MEDPRECEPTOR: DECOUPLED PROCESS SUPERVISION FOR ACTIVE CLINICAL INQUIRY IN MULTI-TURN MEDICAL CONSULTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Effective medical consultation requires not only diagnostic reasoning over existing evidence, but also actively eliciting crucial clinical information across multi-turn interactions. However, both human clinicians and Medical LLMs frequently exhibit process-level drawbacks during clinical inquiry, such as repetitive questioning, misdirected inquiry, omission of key evidence, and premature diagnostic closure. Existing methods typically address this by fine-tuning the Doctor LLM itself, which tightly couples consultation execution with process supervision and impedes reusability across heterogeneous models and human clinicians. To address this limitation, we introduce MedPreceptor, a decoupled process supervision framework that selectively intervenes in multi-turn consultations while leaving the underlying Doctor LLM unchanged. To ensure robust decision-making, we construct prefix-aligned supervision via a two-stage annotation pipeline that eliminates future information leakage, and develop Decision-Guided Selective On-Policy Distillation (DG-OPD). By selectively routing teacher signals conditioned on intervention decisions, DG-OPD trains the supervisor on trajectories from three Doctor LLMs alongside real doctor-patient dialogues, enabling it to learn both when to intervene and what actionable guidance to provide. In addition, we construct the Clinically Annotated Consultation Intervention Benchmark (CACI), comprising 1,664 expert-annotated consultation states for evaluating intervention timing and the quality of corrective advice. Across five Doctor LLMs on MAQuE, MedPreceptor improves diagnostic accuracy by up to 10.27% points; across seven Doctor LLMs on MedDialogRubrics, it increases clinically important information coverage by up to 7.08 % points. On CACI, MedPreceptor achieves 66.11% accuracy, 67.70% F1, and a corrective advice quality score of 7.71/10. These results demonstrate that decoupling process supervision from consultation execution offers an effective, generalizable paradigm for advancing active clinical inquiry across diverse medical LLMs. Code is available at https://anonymous.4open.science/r/MedPreceptor-5E1F.

1 INTRODUCTION

Medical consultation is central to clinical decision-making. While recent Medical LLMs exhibit remarkable capabilities in medical knowledge retrieval and diagnostic reasoning, successful consultation depends equally on active clinical inquiry, i.e., the dynamic process of progressively eliciting patient evidence across multi-turn interactions. In clinical workflows, evidence is rarely presented upfront; a physician must actively probe for symptoms, identify missing clinical context, and dynamically refine inquiry trajectories. However, both human practitioners and clinical LLMs frequently stumble during this sequential process. Prior clinical and NLP literature (Qiao et al., 2026; Gong et al., 2026; 2025; Fang et al., 2026) identifies five widespread process-level pitfalls: *repeated questioning*, *inefficient inquiry*, *misdirected focus*, *omission of key clinical clues*, and *premature diagnostic closure*. As shown in Figure 1, these issues are not unique to artificial models: retrospective analyses of diagnostic errors among human clinicians indicate that flawed data gathering contributes to 84.3% of diagnostic errors, with cognitive biases such as premature closure (18.2%)

