MCPlanner: Hierarchical Adaptive Planning-Based Medical Consultation with Large Language Models

Anonymous ACL submission

Abstract

Medical consultation (MC) is a partially observable interactive decision-making process that requires dynamically adapting inquiry directions according to evolving patient feedback. Although large language models (LLMs) have demonstrated strong planning capabilities in medical applications, existing LLM-based MC methods typically make turn-level independent inquiry decisions without adaptive control over global inquiry directions, making them prone to inefficient local exploration. To address this limitation, we propose a hierarchical adaptive planning-based MC framework—MCPlanner, where a high-level planner that adaptively controls inquiry direction and a low-level planner executes inquiry actions under the given direction. Additionally, MCPlanner incorporates expert-guided distillation and a smoothed confidence calibration strategy to improve high-level planning efficiency and diagnostic accuracy. Experimental results on three public datasets show that MCPlanner outperforms the baselines in overall performance, demonstrating its effectiveness for the MC task.

1 Introduction

Recent years have witnessed growing interest in the medical consultation (MC) task in AI for healthcare (Hou et al., 2023; Tu et al., 2024; Fan et al., 2025; Chen et al., 2026). In this task, an AI doctor interacts with patients over multiple turns to collect critical symptom information and make accurate diagnoses. Compared with traditional deep learning approaches (Zhong et al., 2022; Chen et al., 2022b,a; Yan et al., 2023; Wang et al., 2023; Zhao et al., 2024), large language models (LLMs) exhibit stronger medical knowledge and more interpretable decision-making, making them increasingly adopted for MC (Chen et al., 2025a; Feng et al., 2025; Jia et al., 2025; Dou et al., 2026).

As illustrated in Figure 1-(a), existing LLM-based MC methods typically adopt turn-wise independent strategies to query informative symptoms based on symptom-level associations (Chen et al., 2025a; Wang et al., 2025; Jia et al., 2025; Zhu et al., 2025). However, without explicit control over inquiry direction, LLMs are prone to inefficient local exploration. For example, early-stage non-specific symptoms such as fever and cough may correspond to multiple diseases. Relying solely on symptom-level associations makes it difficult to elicit symptoms that are truly informative for the target disease, thereby falling into suboptimal inquiry paths. To alleviate this issue, UoT (Hu et al., 2024) and MISQ-HF (Chopra and Shah, 2025) treat remaining candidate branches in the search tree as potential inquiry directions. However, these methods update the candidate disease set in a rigid pruning manner after each patient response, which may prematurely eliminate the true disease.

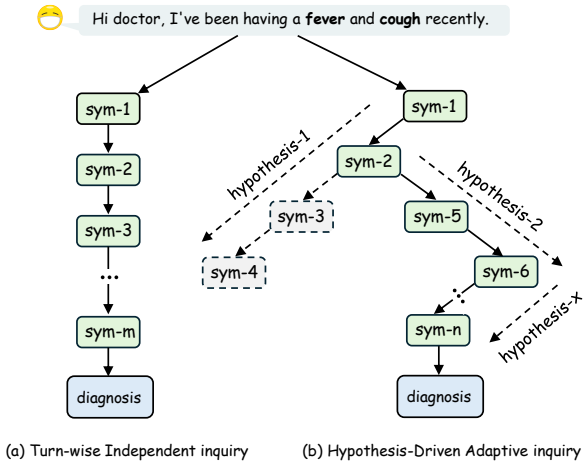

Figure 1: Examples of turn-wise independent Inquiry and hypothesis-driven adaptive inquiry.

In real-world consultations, physicians typically follow an iterative hypothesis–deduction–verification process (Stanley and Campos, 2013). As shown in Figure 1(b), physicians first formulate disease hypotheses from the patient's initial report and then conduct targeted questioning around the

current hypothesis, namely hypothesis-driven adaptive inquiry. If newly acquired evidence supports the hypothesis, the verification proceeds; otherwise, the physician shifts to an alternative hypothesis.

Inspired by this clinical paradigm, we propose MCPlanner, an adaptive hierarchical planning framework for MC. At each turn, a high-level planner dynamically controls consultation flow and inquiry direction, adjusting the direction whenever collected evidence conflicts with the current hypothesis or when remaining symptoms offer limited information gain. Under the selected direction, a low-level planner generates concrete symptom inquiry sequences to collect diagnostic evidence.

To improve high-level planning, we introduce an expert-guided distillation mechanism that enables LLMs to learn adaptive planning strategies from stronger teacher models. For diagnosis, MCPlanner employs KL-divergence-based smoothed in-batch contrastive learning to calibrate diagnostic confidence, improving discrimination among similar diseases while preserving the diagnostic advantage of the ground-truth disease. The main contributions are summarized as follows:

(1) We propose MCPlanner, which simulates the clinical hypothesis–deduction–verification process through hierarchical agent collaboration, improving the adaptivity of LLM-based MC.

(2) We introduce expert-guided distillation for high-level adaptive planning and a KL-divergence-based confidence calibration strategy to enhance disease discrimination.

(3) Experiments on three real-world MC datasets show that MCPlanner consistently outperforms the baselines, achieving accuracies of 95.2%, 90.8%, and 76.1% on DXY, GMD, and CMD, respectively.

2 Task Definition

In the MC task, a virtual doctor starts from the patient's initially reported symptom information, denoted as $\mathcal{S}^{\text{init}} = \{(s_i^{\text{init}}, p_i^{\text{init}})\}_{i=1}^{l_1}$, and progressively acquires additional symptom information through active inquiry, denoted as $\mathcal{S}^{\text{ad}} = \{(s_j^{\text{ad}}, p_j^{\text{ad}})\}_{j=1}^{l_2}$, thereby continuously accumulating diagnostic evidence. Here, l_1 and l_2 denote the numbers of initially reported and additionally collected symptoms, respectively. For both $\mathcal{S}^{\text{init}}$ and $\mathcal{S}^{\text{ad}}$, s denotes a symptom and p denotes its presence status (present or absent).

The consultation process terminates when the virtual doctor is able to determine the diagnosis or when the number of consultation turns reaches a predefined upper bound L. The final diagnosis result d_p is defined as the disease with the highest diagnostic confidence within the candidate disease set $D = \{d_1, d_2, \ldots, d_n\}$:

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

3 Method

As illustrated in Figure 2, MCPlanner begins with PAgent reporting the patient's initial symptoms. DAgent then estimates confidence scores for all candidate diseases and initializes the disease hypothesis set by retaining the top-w candidates. The initial symptoms are simultaneously used to initialize the consultation trajectory.

At each turn, HPlanner assesses the current inquiry direction based on the consultation state, including the consultation trajectory and symptom inquiry plan. If sufficient evidence has been collected, the consultation terminates; otherwise, HPlanner either maintains the current direction or switches to a new one. Under the selected direction, LPlanner generates symptom inquiry actions and interacts with PAgent to collect diagnostic evidence.

When HPlanner terminates the consultation or the maximum turn limit is reached, DAgent re-estimates candidate disease confidence and outputs the most confident disease as the final diagnosis.

To enhance adaptive planning, we introduce an expert-guided distillation mechanism. In addition, we propose a smoothed in-batch contrastive learning strategy for diagnostic confidence calibration, improving LLMs' diagnostic ability.

