choose a symptom from the candidate list that is relatively typical for the top-ranked disease, to help confirm its likelihood.
Strategy 2: If no candidate symptom fits Strategy 1, choose a symptom that is relatively related to the patient's current symptom presentation.
Output format:
If there exists a suitable symptom 'xx' in the candidate symptoms, output: Select 'xx' as the next symptom to inquire about.
If no suitable symptom exists in the candidate symptoms, output: New candidate symptoms are needed.
Please think step by step.

Table 13: The prompt used to select an inquiry action in our DDO framework.

You are a patient simulator. The disease that the simulated patient truly has is {disease_label}.

The symptom knowledge of disease {disease_label} is as follows:
Based on statistics from previously diagnosed cases of {disease_label}, the symptom occurrence frequencies are: {empirical_knowledge}

You need to determine whether the patient is likely to have the symptom {inquiried_symptom} based on the symptom knowledge of the disease. The judgment should be either True or False:
True indicates that the patient is likely to have the symptom.
False indicates that the patient is unlikely to have the symptom.

The criteria for judgment are as follows:
If the symptom {inquiried_symptom} is relatively typical for the disease {disease_label} (i.e., it ranks high in the symptom occurrence frequency), then it is considered likely that the patient has the symptom (judgment should be True).
If the symptom is not typical for the disease, then the patient is considered unlikely to have it (judgment should be False).
Please think step by step and decide whether the patient is likely to have the symptom {inquiried_symptom}.

Table 14: The prompt used to simulating the patient's response in our DDO framework.

You are an experienced medical expert conducting a consultation with a patient.

After several rounds of symptom inquiries, the patient has confirmed the following symptoms: {positive_symptoms}. The patient has denied the following symptoms: {negative_symptoms}. The diseases you suspect are: {candidate_diseases}.

You need to decide whether to continue asking about symptoms to gather diagnostic evidence or to provide a final diagnosis based on the symptoms reported and your diagnostic knowledge of the diseases. The decision should be one of the following two options: (1) Ask about symptoms (2) Diagnose disease. Please provide your decision directly, without any additional explanation.
Decision:

Table 15: The prompt for deciding interaction action in the DP baseline.

You are an experienced medical expert conducting a consultation with a patient.

The symptoms that have already been inquired about and their status are as follows: The symptoms confirmed by the patient: {positive_symptoms}. The symptoms denied by the patient: {negative_symptoms}.

To increase diagnostic confidence, you need to choose a symptom to inquire about, ensuring that it has not been previously inquired about. Please provide the name of the symptom directly, without any additional content.
The symptom to inquire about:

Table 16: The prompt for symptom inquiry in the DP baseline.

You will play the role of a patient diagnosed with {disease}.

Your symptom presentation is as follows: The symptoms you have: {positive_symptoms}. The symptoms you do not have: {negative_symptoms}.

Based on your symptom presentation, please answer truthfully whether you have the symptom {symptom}. The answer should be either 'Yes' or 'No'. Please provide the answer directly without any additional content.
Answer:

Table 17: The prompt for simulating the patient's response in the DP baseline.

You are an experienced medical expert, currently conducting a consultation with a patient.

After several rounds of symptom inquiries: The symptoms confirmed by the patient are: {positive_symptoms}. The symptoms denied by the patient are: {negative_symptoms}. The diseases you suspect include: {candidate_diseases}.

Please select three diseases from the suspected list above as the diagnosis results, ordered from most to least likely. Provide the answer in the form of a Python string list, and do not include any additional content.
Top three most likely diseases:

Table 18: The prompt for disease diagnosis in the DP baseline.