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

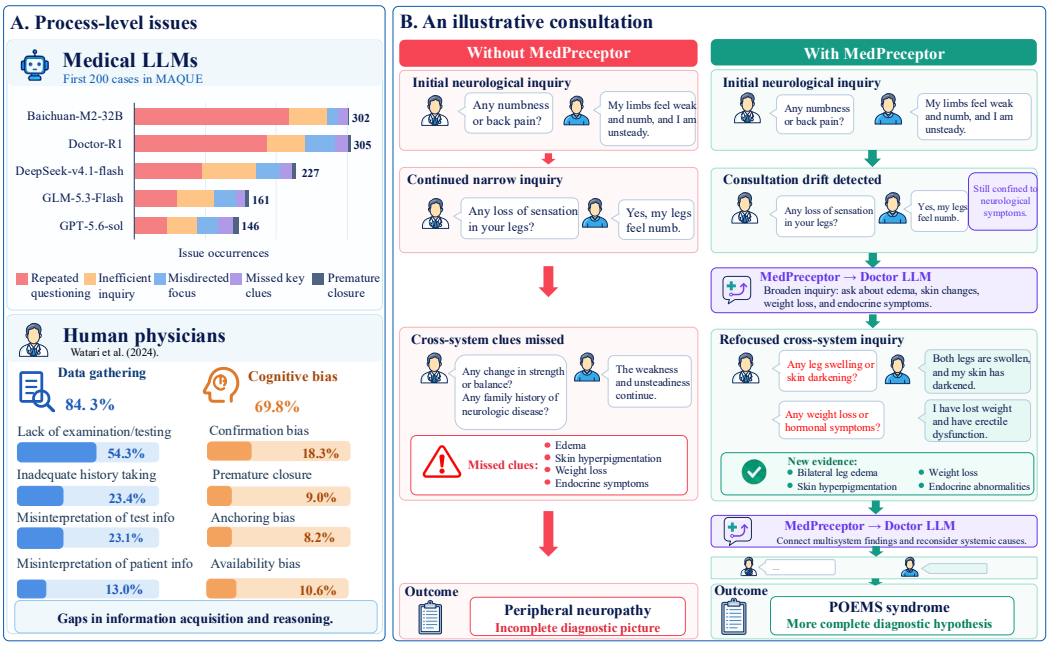


Figure 1: **Motivation of process supervision during multi-turn medical consultations**. **(A) Process-level issues.** Doctor LLMs present five issue on 200 MAQuE test cases: *repeated questioning*, *inefficient inquiry*, *misdirected focus*, *omission of key clinical clues*, and *premature diagnostic closure*. Studies on human physician show that data gathering (84.3%) and cognitive bias (69.8%) are major contributing factors, with inadequate history taking and premature closure as examples. Both reveal gaps in information acquisition and reasoning. **(B) A comparative example.** Without MedPreceptor, narrow inquiry misses cross-system clues and yields an incomplete peripheral neuropathy diagnosis; with it, selective intervention broadens inquiry to recover missing evidence and support a more complete POEMS syndrome diagnosis.

and anchoring bias (8.2%) frequently truncating necessary inquiry (Watari et al., 2024; Singh et al., 2013; Beckman & Frankel, 1984; Angus et al., 2012).

Existing methods typically address these inquiry failures by fine-tuning the Doctor LLM itself, either employing reinforcement learning for strategic question selection, adaptive evidence acquisition, or trajectory-based clinical action planning (Lai et al., 2026; Feng et al., 2025; Hsu et al., 2026). While effective in isolated settings, these approaches tightly couple consultation execution with process regulation within the Doctor LLM's internal parameters. This coupling presents significant drawbacks: it renders the supervisory capability non-reusable across heterogeneous or proprietary models, risks disrupting the underlying model's inherent diagnostic reasoning style, and requires expensive retraining whenever the underlying base model is updated. We therefore propose a complementary, decoupled paradigm: an **external process supervisor** that operates alongside any arbitrary Doctor LLM, silently monitoring consultation progression and intervening selectively only when inquiry deviates from sound clinical paths.

Realizing such a decoupled supervisor presents three key practical challenges. First, constructing reliable supervision is inherently difficult because many consultation pitfalls are straightforward to identify only in hindsight after observing the complete dialogue trajectory. In a realistic interaction, however, an intervention decision can only rely strictly on the dialogue prefix accumulated up to that point, making it essential that the supervisor learns to evaluate inquiry without leaking or relying on future clinical information. Second, different Doctor LLMs naturally follow varied yet clinically reasonable inquiry strategies. A practical supervisor cannot simply enforce a rigid, monolithic question sequence, but must instead identify genuine consultation pitfalls across diverse conversational styles and clinical habits. Finally, the supervisor must jointly learn both when and how to intervene,

accurately deciding to remain unobtrusively silent when the consultation is progressing well, while delivering concise, clinically actionable guidance whenever an intervention is truly warranted.

To address these challenges, we introduce **MedPreceptor**, a decoupled process supervisor for multi-turn medical consultations. To mitigate hindsight leakage, we develop a two-stage annotation pipeline that first identifies candidate intervention points using complete trajectories and then re-adjudicates them strictly under prefix-only constraints. To improve cross-model generalizability, we construct a diverse training corpus spanning trajectories from three Doctor LLMs and authentic doctor-patient dialogues, with annotations focused on model-agnostic inquiry failures. We further propose **Decision-Guided Selective On-Policy Distillation (DG-OPD)** to jointly learn intervention timing and guidance generation. DG-OPD conditions teacher supervision on the student's sampled intervention decision, distilling high-quality guidance when intervention is warranted while suppressing unnecessary guidance when no intervention is warranted. As a model-agnostic supervisor, MedPreceptor can guide different Doctor LLMs without modifying their underlying weights.