3.1 High-Level Planning

In MCPlanner, HPlanner dynamically assesses the rationality of the current inquiry direction at each turn and decides whether to switch it or terminate the consultation or. At turn t, HPlanner receives two types of consultation state information:

(1) Consultation trajectory τ_t. The consultation trajectory records historical interactions and helps HPlanner assess whether the current inquiry deviates from the correct direction. It is defined as:

$$\tau_t = (o_0, (\tilde{z}_1, a_1, o_1), \ldots, (\tilde{z}_{t-1}, a_{t-1}, o_{t-1})), \quad (2)$$

where o_0 denotes the initial symptom information; a_i denotes the inquiry action at turn i; and o_i denotes the PAgent's feedback. $\tilde{z}_i = (z_i, d_i)$ denotes

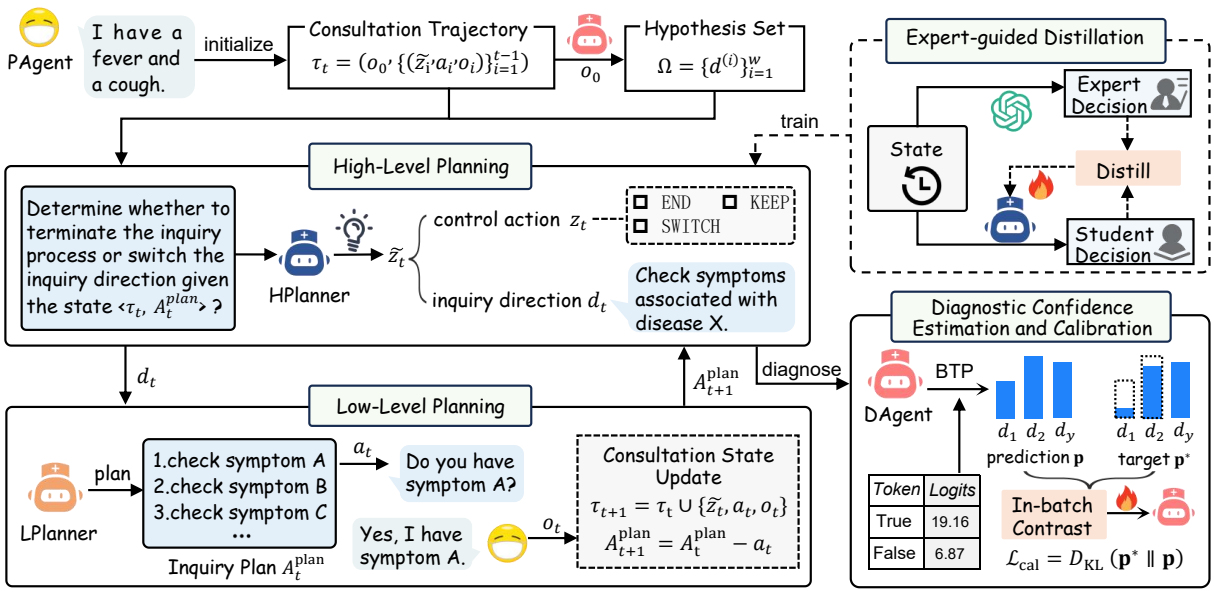

Figure 2: An adaptive hierarchical planning framework (MCPlanner) for medical consultation.

the high-level decision at turn i, $d_i \in \mathcal{D}$ is the inquiry direction and $z_i \in \{\text{END}, \text{KEEP}, \text{SWITCH}\}$ is the flow control action:

- END: sufficient evidence supports d_i and excludes other candidate diseases;
- KEEP: current evidence supports d_i, while remaining symptoms are still informative;
- SWITCH: evidence is inconsistent with d_i, or remaining symptoms are less informative.

(2) Symptom inquiry plan $\mathcal{A}_t^{\text{plan}}$. When evaluating the value of the current inquiry direction, HPlanner considers not only its consistency with the collected evidence but also the information gain of the remaining planned inquiries. This plan consists of the unexecuted symptom inquiry actions under the current direction:

$$\mathcal{A}_t^{\text{plan}} = \left(a_t^{(1)}, a_t^{(2)}, \ldots, a_t^{(k_t)}\right), \tag{3}$$

where $a_t^{(j)}$ denotes the j-th inquiry action at turn t, and k_t denotes the number of unexecuted actions.

Based on these two inputs, HPlanner outputs its decision according to policy π^H:

$$(z_t, d_t) \sim \pi^H(\cdot \mid \tau_t, \mathcal{A}_t^{\text{plan}}), \tag{4}$$

where z_t denotes the flow control action and d_t denotes the inquiry direction at turn t. If $z_t =$ SWITCH, a new candidate disease is selected from the disease hypothesis set as d_t, based on its consistency with the collected symptom evidence and the informativeness of its remaining symptoms.

3.2 Low-Level Planning

When the inquiry direction changes, LPlanner is responsible for generating a symptom inquiry action sequence under the current direction d_t. In addition, LPlanner interacts with PAgent to progressively collect diagnostic evidence.

3.2.1 Symptom Inquiry Plan Generating

To generate a symptom inquiry plan, LPlanner takes the current inquiry direction d_t and the corresponding clinical knowledge $\mathcal{K}$ as input, where $\mathcal{K}$ is constructed from symptom occurrence statistics of candidate diseases in the training set.

Based on these inputs, LPlanner generates the symptom inquiry plan:

$$\mathcal{A}_t^{\text{plan}} = \left[a_t^{(1)}, a_t^{(2)}, \ldots, a_t^{(K)}\right] \sim \pi_{\text{plan}}^L(\cdot \mid d_t, \mathcal{K}), \tag{5}$$

where $a_t^{(k)}$ denotes the k-th symptom inquiry action in the plan, and π_{plan}^L denotes the low-level planning policy. Specifically, LPlanner prioritizes symptoms that are more typical of the disease corresponding to the current inquiry direction. When multiple symptoms exhibit similar typicality, those with higher discriminative value are prioritized to better distinguish similar diseases.

3.2.2 Inquiry Action Execution

LPlanner sequentially executes the inquiry actions in $\mathcal{A}_t^{\text{plan}}$. Let a_t denote the action executed at turn t, and let s_t denote the corresponding queried symptom. After executing this action, the symptom in-

quiry plan is updated as:

$$\mathcal{A}_{t+1}^{\text{plan}} = \mathcal{A}_t^{\text{plan}} - a_t, \tag{6}$$

where "$-$" denotes the element removal operation.

PAgent simulates the patient response based on the original medical consultation record $\mathcal{P}$. Specifically, PAgent first checks whether the presence status of symptom s_t is explicitly recorded in $\mathcal{P}$. If available, the recorded status is directly returned. Otherwise, PAgent infers its presence status according to the disease d_l suffered by the patient: if s_t rarely occurs in disease d_l, it is inferred as absent; otherwise, it is inferred as present.

After receiving patient feedback o_t, the consultation trajectory is updated as:

$$\tau_{t+1} = \tau_t \cup \{(\tilde{z}_t, a_t, o_t)\}. \tag{7}$$

3.3 Expert-guided Distillation

Since lightweight LLMs (e.g., 7B) often exhibit unreliable flow control, we introduce expert-guided distillation to enhance their adaptive planning.