Our main contributions are summarized as follows:

- We present MedPreceptor, a decoupled process supervision framework that selectively intervenes in multi-turn clinical consultations without modifying the underlying Doctor LLM, enabling model-agnostic supervision across different Doctor LLMs.
- We develop a two-stage prefix-aligned annotation pipeline to prevent hindsight leakage and introduce Decision-Guided Selective On-Policy Distillation (DG-OPD), combining prefix-aligned supervision with selective policy distillation to jointly optimize intervention timing and actionable clinical guidance.
- We construct the Clinically Annotated Consultation Intervention Benchmark (CACI), comprising 1,664 expert-annotated consultations to evaluate intervention timing and advice quality. Across multiple existing medical LLMs, MedPreceptor improves diagnostic accuracy by up to 10.27 percentage points on MAQuE and clinical evidence coverage by up to 7.08 percentage points on MedDialogRubrics. On CACI, MedPreceptor achieves 66.11% accuracy, 67.70% F1, and an advice quality score of 7.71/10, with consistent improvements across different Doctor LLMs.

2 Related Work

Interactive and active medical consultation. Medical LLMs are increasingly moving beyond static question answering toward multi-turn consultation, which provides a more realistic formulation of clinical practice. In this setting, models must progressively acquire patient information and update clinical reasoning as new evidence emerges. However, existing methods primarily aim to improve final diagnostic accuracy or sequential decision making, treating information acquisition as an intermediate means. For example, AgentClinic places the Doctor in an interactive environment and requires it to adapt actions to newly revealed information (Schmidgall et al., 2026). MedAction formulates diagnosis as active evidence acquisition, while Doctor-R1 trains Doctors to ask high-value questions and conduct strategic multi-turn inquiry (Hsu et al., 2026; Lai et al., 2026). Some benchmarks further evaluate information-gathering quality, such as coverage and relevance, but the optimization target remains the Doctor itself. In contrast, MedPreceptor treats the ongoing information-acquisition process as the direct object of external supervision.

Simulated dialogue, external feedback, and selective process supervision. Simulated dialogue provides both training signals and evaluation settings for consultation LLMs. AMIE improves clinical history taking and diagnostic dialogue through simulated self-play and automated feedback, while CRAFT-MD and MAQuE evaluate multi-turn consultation by simulating Doctor–Patient interactions and diverse patient behaviors (Tu et al., 2025; Johri et al., 2025; Gong et al., 2025). Beyond simulated interaction, external critics and planners provide a modular alternative to directly modifying the main LLM. Critic-based methods evaluate generated responses and provide feedback for refinement: CriticBench studies generation–critique–correction, while MultiCritique and RCO train critics to produce more effective feedback (Lin et al., 2024; Lan et al., 2025; Yu et al., 2025). In dialogue, PsyAdvisor and straQ* use plug-and-play planners that observe dialogue state and propose subsequent strategies (Hu et al., 2025; Wang et al., 2025). These approaches share our modular use of external supervision but target different problems: critics refine generated responses, while planners decide next actions or strategies. Medical consultation additionally requires selec-