3.3.1 Expert Demonstration Generation

Expert demonstrations are constructed by allowing a teacher model to execute the MCPlanner consultation process on the training set. For the i-th sample $\mathcal{P}_i^{\text{train}}$, at turn t, the teacher model produces a consultation trajectory τ_t^T and the remaining symptom inquiry plan $\mathcal{A}_t^T$. It then outputs a high-level planning decision according to policy π_T^H:

$$(z_t^T, d_t^T) \sim \pi_T^H(\cdot \mid \tau_t^T, \mathcal{A}_t^T), \tag{8}$$

where z_t^T and d_t^T denote the flow control action and inquiry direction, respectively.

This yields a set of expert demonstrations:

$$\xi_i^T = \{(\tau_{i,t}^T, \mathcal{A}_{i,t}^T, z_{i,t}^T, d_{i,t}^T)\}_t^{l_i}, \tag{9}$$

where l_i is the total number of consultation turns.

To improve supervision quality, we filter out suboptimal demonstrations and retain only those where the teacher model either produces the correct final diagnosis or improves the ranking of the ground-truth disease. The remaining samples form the final expert demonstration dataset $\mathcal{D}_{\text{demo}}$.

3.3.2 Distillation Learning

We fine-tune the student model using the expert demonstrations to optimize its high-level planning policy π_S^H. Specifically, the distillation is performed by maximizing the conditional likelihood of reproducing teacher decisions under the same state. The distillation objective is defined as:

$$\begin{aligned}\mathcal{L}_{\text{distill}} = \mathbb{E}_{(\tau^T, \mathcal{A}^T, z^T, d^T) \sim \mathcal{D}_{\text{demo}}} \\ \left[-\log \pi_S^H(z^T, d^T \mid \tau^T, \mathcal{A}^T)\right].\end{aligned} \tag{10}$$

3.4 Diagnostic Confidence Estimation and Calibration

3.4.1 Diagnostic Confidence Estimation

In MCPlanner, DAgent estimates the diagnostic confidence of each candidate disease using the logits-based BTP method (Jia et al., 2025):

For each candidate disease $d_i \in D$, DAgent outputs a binary decision token (`True`/`False`) conditioned on the observed symptoms, indicating whether d_i is considered a plausible diagnosis or not. The logits corresponding to `True` and `False`, denoted as logit_T and logit_F, are extracted to compute the diagnostic confidence of d_i:

$$conf_i = \frac{\exp(\text{logit}_\text{T})}{\exp(\text{logit}_\text{T}) + \exp(\text{logit}_\text{F})}. \tag{11}$$

3.4.2 Diagnostic Confidence Calibration

To improve domain adaptation, DDO (Jia et al., 2025) uses KL-divergence-based in-batch contrastive learning for confidence calibration, constructing an approximately one-hot target distribution to maximize the confidence gap between the ground-truth disease and other candidates. However, since candidate diseases often share overlapping symptoms, such sharp targets may increase optimization difficulty and destabilize confidence ranking. To address this issue, inspired by the smoothed token prediction strategy of Li et al. (Li et al., 2025b), we propose a smoothed confidence calibration method.

Specifically, DAgent first estimates confidence scores for all candidate diseases to obtain the predicted distribution $\mathbf{p}$. We then identify the strongest competing disease d_c, i.e., the non-ground-truth disease with the highest predicted probability. Let p_l and p_c denote the predicted probabilities of the ground-truth disease d_l and d_c, respectively. We enforce a margin constraint to keep the ground-truth probability slightly higher:

$$p_l^* = \min\Big(\max(p_l,\, p_c + margin),\, 1-\epsilon\Big), \tag{12}$$

where ϵ is a small constant. The remaining probabilities are proportionally rescaled to obtain the smoothed target distribution $\mathbf{p}^*$.

Finally, confidence calibration is achieved by minimizing the KL divergence between the predicted distribution and the target distribution:

$$\mathcal{L}_{\text{cal}} = D_{\text{KL}}(\mathbf{p}^* \parallel \mathbf{p}) \,. \tag{13}$$

4 Experiments

4.1 Baselines

4.1.1 Traditional Deep Learning-based Methods

For traditional deep learning-based MC methods, we select three representative baselines: **EBAD** (Yan et al., 2023), **MTDiag** (Hou et al., 2023), and **HAIformer** (Zhao et al., 2024). EBAD reduces decision randomness in reinforcement learning-based consultation through a knowledge graph-based pruning strategy. MTDiag formulates MC as a multi-task learning problem by separately optimizing symptom inquiry and disease diagnosis. HAIformer improves consultation decision quality through a human-AI collaborative mechanism.

4.1.2 LLM-Based Methods

For LLM-based MC methods, we compare MCPlanner with four baselines. **UoT** (Hu et al., 2024) plans subsequent inquiries by estimating information gain. **MISQ-HF** (Chopra and Shah, 2025) extends UoT by introducing a reward mechanism that learns from successful experiences. **DDO** (Jia et al., 2025) designs differentiated optimization objectives for symptom inquiry and disease diagnosis. **Direct Prompting (DP)** directly leverages the internal knowledge of LLMs for consultation reasoning, without additional prompt engineering or parameter adaptation.

4.2 Datasets

We evaluate the proposed MCPlanner framework 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.

Dataset	DXY	GMD	CMD
# Total 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. Avg. Symptoms denotes average recorded symptoms per sample.

4.3 Evaluation Metrics

(1) Diagnostic Accuracy. We adopt diagnostic accuracy (Acc) as the primary evaluation metric to measure the model's ability to correctly identify the ground-truth disease from the candidate disease set. Given N samples, where d_i^{l} and d_i^{p} denote the ground-truth and predicted disease of the i-th sample, respectively, Acc is defined as:

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}\left(d_i^{\text{p}} = d_i^{\text{l}}\right), \tag{14}$$

where $\mathbf{I}(\cdot)$ is the indicator function, which equals 1 if the prediction matches the ground-truth disease and 0 otherwise.

(2) Average Consultation Turns. The average number of consultation turns (Avg.n) reflects both the amount of symptom information collected and the efficiency of flow control. Given N samples, where T_i denotes the number of symptom inquiry turns for the i-th sample, Avg.n is defined as:

$$\text{Avg.n} = \frac{1}{N} \sum_{i=1}^{N} T_i. \tag{15}$$

4.4 Implementation Details

We re-implemented all baselines based on their publicly available code. Since MTDiag does not provide official code, we directly report the results from its original paper. For UoT, the pruning ratio is set to -0.5 to balance the computational cost of large-scale simulated interactions. For MISQ-HF, the maximum number of MCTC explorations is set to 30, and the reward increment ratio during feedback is set to 0.5. In both UoT and MISQ-HF, consultation termination is triggered when the doctor agent first explicitly asks whether the patient has a specific disease. The maximum number of symptom inquiry turns is uniformly set to 10. The LLM backbones used in our experiments include DeepSeek-V3 (Liu et al., 2025), Qwen3-8B, Qwen2.5-14B-Instruct, and Qwen2.5-7B-Instruct (Yang et al., 2025). More implementation details are provided in Appendix B.