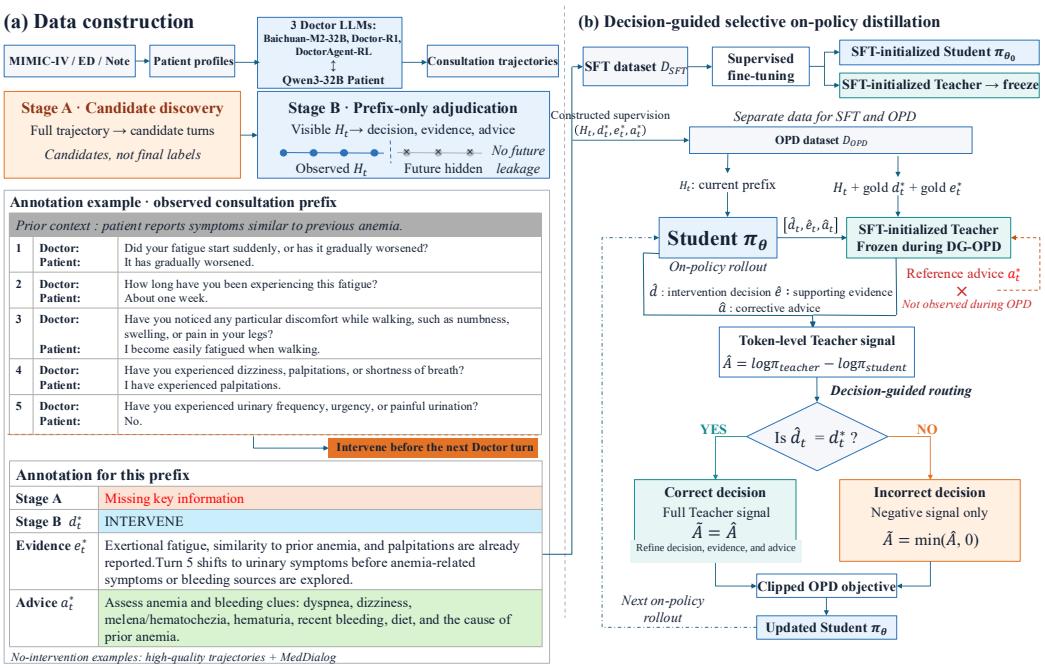


Figure 2: Overview of the MedPreceptor training framework.

tive process supervision—determining whether the current inquiry exhibits a process-level issue that warrants intervention, while remaining silent when the Doctor follows a clinically reasonable path. MedPreceptor focuses on this selective supervision of the ongoing information-acquisition process.

3 MedPreceptor Design

This section presents the overall framework of **MedPreceptor**, which uses selective external intervention to improve information acquisition in multi-turn medical consultation. The task involves two core capabilities: deciding *when* to intervene and, when necessary, *how* to provide effective guidance. We first introduce the application scenarios, followed by the task formulation. Then we describe prefix-aligned intervention supervision in Section 3.3, and introduce **Decision-Guided Selective On-Policy Distillation (DG-OPD)** in Section 3.4. Figure 2 shows the training framework.

3.1 Application Scenarios

MedPreceptor is designed for multi-turn medical consultations, whether conducted by a human clinician or a medical LLM. In a physical hospital or clinic, the system monitors the spoken conversation between a patient and a doctor. When it identifies pitfalls in the dialogue, such as missed clinical clues or premature conclusions, it provides a prompt to the physician with specific recommendations for a more effective inquiry. When a patient consults a medical LLM for a diagnosis, the system oversees the textual exchange. It continuously evaluates the conversation and directly guides the artificial intelligence whenever the questioning loses focus. This targeted assistance ensures the underlying model asks the necessary follow-up questions to achieve an accurate diagnosis.

3.2 Problem Formulation

We formulate **MedPreceptor** as an external intervention policy for multi-turn medical consultation. Let $\mathcal{H}_t$ denote the consultation history available at time t, containing only the Doctor–Patient interactions that have already occurred. Given $\mathcal{H}_t$, MedPreceptor decides whether intervention is needed and generates supporting evidence e_t. When $d_t = 0$, no intervention is needed and $a_t = \text{None}$.

When $d_t = 1$, e_t identifies the current issue, and MedPreceptor further generates corrective advice a_t to guide the Doctor's subsequent information gathering.

We denote the MedPreceptor policy by π_θ and factorize its output distribution as

$$\pi_\theta(d_t, e_t, a_t \mid \mathcal{H}_t) = p_\theta(d_t \mid \mathcal{H}_t)\, \pi_\theta(e_t, a_t \mid \mathcal{H}_t, d_t).$$

Here, $p_\theta(d_t \mid \mathcal{H}_t)$ determines whether intervention is needed, while $\pi_\theta(e_t, a_t \mid \mathcal{H}_t, d_t)$ generates the supporting evidence for the decision and, when intervention is needed, the corresponding corrective advice.

3.3 Prefix-Aligned Intervention Supervision

We select clinical cases with sufficiently complete histories and explicit diagnoses from MIMIC-IV, MIMIC-IV-ED, and MIMIC-IV-Note (Johnson et al., 2024; 2023a;b). Following the data organization of MAQuE (Gong et al., 2025), each case is converted into a structured Patient profile. Clinical information is revealed progressively by the Patient Simulator in response to the Doctor's questions, while gold information, including the final diagnosis, remains hidden from the Doctor. We use Qwen3-32B (Yang et al., 2025) as the Patient Simulator and Baichuan-M2-32B, Doctor-R1, and DoctorAgent-RL (Baichuan-M2 Team, 2025; Lai et al., 2026; Feng et al., 2025) as Doctor LLMs. Each Doctor conducts a multi-turn consultation with the simulated Patient for up to 16 turns, producing Doctor–Patient trajectories with diverse information-gathering strategies.

3.3.1 Two-Stage Intervention Annotation

To obtain reliable intervention annotations, we use a two-stage annotation process. We first use the full consultation trajectory to identify candidate intervention points, and then re-evaluate each point using only the dialogue prefix available at that time to avoid future information leakage.

Stage A: Full-trajectory candidate discovery. The Annotation LLM reviews each complete Doctor–Patient trajectory and marks potential intervention points based on the common process-level issues summarized in Appendix C.2, including *repeated questioning*, inefficient inquiry, misdirected focus, omission of key clinical clues, and premature diagnostic closure. These categories are used only to identify candidate points and are not prediction targets for MedPreceptor.

Stage B: Prefix-only adjudication. For each candidate point t, we remove all subsequent dialogue and keep only the consultation history $\mathcal{H}_t$ available at that time. The Annotation LLM then re-evaluates whether intervention is needed using only this prefix. For confirmed intervention points, it further generates supporting evidence $e_t^\star$ and corrective advice $a_t^\star$, represented as

$$(\mathcal{H}_t, d_t^\star, e_t^\star, a_t^\star), \qquad d_t^\star = 1.$$

In this way, the full trajectory is used only to locate candidate points, while the final supervision is generated from information available at the actual intervention point, avoiding future-information leakage.

3.3.2 Non-Intervention Annotation

To train MedPreceptor to remain silent when intervention is unnecessary, we use the Annotation LLM to construct no-intervene supervision from two sources: high-quality LLM-generated consultation trajectories with correct final diagnoses and no clear process-level issues, and high-quality multi-turn medical conversations from MedDialog (Zeng et al., 2020). For each extracted consultation prefix $\mathcal{H}_t$, the Annotation LLM assigns $d_t^\star = 0$ and generates supporting evidence $e_t^\star$ explaining why no intervention is needed, without generating corrective advice.

3.4 Decision-Guided Selective On-Policy Distillation

3.4.1 Supervised Policy Initialization

Based on the prefix-aligned intervention supervision constructed in the previous section, we first perform supervised fine-tuning on both the student and teacher using the annotated data $(\mathcal{H}_t, d_t^\star, e_t^\star, a_t^\star)$. The training set contains both intervention and non-intervention instances, allowing both models to

learn whether intervention is needed from the current consultation prefix and to generate evidence supporting the decision. For intervention instances, they additionally learn to generate corrective advice. After supervised fine-tuning, we obtain the initial student policy π_{θ_0} and a task-adapted teacher. The subsequent on-policy distillation starts from π_{θ_0}, with the student generating rollouts from its current policy. The teacher remains frozen during this stage and is used only to score the student outputs and provide distillation signals.

3.4.2 DECISION-GUIDED ON-POLICY DISTILLATION

Building on the supervised fine-tuning stage, we further optimize the initial Student policy π_{θ_0}. Direct reinforcement learning is not well suited to this task, because the benefit of an intervention often becomes clear only after several subsequent Doctor–Patient interactions. As a result, rewards based on final diagnosis or overall consultation quality cannot reliably measure the contribution of a specific intervention or its advice, leading to difficult credit assignment. We therefore adopt on-policy distillation, which provides direct Teacher supervision on outputs generated by the Student's current policy without relying on a hard-to-design scalar reward.

However, directly applying standard distillation to all on-policy outputs is also not suitable for MedPreceptor. Prior studies have shown that Teacher guidance on Student-generated trajectories is not always reliable; when the current trajectory has already deviated from a valid solution path, the distillation signal may further push the Student in the wrong direction and degrade performance (Gan et al., 2026; Fu et al., 2026). MedPreceptor first needs to determine whether intervention is needed, and corrective advice is meaningful for optimization only when the intervention decision is correct and intervention is indeed required. Therefore, we propose DG-OPD, which selectively performs subsequent distillation based on the Student's intervention decision.

Specifically, at on-policy training iteration k, the current Student policy generates

$$\hat{\mathbf{s}}_t^{(k)} = [\hat{d}_t^{(k)}; \hat{e}_t^{(k)}; \hat{a}_t^{(k)}] \sim \pi_{\theta_k}(\cdot \mid \mathcal{H}_t),$$

where $\hat{d}_t^{(k)}$, $\hat{e}_t^{(k)}$, and $\hat{a}_t^{(k)}$ denote the generated intervention decision, supporting evidence, and corrective advice, respectively.