4.5 Main Results

Table 2 presents the comparative results of MCPlanner and baseline methods on three real-world MC datasets. In terms of final diagnostic accuracy, MCPlanner achieves the best performance across all

Methods	LLM backbones	DXY			GMD			CMD		
		Acc_{init}	Acc	Avg.n	Acc_{init}	Acc	Avg.n	Acc_{init}	Acc	Avg.n
Traditional Deep Learning-based Methods										
EBAD (Yan et al., 2023)	-	69.2	72.1	7.0	70.7	78.7	7.4	64.0	66.0	9.1
MTDiag (Hou et al., 2023)	-	-	81.9	9.6	-	85.9	9.6	-	-	-
HAIformer (Zhao et al., 2024)	-	70.2	88.5	1.8	72.8	90.4	1.9	66.4	71.1	3.6
LLM-based Methods										
DP	DeepSeek-V3	62.1	67.0	9.8	64.4	71.1	9.9	49.0	47.6	9.6
	Qwen3-8B	32.0	39.8	3.5	59.0	61.5	8.5	30.8	36.9	9.3
	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)	DeepSeek-V3	65.4	68.3	0.5	76.6	70.0	0.1	37.3	37.0	0.1
	Qwen3-8B	49.0	67.3	0.7	66.9	72.8	0.8	32.5	34.0	2.3
	Qwen2.5-14B-Instruct	55.8	60.6	0.2	72.0	61.1	0.6	31.4	32.6	1.3
	Qwen2.5-7B-Instruct	54.8	61.5	2.2	63.2	71.1	0.5	30.8	32.5	1.6
MISQ-HF (Chopra and Shah, 2025)	DeepSeek-V3	76.9	73.1	0.5	**76.6**	77.8	0.9	37.3	31.0	2.1
	Qwen3-8B	70.2	58.7	0.9	67.8	64.0	1.2	32.7	31.2	1.8
	Qwen2.5-14B-Instruct	69.2	69.2	0.6	72.0	75.7	1.5	31.6	43.8	3.0
	Qwen2.5-7B-Instruct	66.3	66.3	0.8	65.3	75.3	2.0	31.0	31.4	3.4
DDO (Jia et al., 2025)	Qwen3-8B	64.4	90.4	10.0	71.1	82.8	9.1	64.2	71.1	10.0
	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
MCPlanner(Ours)	Qwen3-8B(Distill)	**76.9**	**95.2**	8.1	73.2	88.7	6.8	62.6	69.7	9.7
	Qwen3-8B	76.9	91.3	9.9	73.2	88.7	9.3	62.6	65.5	8.8
	Qwen2.5-14B-Instruct(Distill)	71.2	93.3	8.3	69.9	87.0	6.4	**69.6**	**76.1**	9.8
	Qwen2.5-14B-Instruct	71.2	92.3	10.0	69.9	85.8	8.5	69.6	76.0	9.9
	Qwen2.5-7B-Instruct(Distill)	68.3	88.5	8.4	70.3	87.9	7.3	61.3	68.4	9.2
	Qwen2.5-7B-Instruct(DeepSeek-V3)	68.3	85.6	10.0	70.3	**90.8**	9.8	61.3	69.1	9.8
	Qwen2.5-7B-Instruct(GPT-5-mini)	68.3	87.5	8.0	70.3	90.0	6.7	61.3	68.2	8.8
	Qwen2.5-7B-Instruct	68.3	80.8	10.0	70.3	82.4	7.1	61.3	68.0	9.3

Table 2: Main experimental results of MCPlanner and baseline methods on three MC datasets. Acc_{init} denotes the initial diagnostic accuracy. Bold indicates the best diagnostic accuracy, with all accuracy values reported as percentages. For MCPlanner, the model name is followed by the backbone model of HPlanner in parentheses. Distill denotes Qwen2.5-7B-Instruct after expert-guided distillation.

three datasets, reaching 95.2%, 90.8%, and 76.1% on DXY, GMD, and CMD, respectively. Compared with the initial diagnostic accuracy, multi-turn symptom inquiry consistently improves diagnostic performance. Taking Qwen3-8B as an example, the accuracy gains on DXY, GMD, and CMD are 16.4%, 15.5%, and 5.0%, respectively, demonstrating that the hierarchical adaptive planning mechanism enables LLMs to collect informative diagnostic evidence.

Compared with traditional baselines, MCPlanner consistently achieves higher final diagnostic accuracy across all three datasets. Moreover, the improvement from initial to final diagnostic accuracy is generally larger than that of traditional methods. This indicates that LLMs can achieve superior consultation performance by simulating the adaptive clinical "hypothesis–deduction–verification" consultation process.

Compared with existing LLM-based MC methods, MCPlanner delivers superior consultation performance, validating the effectiveness of the Hypothesis-Driven Adaptive Inquiry strategy for improving LLM-based consultation. DP and DDO tend to continue consultation until reaching the maximum turn limit, whereas MCPlanner effectively reduces the average number of consultation turns while maintaining strong diagnostic performance. This is because HPlanner can accurately assess the sufficiency of diagnostic evidence under the current inquiry direction based on changes in the consultation state, enabling timely termination once sufficient evidence has been collected and thus avoiding unnecessary inquiries.

UoT and MISQ-HF progressively prune candidate diseases based on patient feedback and use the remaining candidates as guidance for symptom inquiry. However, they lack adaptive error correction when the inquiry direction substantially deviates from the correct diagnosis, making them prone to inefficient exploration along suboptimal directions and ultimately degrading consultation performance. By jointly considering historical consultation interactions and future symptom inquiry plans, MCPlanner adaptively determines whether the inquiry direction should be adjusted, avoiding inefficient symptom inquiries caused by overly aggressive or overly conservative direction switching.

4.6 Ablation Study

Table 3 reports the ablation results of MCPlanner with Qwen2.5-7B-Instruct as the backbone.

Methods	DXY		GMD		CMD	
	Acc	Avg.n	Acc	Avg.n	Acc	Avg.n
MCPlanner(Ours)	**88.5**	8.4	**87.9**	7.3	**68.4**	9.2
w/o control	70.2	5.6	74.5	5.9	62.3	5.5
w/o plan	75.0	9.9	71.5	10.0	61.5	10.0
w/o distill	80.8	10.0	82.4	7.1	68.0	9.3
w/o calibration	81.7	8.4	82.8	7.3	55.7	9.2

Table 3: Ablation results of MCPlanner.

(1) Removing High-Level Adaptive Planning (*w/o* control). Removing HPlanner reduces diagnostic accuracy to 70.2%, 74.5%, and 62.3% on DXY, GMD, and CMD, respectively, while significantly shortening consultation length. This suggests that without high-level guidance, the model struggles to generate informative inquiry plans from initial symptoms alone.

(2) Removing Hierarchical Planning (*w/o* plan). When hierarchical planning is removed and the LLM directly performs inquiry decisions, diagnostic accuracy drops substantially and consultation turns approach the maximum limit. This indicates that explicit planning is essential for efficient symptom exploration in complex consultation scenarios.

(3) Removing Expert-guided Distillation (*w/o* distill). Removing expert-guided distillation degrades performance on all datasets. In particular, the average consultation turns on DXY increase from 8.4 to 10.0, showing that distillation improves the adaptivity of HPlanner in lightweight LLMs.

(4) Removing Confidence Calibration (*w/o* calibration). Removing confidence calibration lowers diagnostic accuracy across all datasets, with the largest drop on CMD. This suggests that calibration improves discrimination between similar diseases by producing more reliable confidence estimates.

4.7 Analysis of High-Level Planning with Different LLM Backbones

HPlanner directly affects both the efficiency of symptom acquisition and the final diagnostic performance of MCPlanner. We therefore evaluate different backbone models as HPlanner on the DXY dataset, using Qwen2.5-7B-Instruct with diagnostic confidence calibration as DAgent.