We denote the fixed teacher policy as π_{te}. During distillation, the teacher is provided with the consultation context $\mathcal{H}_t$, the reference intervention decision $d_t^\star$, and the reference supporting evidence $e_t^\star$, but *not* the reference corrective advice. The teacher therefore guides the Student's generated advice based on the correct intervention decision and its supporting evidence, rather than directly imitating a reference suggestion. For no-intervene outputs, the corrective advice is empty. Following standard on-policy distillation, for each student-generated token $\hat{s}_{t,j}^{(k)}$ in the semantic output fields (intervention decision, supporting evidence, and corrective advice), we define the token-level Teacher signal as

$$\hat{A}_{t,j}^{(k)} = \log \pi_{\mathrm{te}}\left(\hat{s}_{t,j}^{(k)} \mid \mathcal{H}_t, d_t^\star, e_t^\star, \hat{s}_{t,<j}^{(k)}\right) - \log \pi_{\theta_k}\left(\hat{s}_{t,j}^{(k)} \mid \mathcal{H}_t, \hat{s}_{t,<j}^{(k)}\right).$$

We then modify this signal according to whether the student makes the correct intervention decision:

$$\widetilde{A}_{t,j}^{(k)} = \begin{cases} \hat{A}_{t,j}^{(k)}, & \hat{d}_t^{(k)} = d_t^\star, \\ \min\left(\hat{A}_{t,j}^{(k)}, 0\right), & \hat{d}_t^{(k)} \neq d_t^\star. \end{cases}$$

When the intervention decision is correct, we retain the full Teacher signal to refine the Student-generated evidence and corrective advice. Importantly, the Teacher does not observe a reference advice; its guidance is conditioned only on the correct decision and supporting evidence. When the decision is incorrect, we retain only negative signals, suppressing Student-generated tokens that receive lower probability under the Teacher without positively reinforcing content generated along an incorrect decision path. Since the intervention decision itself is part of $\hat{\mathbf{s}}_t^{(k)}$, it also receives this corrective signal.

Finally, we replace the standard OPD token-level signal with $\widetilde{A}_{t,j}^{(k)}$ and optimize the Student using the standard clipped OPD objective. Thus, the intervention decision controls how Teacher supervision is applied while the training remains fully on-policy. The detailed algorithm is provided in Algorithm 1.

Table 1: Results on the MAQuE validation set across diagnostic accuracy and consultation-quality metrics.

Doctor	Preceptor	Doctor Acc.	GPT-5.6 Acc.	Coverage	Relevance	Adherence	Coherence	Clarity	Empathy
Baichuan-M2-32B	–	50.45	66.34	0.313	0.879	0.846	0.576	0.682	0.404
	✓	**60.72**	**68.24**	**0.3618**	**0.9008**	**0.8752**	**0.6503**	**0.6817**	**0.4094**
Doctor-R1	–	46.00	67.48	0.333	0.902	0.900	0.678	0.706	0.408
	✓	**54.79**	**69.52**	**0.350**	**0.908**	**0.910**	**0.686**	**0.712**	0.408
DeepSeek-v4.1-flash	–	56.97	69.79	0.354	0.872	**0.945**	0.817	0.803	0.523
	✓	**62.14**	**71.97**	**0.363**	**0.879**	0.943	**0.818**	0.803	0.523
GPT-5.6-sol	–	71.97	71.97	0.209	0.927	0.958	0.873	0.592	0.392
	✓	**74.62**	**74.62**	**0.233**	**0.937**	**0.966**	**0.874**	**0.613**	**0.394**
GLM-5.3-Flash	–	68.59	72.00	0.365	0.870	0.986	**0.881**	**0.802**	0.499
	✓	**73.83**	**73.00**	**0.395**	**0.873**	**0.987**	0.876	0.798	**0.501**

4 Experimental Evaluation

4.1 Evaluation Setup

We primarily evaluate MedPreceptor on two public multi-turn medical consultation benchmarks, MAQuE (Gong et al., 2025) and MedDialogRubrics (Gong et al., 2026), which assess overall consultation quality and clinically important information gathering, respectively. For each Doctor, we compare performance with and without MedPreceptor. MedPreceptor continuously monitors the consultation and provides guidance to the Doctor when intervention is needed.

To directly evaluate intervention capability, we further construct the **Clinically Annotated Consultation Intervention Benchmark (CACI)**. CACI covers six clinical specialties and contains 1,664 consultation states, including 720 intervention-required states with specific corrective advice and 944 no-intervention states. All samples are independently reviewed by three master's students in clinical medicine and reconciled through the protocol described in Appendix B.3. They are used to evaluate intervention timing and advice quality.