Figure 3 shows the distribution of consultation termination reasons. Most cases terminate after reaching the maximum turn limit, suggesting that MCPlanner often requires iterative "hypothesis–deduction–verification" interactions to collect sufficient evidence. In some cases, the initial self-reported symptoms already provide enough diagnostic evidence, allowing early termination. For example, GPT-5-mini terminates early in 34.3% of cases. In contrast, DeepSeek-V3 and the original Qwen2.5-7B-Instruct tend to continue inquiry. After distillation, Qwen2.5-7B-Instruct achieves a higher early termination ratio while maintaining strong diagnostic accuracy, indicating that lightweight models can learn more reliable termination strategies from expert demonstrations.

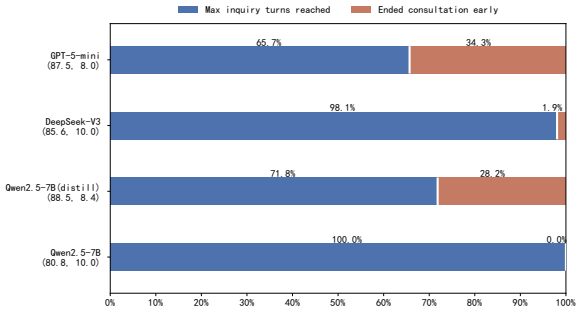

Figure 3: Distribution of consultation termination reasons for different LLM backbones of HPlanner. The values in parentheses indicate diagnostic accuracy and average consultation turns.

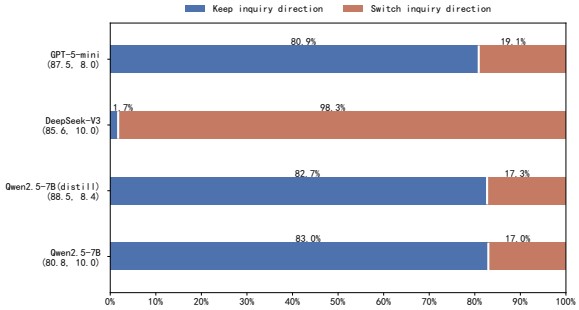

Figure 4: Distribution of direction control actions for different LLM backbones of HPlanner. The values in parentheses indicate diagnostic accuracy and average consultation turns.

Figure 4 presents the distribution of direction control actions for non-terminated consultations. DeepSeek-V3 switches inquiry directions much more frequently than GPT-5-mini, with a switching ratio of 98.3%. However, it achieves lower diagnostic accuracy and requires more consultation turns, suggesting that frequent switching does not necessarily improve consultation performance. Although the original and distilled Qwen2.5-7B-Instruct show similar action distributions, the distilled model achieves substantially better diagnostic performance, demonstrating that expert-guided distillation helps lightweight models learn more

Methods	DXY		GMD		CMD	
	Acc_{init}	Acc	Acc_{init}	Acc	Acc_{init}	Acc
Numerical	67.3	79.8	56.5	82.4	41.9	67.6
Numerical-SC	57.7	75.0	52.7	75.7	37.7	56.5
FirstLogit	59.6	80.8	61.5	81.2	56.8	63.2
AvgLogit	43.3	61.5	59.8	80.3	25.2	30.5
BTP	63.5	81.7	63.6	82.8	54.2	55.7
BTP + one-hot calibration	66.3	85.6	69.5	82.8	60.4	62.4
BTP + soft calibration(Ours)	**68.3**	**88.5**	**70.3**	**87.9**	**61.3**	**68.4**

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

appropriate timing for inquiry direction switching.

4.8 Analysis of Diagnostic Performance under Different Confidence Estimation Methods

Table 4 presents the diagnostic results of different confidence estimation methods under identical symptom inputs. The backbone model is Qwen2.5-7B-Instruct.

(1) Text decoding-based confidence estimation methods. Numerical and Numerical-SC (Li et al., 2024) directly prompt the model to generate confidence scores within the range $[0, 1]$, where Numerical-SC further applies self-consistency (Wang et al., 2022) by averaging multiple generated confidence values. These text decoding-based methods generally underperform most logits-based methods. This is likely because text decoding introduces generation randomness, whereas the BTP method used in this work relies only on the logits of the first output token, thereby avoiding this issue.

(2) Logits-based confidence estimation methods. Compared with BTP and its calibrated variants, FirstLogit and AvgLogit (Ma et al., 2025) achieve relatively lower diagnostic performance. This is likely because they typically estimate confidence for all candidate diseases simultaneously within a single generation process, making confidence estimation more susceptible to interference from contextual information associated with other candidate diseases. In contrast, BTP independently estimates the confidence of each candidate disease, effectively avoiding such interference.

Confidence calibration based on one-hot target distributions can improve the diagnostic performance of BTP to some extent. Building upon this, we further smooth the target distribution used in in-batch contrastive learning, which leads to additional performance gains. This demonstrates the effectiveness of the proposed smoothed confidence calibration strategy.

Model Variant	Acc_{init}	Acc
Trained on DXY	70.3	84.9
Trained on GMD	**70.3**	**87.9**
Trained on CMD	70.3	84.5
Untrained	63.6	81.2

Table 5: Cross-dataset transfer results of MCPlanner

4.9 Cross-Dataset Generalization Analysis

We evaluate the cross-dataset generalization of MCPlanner on GMD using Qwen2.5-7B-Instruct as the backbone. Four model variants are considered: models trained on DXY, GMD, and CMD, respectively, and an untrained variant without confidence calibration or expert-guided distillation.

The cross-dataset transfer results are reported in Table 5. The model trained on GMD achieves the best performance, likely because training and testing on the same data source allows the model to better capture the underlying data distribution, thereby improving in-domain performance. Meanwhile, although the models trained on DXY and CMD perform worse than the GMD-trained model, both outperform the untrained model. This suggests that MCPlanner learns not only dataset-specific distributions but also transferable consultation strategies from clinical consultation data, demonstrating its cross-dataset generalization capability and potential for continual learning.

5 Conclusion

Existing LLM-based MC methods lack adaptive inquiry direction control, which often causes models to become trapped in suboptimal inquiry processes. To address this issue, we propose MCPlanner, a hierarchical adaptive planning framework that simulates the clinical "hypothesis–deduction–verification" consultation process through collaboration between high-level and low-level planning agents. This effectively improves both inquiry direction adaptivity and overall consultation performance. For the core HPlanner, we further introduce expert-guided distillation, which effectively improves the adaptive planning capability of LLMs. In addition, we propose a smoothed in-batch contrastive learning confidence calibration strategy, further improving diagnostic accuracy among highly similar candidate diseases. Experimental results on three MC datasets show that MCPlanner consistently outperforms the baselines, demonstrating its effectiveness for improving LLM-based consultation performance.

Limitations

MCPlanner still has several limitations that warrant future improvement: (1) The consultation information considered is relatively limited. Current inquiry mainly focuses on symptom presence, while other important clinical factors, such as medical history and family history, are not yet incorporated. (2) The model currently generates explicit decision actions rather than adapting them into the diverse natural language expressions commonly used in real-world clinical consultation. (3) Expert-guided distillation relies on costly expert model inference, motivating future exploration of more cost-efficient distillation strategies.