4.2 Main Results

4.2.1 General Consultation Performance

We evaluate overall multi-turn consultation performance on MAQuE (Gong et al., 2025). For each Doctor, we compare complete consultation trajectories generated with and without MedPreceptor under the same cases and consultation setting. We report two diagnostic accuracy metrics: the original Accuracy and a unified Accuracy obtained by re-evaluating all consultations with GPT-5.6-sol. We also use GLM-5.3-Flash to evaluate process-level consultation quality.

As shown in Table 1, MedPreceptor improves diagnostic performance across all five Doctor LLMs. For Doctor Acc., Baichuan-M2-32B improves from 50.45% to 60.72%, Doctor-R1 from 46.00% to 54.79%, and GPT-5.6-sol from 71.97% to 74.62%. GPT-5.6 Acc. shows the same trend, with DeepSeek-v4.1-flash improving from 69.79% to 71.97% and Doctor-R1 from 67.48% to 69.52%. MedPreceptor also improves information gathering. Coverage increases from 0.313 to 0.362 for Baichuan-M2-32B and from 0.365 to 0.395 for GLM-5.3-Flash, while Relevance improves across all five Doctors. The remaining quality metrics are largely stable, with only minor fluctuations. Overall, MedPreceptor improves diagnostic performance and information acquisition across different Doctor LLMs while largely preserving consultation quality.

4.2.2 MedDialogRubrics Consultation Performance

We further evaluate MedPreceptor on MedDialogRubrics (Gong et al., 2026), a benchmark for multi-turn medical consultation that provides case-specific, expert-refined "must-ask" rubrics. These rubrics characterize clinically important information that should be elicited during each consultation, making the benchmark particularly suitable for evaluating the information-acquisition process. We report **Coverage** by first computing, for each case, the fraction of must-ask rubrics satisfied in the complete consultation and then averaging this value across cases. We additionally report the average number of consultation turns (**Avg. Turns**). To examine rubric completion relative to con-

Table 2: Results on MedDialogRubrics.

Doctor	Preceptor	Cov.	Turns	Rub./Turn	Doctor	Preceptor	Cov.	Turns	Rub./Turn
Baichuan-M2-32B	–	32.65	6.94	0.633	Gemini-3.8-Flash	–	56.09	3.51	2.165
	✓	**39.73**	7.30	**0.733**		✓	**60.69**	3.73	**2.205**
Doctor-R1	–	52.65	10.56	0.671	GPT-5.6-Sol	–	51.88	5.69	**1.234**
	✓	**58.80**	10.87	**0.728**		✓	**56.12**	7.38	1.029
DeepSeek-V4.1-Flash	–	58.93	10.33	0.769	Grok-4.3	–	40.01	2.82	1.915
	✓	**63.96**	10.14	**0.852**		✓	**44.44**	3.07	**1.956**
GLM-5.3-Flash	–	58.05	6.62	1.185					
	✓	**62.82**	6.81	**1.245**					

Table 3: Results on the Clinically Annotated Consultation Intervention Benchmark (CACI). Intervention decision metrics are evaluated on 1,664 consultation states, while corrective advice quality is evaluated on the 720 intervention-required states. Precision, recall, and F1 treat intervention-required states as the positive class.

	Intervention Decision				Corrective Advice Quality					
Model	**Acc.**	**Prec.**	**Recall**	**F1**	**Score**	**Align.**	**Act.**	**Ground.**	**Eff.**	**Style**
MedPreceptor	**66.11**	57.60	82.08	**67.70**	7.710	**1.469**	**1.854**	1.499	1.085	1.803
Baichuan-M2-32B	45.79	43.78	88.89	58.66	7.107	1.078	1.758	1.500	0.853	1.918
Doctor-R1	43.21	43.22	**99.58**	60.28	7.015	0.969	1.658	**1.575**	0.883	**1.929**
Gemma-4-31B-IT	61.96	54.81	68.89	61.05	5.999	1.044	1.462	0.974	0.917	1.601
DeepSeek-V4.1-Flash	62.86	**57.61**	53.61	55.54	6.618	1.269	1.671	1.143	0.851	1.683
GPT-5.6-Sol	62.80	56.80	58.61	57.69	**7.843**	1.438	1.779	1.428	**1.331**	1.868

sultation length, we also report **Rubrics/Turn**, computed as the number of satisfied rubrics in the final trajectory divided by the number of consultation turns.