Ethical Consideration

Given the potential hallucination behavior of large language models, the outputs generated by MCPlanner may contain factually inaccurate information. Consequently, the proposed framework is intended exclusively for research purposes and is not designed for direct clinical deployment. All real-world medical decisions should be made by licensed medical professionals.

References

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. 2024. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 17682–17690.

Junying Chen, Chi Gui, Anningzhe Gao, Ke Ji, Xidong Wang, Xiang Wan, and Benyou Wang. 2025a. Cod, towards an interpretable medical agent using chain of diagnosis. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 14345–14368, Vienna, Austria. Association for Computational Linguistics.

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

Sirry Chen, Jieyi Wang, Wei Chen, and Zhongyu Wei. 2026. Speechmedassist: Efficiently and effectively adapting speech language models for medical consultation. *arXiv preprint arXiv:2601.04638*.

Wei Chen, Cheng Zhong, Jiajie Peng, and Zhongyu Wei. 2022b. Dxformer: a decoupled automatic diagnostic system based on decoder–encoder transformer with dense symptom representations. *Bioinformatics*, 39(1):btac744.

Zhenfang Chen, Delin Chen, Rui Sun, Wenjun Liu, and Chuang Gan. 2025b. Armap: Scaling autonomous agents via automatic reward modeling and planning. In *International Conference on Learning Representations*.

Harshita Chopra and Chirag Shah. 2025. Feedback-aware mcts for goal-oriented information seeking. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, San Diego, California, USA. Curran Associates, Inc.

Chengfeng Dou, Fan Yang, Fei Li, Jiyuan Jia, Qiang Ju, Shuai Wang, Tianpeng Li, Xiangrong Zeng, Yijie Zhou, Hongda Zhang, et al. 2026. Baichuan-m3: Modeling clinical inquiry for reliable medical decision-making. *arXiv preprint arXiv:2602.06570*.

Lutfi Eren Erdogan, Nicholas Lee, Sehoon Kim, Suhong Moon, Hiroki Furuta, Gopala Anumanchipalli, Kurt Keutzer, and Amir Gholami. 2025. Plan-and-act: Improving planning of agents for long-horizon tasks. In *International Conference on Machine Learning*, pages 15419–15462. PMLR.

Zhihao Fan, Lai Wei, Jialong Tang, Wei Chen, Wang Siyuan, Zhongyu Wei, and Fei Huang. 2025. Ai hospital: Benchmarking large language models in a multi-agent medical interaction simulator. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10183–10213, Abu Dhabi, UAE. Association for Computational Linguistics.

Yichun Feng, Jiawei Wang, Lu Zhou, Zhen Lei, and Yixue Li. 2025. Doctoragent-rl: A multi-agent collaborative reinforcement learning system for multi-turn clinical dialogue. *arXiv preprint arXiv:2505.19630*.

Dayuan Fu, Biqing Qi, Yihuai Gao, Che Jiang, Guanting Dong, and Bowen Zhou. 2024. Msi-agent: Incorporating multi-scale insight into embodied agents for superior planning and decision-making. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 643–659.

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, Washington, D.C., USA. Association for the Advancement of Artificial Intelligence.

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.

Zhihao Jia, Mingyi Jia, Junwen Duan, and Jianxin Wang. 2025. Ddo: Dual-decision optimization for llm-based medical consultation via multi-agent collaboration. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 26380–26397.

Jinghan Li, Zhicheng Sun, and Yadong Mu. 2025a. Closed-loop long-horizon robotic planning via equilibrium sequence modeling. In *International Conference on Machine Learning*, pages 35884–35908. PMLR.

Jingyao Li, Senqiao Yang, Sitong Wu, Han Shi, Chuanyang Zheng, Hong Xu, and Jiaya Jia. 2025b. Logits-based finetuning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 14763–14775.

Stella Li, Vidhisha Balachandran, Shangbin Feng, Jonathan Ilgen, Emma Pierson, Pang Wei W Koh, and Yulia Tsvetkov. 2024. Mediq: Question-asking llms and a benchmark for reliable interactive clinical reasoning. *Advances in Neural Information Processing Systems*, 37:28858–28888.

Xinyi Ling, Ye Liu, Reza Averly, and Xia Ning. 2026. Uncertainty as a planning signal: Multi-turn decision making for goal-oriented conversation. *arXiv preprint arXiv:2604.03924*.

Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, et al. 2025. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*.

Hongcheng Liu, Yusheng Liao, Siqv Ou, Yuhao Wang, Heyang Liu, Yanfeng Wang, and Yu Wang. 2024. Med-pmc: Medical personalized multi-modal consultation with a proactive ask-first-observe-next paradigm. *arXiv preprint arXiv:2408.08693*.

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 Thirty-First International Joint Conference on Artificial Intelligence*, pages 4266–4272, Vienna, Austria. International Joint Conferences on Artificial Intelligence Organization.

Mingyu Derek Ma, Yanna Ding, Zijie Huang, Jianxi Gao, Yizhou Sun, and Wei Wang. 2025. Inferring from logits: Exploring best practices for decoding-free tive candidate selection. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33125–33144, Vienna, Austria. Association for Computational Linguistics.

Shuofei Qiao, Ningyu Zhang, Runnan Fang, Yujie Luo, Wangchunshu Zhou, Yuchen Jiang, Chengfei Lv, and Huajun Chen. 2024. Autoact: Automatic agent learning from scratch for qa via self-planning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3003–3021.

Zhiyao Ren, Yibing Zhan, Baosheng Yu, Liang Ding, and Dacheng Tao. 2024. Healthcare copilot: Eliciting the power of general llms for medical consultation. *arXiv preprint arXiv:2402.13408*.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36:8634–8652.

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

Haotian Sun, Yuchen Zhuang, Lingkai Kong, Bo Dai, and Chao Zhang. 2023. Adaplanner: Adaptive planning from feedback with language models. *Advances in neural information processing systems*, 36:58202–58245.

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*.

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*.

Huimin Wang, Wai-Chung Kwan, Kam-Fai Wong, and Yefeng Zheng. 2023. Coad: Automatic diagnosis through symptom and disease collaborative generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6348–6361, Toronto, Canada. Association for Computational Linguistics.

Qipeng Wang, Rui Sheng, Yafei Li, Huamin Qu, Yushi Sun, and Min Zhu. 2025. Medkgi: Iterative differential diagnosis with medical knowledge graphs and information-guided inquiring. *arXiv preprint arXiv:2512.24181*.

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*.

Zihao Wang, Shaofei Cai, Anji Liu, Yonggang Jin, Jinbing Hou, Bowei Zhang, Haowei Lin, Zhaofeng He, Zilong Zheng, Yaodong Yang, et al. 2024. Jarvis-1: Open-world multi-task agents with memory-augmented multimodal language models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3):1894–1907.

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.

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, New York, New York, USA. Association for the Advancement of Artificial Intelligence.

Jiao Xu, Junwei Liu, Jiangwei Lao, Qi Zhu, Yunpeng Zhao, Congyun Jin, Shinan Liu, Zhihong Lu, Lihe Zhang, Xin Chen, et al. 2026. Pulsemind: A multimodal medical model for real-world clinical diagnosis. *arXiv preprint arXiv:2601.07344*.

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*, pages 7346–7353, Honolulu, Hawaii, USA. Association for the Advancement of Artificial Intelligence.

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, Istanbul, Turkey. IEEE.

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

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Bolin Zhang, Shengwei Wang, Yangqin Jiang, Dianbo Sui, Zhiying Tu, and Dianhui Chu. 2025. Ask and retrieve knowledge: Towards proactive asking with imperfect information in medical multi-turn dialogues. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1055–1065, Padua, Italy. Association for Computing Machinery.

Shaoqing Zhang, Zhuosheng Zhang, Kehai Chen, Xinbei Ma, Muyun Yang, Tiejun Zhao, and Min Zhang. 2024. Dynamic planning for llm-based graphical user interface automation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1304–1320.

Wanpeng Zhang and Zongqing Lu. 2024. Adarefiner: Refining decisions of language models with adaptive feedback. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 782–799.

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 *Proceedings of the 27th European Conference on Artificial Intelligence*, pages 1495–1502, Santiago de Compostela, Spain. IOS Press BV.

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.

Zihan Zhou, Yinan Liu, Yuyang Xie, Bin Wang, Xiaochun Yang, and Zezheng Feng. 2026. Diaglink: A dual-user diagnostic assistance system by synergizing experts with llms and knowledge graphs. *arXiv preprint arXiv:2601.20311*.

Jiayuan Zhu, Jiazhen Pan, Yuyuan Liu, Fenglin Liu, and Junde Wu. 2025. Ask patients with patience: Enabling llms for human-centric medical dialogue with grounded reasoning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 2846–2857, Suzhou, China. Association for Computational Linguistics.

A Related Work

A.1 Medical Consultation Task

Early MC methods modeled the consultation process as a Markov decision process and constructed doctor agents using deep reinforcement learning (Tang et al., 2016; Wei et al., 2018; Xu et al., 2019), but suffered from high decision randomness (Xia et al., 2020). Sequence generation-based methods treat each consultation action as an individual token and significantly improve performance on this task through large-scale supervised training (Chen et al., 2022b,a; Wang et al., 2023; Hou et al., 2023; Zhao et al., 2024). In recent years, owing to their strong domain generalization capabilities, LLMs have been widely adopted for MC (Fan et al., 2025; Li et al., 2024; Ren et al., 2024). To improve consultation performance in complex disease

scenarios, researchers have introduced techniques such as retrieval-augmented generation (Zhang et al., 2025; Zhou et al., 2026), uncertainty-aware planning (Hu et al., 2024; Chopra and Shah, 2025), supervised fine-tuning (Tu et al., 2024; Chen et al., 2025a; Jia et al., 2025), and reinforcement learning (Feng et al., 2025; Dou et al., 2026) into LLM-based MC. Beyond text modality, recent studies have also explored MC in image modality (Liu et al., 2024; Xu et al., 2026) and speech modality (Chen et al., 2026).

A.2 LLMs in Dynamic Planning

Traditional LLM planning and reasoning methods (Wei et al., 2022; Yao et al., 2023; Besta et al., 2024) are primarily designed for static planning tasks, such as mathematical problem solving and programming, where the environment is fully observable. In partially observable environments, such as robotic navigation, web interaction, and MC, LLMs are required to perform real-time self-refinement of their planning process (Li et al., 2025a; Zhang et al., 2024). Methods based on the thought-action-observation paradigm (Yao et al., 2022; Shinn et al., 2023; Qiao et al., 2024) reason over environmental feedback to dynamically determine subsequent actions. Some studies explicitly model planning sequences and iteratively update action execution plans after each environmental feedback. Hierarchical planning has also shown strong performance in complex tasks (Sun et al., 2023; Zhang and Lu, 2024; Erdogan et al., 2025). Selection-based methods (Chen et al., 2025b; Ling et al., 2026; Chopra and Shah, 2025) simulate future rollouts to select actions with maximal information gain. Other approaches leverage previously successful experiences to guide planning decisions, thereby improving the efficiency of dynamic planning in LLMs (Fu et al., 2024; Wang et al., 2024).

B Other Experimental Details

B.1 Dataset 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. The train/dev/test splits for DXY, GMD, and CMD are 318/103/103, 1912/239/239, and 3379/671/1342, respectively.

MCPlanner uses the same calibration sub-trajectory data as DDO. The numbers of confidence calibration training samples for the DXY, GMD, and CMD datasets are 2,185, 13,598, and 54,608, respectively.

When constructing distillation training data, considering the relatively large training scales of the GMD and CMD datasets, we perform stratified sampling over disease categories to improve distillation efficiency. Specifically, an upper bound is imposed on the number of samples for each disease, while representative consultation records are retained to construct a distillation subset. After preprocessing, we obtain 2,028, 2,074, and 3,367 expert demonstrations for distillation training on DXY, GMD, and CMD, respectively.

B.2 Implementation Details of MCPlanner

During the Diagnostic Confidence Estimation and Calibration stage, efficient fine-tuning is performed using LoRA. Hyperparameter tuning and checkpoint selection are based on Top-k diagnostic accuracy and disease-level diagnostic performance. The key hyperparameters are set as follows: the softmax temperature in the KL-divergence loss is fixed at 1.0, the maximum number of training steps is 1000, the learning rate is 5×10^{-5}, the batch size is 8, and the LoRA rank is set to 16. The minimum confidence margin parameter $margin$ between the ground-truth disease and the strongest competing disease is adjusted according to disease similarity within each dataset. Since candidate diseases in DXY mainly consist of pediatric diseases with high clinical similarity, we set $margin = 0.05$. For GMD and CMD, which contain diseases from more diverse clinical departments, $margin$ is set to 0.15.

During the expert-guided distillation stage, dis-

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

tillation data are first constructed from the training set. GPT-5-mini is used as the teacher model, while Qwen2.5-7B-Instruct serves as the student model. To improve model selection efficiency, distilled model selection is not performed on the full validation set. Instead, we construct a hard-sample validation subset. Specifically, the undistilled model first performs the adaptive hierarchical consultation process on the validation set, and samples with incorrect final diagnoses are selected as hard samples. Hyperparameter tuning and checkpoint selection are then conducted on this subset. LoRA is also used for efficient fine-tuning during distillation. The key hyperparameters are set as follows: the number of training epochs is 3, the learning rate is 1×10^{-4}, the batch size is 8, and the LoRA rank is set to 16.

During inference, the initial candidate disease window size is set to 5 based on initial Top-k diagnostic accuracy, balancing label coverage and consultation efficiency. The maximum consultation turn limit is set to $L = 10$.

B.3 Model Deployment

Qwen3-8B and Qwen2.5-7B-Instruct were locally deployed on a single NVIDIA RTX 3090 GPU. Qwen2.5-14B-Instruct was locally deployed on a single NVIDIA Tesla V100 GPU. GPT-5-mini[4] and DeepSeek-V3[5] were accessed via online APIs.

C Standard Errors and Confidence Intervals

Table 6 reports the randomness analysis of MCPlanner under multiple random seeds, including the mean diagnostic accuracy, standard error, and 95% confidence interval on three MC datasets. The backbone model is Qwen2.5-7B-Instruct after expert-guided distillation and Diagnostic Confidence Estimation and Calibration.

Dataset	Mean Accuracy	Standard Error	95% Confidence Interval (%)
DXY	84.4	1.1	3.0
GMD	87.2	0.6	1.6
CMD	68.1	0.2	0.8

Table 6: Standard Errors and Confidence Intervals of MCPlanner on three MC datasets

The results show that MCPlanner achieves relatively small standard errors across all three datasets (with a maximum of 1.1) and narrow confidence intervals, with the confidence interval on CMD being only 0.76. This indicates that MCPlanner exhibits stable performance under different random seeds. It can also be observed that the standard error and confidence interval on DXY are slightly larger than those on GMD and CMD. This is likely because DXY has a relatively smaller sample size, leading to slightly lower performance stability compared with the other two datasets. The main experimental results are reported using the best-performing random seed.

D Case Study

Table 7 presents an adaptive consultation case of MCPlanner. The patient's ground-truth disease is pneumonia, and the initial symptom report includes nasal congestion, fever, rhinorrhea, cough, sneezing, and sputum production. Since these symptoms are common across multiple diseases, such as upper respiratory tract infection, pneumonia, and allergic rhinitis, pneumonia is ranked only third among the initial candidate diseases estimated by DAgent.

HPlanner initially selects upper respiratory tract infection as the inquiry direction. Although early patient feedback is largely consistent with this hypothesis, it remains insufficient to distinguish it from competing diseases such as pneumonia, leading HPlanner to continue evidence collection.

As consultation proceeds, the remaining candidate symptoms provided by LPlanner become less informative. HPlanner therefore switches the inquiry direction to pneumonia, which offers higher diagnostic value.

During subsequent interactions, the key symptom "pulmonary shadow" is confirmed, providing decisive evidence. The consultation is then terminated, and pneumonia rises to the top of the diagnostic ranking.

This case shows that MCPlanner can dynamically update disease hypotheses and adaptively adjust inquiry directions through iterative evidence collection and verification. Moreover, the hierarchical planning mechanism improves efficiency by avoiding full-model reasoning at every turn for symptom selection.

E Prompt Templates

Tables 8 to 11 present the prompt templates used in MCPlanner.

[4] https://api.chatanywhere.org/
[5] http://i.hpc.csu.edu.cn/csu

Original Medical Consultation Record		
Ground-truth Disease: Pneumonia **Complete Symptom Information:** Present: nasal congestion, fever, rhinorrhea, cough, sneezing, sputum production, irritability, dyspnea		
Initial Information		
Initial Symptom Information: Present: nasal congestion, fever, rhinorrhea, cough, sneezing, sputum production **Initial Candidate Disease Ranking:** 1. Upper Respiratory Tract Infection; 2. Allergic Rhinitis; 3. Pneumonia; 4. Hand-Foot-and-Mouth Disease; 5. Pediatric Diarrhea		
High-Level Planning Decision and Rationale	**Low-Level Inquiry Sequence**	**Interaction Record**
Initialize the inquiry direction as Upper Respiratory Tract Infection (highest initial diagnostic confidence)	[Vomiting, Throat Discomfort, Irritability, Lethargy, Dyspnea, Loss of Appetite]	Vomiting: Present Throat Discomfort: Absent Irritability: Present Lethargy: Absent Dyspnea: Present
Switch the inquiry direction to Pneumonia (limited information gain from remaining symptoms; switch to a more discriminative direction)	[Foaming at the Mouth, Moist Rales, Pulmonary Shadow, Loss of Appetite]	Foaming at the Mouth: Present Moist Rales: Absent Pulmonary Shadow: Present
Terminate consultation and diagnose Pneumonia (key evidence "Pulmonary Shadow" confirms the diagnosis and excludes other candidates)	–	–

Table 7: An adaptive consultation case of MCPlanner. For clarity, only brief reasoning content is shown.

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

## Patient Symptom Information

- Present symptoms: {positive_symptoms}.

- Absent symptoms: {negative_symptoms}.

## Symptom Knowledge of Disease '{candidate_disease}'

- Symptom occurrence frequencies estimated from historical consultation records diagnosed as '{candidate_disease}': {empirical_knowledge}

Based on the patient's symptom information and the symptom knowledge of disease '{candidate_disease}', determine whether this disease can be diagnosed.

The presence of disease-related symptoms increases diagnostic confidence, while their absence decreases it.

The impact of each symptom on diagnostic confidence increases with its typicality for the disease.

The output must be either True or False: output True if the disease is considered diagnosable, otherwise output False.

Output only the judgment result without any additional explanation.
```

Table 8: DAgent's prompt template

You are an experienced medical expert. Your task is to adaptively control the inquiry direction during multi-turn medical consultation based on the currently available information.

[Task Description]

Definition of Inquiry Direction

The inquiry direction refers to the tendency to focus symptom inquiry on a specific disease during consultation.

Definition of Flow Control Actions

1. END: The diagnosis has been determined and the consultation should terminate.

2. KEEP: The current inquiry direction remains worth exploring and should be maintained.

3. SWITCH: The current inquiry direction is no longer worth exploring and should be replaced with a new one.

Decision Rules

1. If the currently observed symptoms strongly support the current inquiry direction and can exclude other candidate diseases, select END.

2. Otherwise, determine whether the inquiry direction should be switched.

Select KEEP if both of the following conditions are satisfied; otherwise, select SWITCH:

(1) The observed symptoms are generally consistent with the expected manifestations under the current inquiry direction.

(2) There remain highly informative symptoms to query under the current inquiry direction.

3. If SWITCH is selected, provide a new inquiry direction:

(1) The new direction must be selected from the current candidate diseases, preferably an unexplored one.

(2) It should be as consistent as possible with the currently observed symptoms.

[Known Information]

Patient's Initial Reported Symptoms

{self_report_str}

Historical Inquiry Directions and Queried Symptoms

{directions_symptoms_str}

Remaining Symptoms under the Current Inquiry Direction

{planning_symptoms_str}

Candidate Diseases and Their Symptom Occurrence Frequencies

{sub_disease_knowledge}

[Output Requirement]

Adaptively determine the inquiry direction based on the task description and known information, and provide the main rationale.

Output in the following format:

json: {"action": "SWITCH", "new_direction": "Disease B", "reason": "Decision rationale"}

Table 9: HPlanner's prompt template

You are an experienced medical expert. Your task is to plan the subsequent symptom inquiry sequence based on the known information during consultation.
[Known Information]
Current Inquiry Direction and Symptom Occurrence Frequencies
{masked_direction_knowledge}
Relatively Distinctive Symptoms under the Current Inquiry Direction
{typical_symptoms}
[Decision Rules]
1. Symptoms in the inquiry sequence should be relatively common under the current inquiry direction.
2. If multiple symptoms have similar typicality, more distinctive symptoms should be prioritized.
[Output Requirement]
Based on the known information and decision rules, plan the subsequent symptom inquiry sequence.
Output in the following format without any explanation:
python: ["Symptom 1", "Symptom 2", "Symptom 3", "Symptom 4", "Symptom 5", ...]

Table 10: LPlanner's prompt template

You are a patient simulator. The patient you simulate is diagnosed with '{disease_label}'.
The symptom knowledge of disease '{disease_label}' is as follows:
- Symptom occurrence frequencies estimated from historical consultation records diagnosed as '{disease_label}': {disease_knowledge}
Your task is to determine whether the patient is likely to exhibit symptom '{symptom}' based on the disease symptom knowledge.
The output must be either True or False: True indicates the symptom is likely present, while False indicates it is unlikely present.
The decision criteria are as follows:
- If the symptom is relatively typical for disease '{disease_label}' (i.e., ranked high in occurrence frequency), output True.
- Otherwise, output False.
Think step by step and determine whether symptom '{symptom}' is likely present.

Table 11: PAgent's prompt template