As shown in Table 2, we evaluate MedPreceptor across seven Doctor LLMs covering both open- and closed-source models with different scales and capabilities. MedPreceptor improves must-ask rubric Coverage for all models, including Baichuan-M2-32B from 32.65% to 39.73%, Doctor-R1 from 52.65% to 58.80%, and DeepSeek-V4.1-Flash from 58.93% to 63.96%. Rubrics/Turn also improves for most models, for example from 0.769 to 0.852 for DeepSeek-V4.1-Flash and from 1.185 to 1.245 for GLM-5.3-Flash. Overall, MedPreceptor improves the acquisition of clinically important information across diverse Doctor LLMs.

4.2.3 Clinically Annotated Consultation Intervention Evalution

Based on the CACI evaluation set, we assess two core capabilities: **when to intervene** and **how to intervene**. For intervention decisions, each model is given the consultation prefix available at the current state, and its prediction is compared with the human annotation. We report accuracy, precision, recall, and F1, treating intervention-required states as the positive class. For corrective advice, we evaluate the 720 states annotated as requiring intervention. GPT-5.6-sol scores the quality of the generated advice using the visible case information, consultation prefix, human reference advice, and model-generated advice, without re-evaluating whether intervention is needed. The evaluation covers five dimensions: Alignment, Actionability, Grounding, Efficiency, and Style & Safety. Each dimension is scored from 0 to 2, for a total score of 10; detailed scoring criteria are provided in Table 10.

As shown in Table 3, MedPreceptor achieves the highest intervention-decision accuracy at 66.11%, with 57.60% precision, 82.08% recall, and 67.70% F1 for intervention-required states. The relatively high recall indicates that MedPreceptor detects most states requiring intervention, while the lower precision shows that unnecessary interventions remain a limitation. For corrective advice, it obtains a score of 7.710, second only to GPT-5.6-Sol, and achieves the best Alignment and Actionability scores. Overall, MedPreceptor performs strongly on both when to intervene and how to intervene.

Table 4: Ablation study of MedPreceptor training strategies on MAQuE. Base and Teacher denote Qwen3.5-4B Base and Qwen3.5-27B SFT, respectively.

Method	Doctor Acc.	GPT Acc.	Cov.	Rel.	Adh.	Coh.	Cla.	Emp.	Q.
No Preceptor	50.45	66.34	0.3130	0.8790	0.8460	0.5760	0.6820	0.4040	14.85
Base	52.55	64.07	0.3313	0.8983	0.8588	0.6119	0.6766	0.4093	15.02
SFT	57.52	66.83	0.3546	0.8948	0.8687	0.6371	0.6821	0.4097	14.71
OPD	55.97	65.93	0.3502	0.8928	0.8598	0.6279	**0.6863**	0.4087	14.85
DG-OPD	**60.72**	**68.24**	**0.3618**	**0.9008**	**0.8752**	**0.6503**	0.6817	0.4094	14.72
Teacher	60.34	66.55	0.3486	0.8918	0.8722	0.6385	0.6837	**0.4100**	**14.64**

Table 5: Ablation results on the Clinically Annotated Consultation Intervention Benchmark (CACI). Precision, Recall, and F1 are computed by treating intervention-required states as the positive class. Score is out of 10; component scores are out of 2.

Method	Acc.	Precision	Recall	F1	Score	Align.	Act.	Ground.	Eff.	Style
DG-OPD	**66.11**	**57.60**	82.08	**67.70**	**7.710**	**1.469**	**1.854**	**1.499**	1.085	1.803
Base	43.27	43.27	**100.00**	60.40	5.757	1.001	1.501	0.915	0.753	1.586
SFT	58.77	51.66	73.47	60.66	7.392	1.411	1.696	1.283	**1.213**	1.789
OPD	59.44	54.62	36.94	44.07	6.844	1.089	1.736	1.485	0.608	**1.926**
Teacher	61.84	54.98	65.14	59.63	7.304	1.386	1.796	1.369	0.989	1.764

4.3 Ablation Studies

On MAQuE, we fix Baichuan-M2-32B as the Doctor model to compare different Preceptor training strategies. To isolate the effect of decision-guided negative-signal truncation, standard OPD and DG-OPD use identical supervision information, SFT initialization, on-policy data, and training budget, and differ only in whether positive Teacher signals are removed after an incorrect intervention decision; full configuration details are provided in Appendix A.2. Compared with No Preceptor, the Base model already improves several consultation-quality metrics, although its diagnostic performance is less stable. After SFT, Doctor Acc. increases from 52.55% to 57.52%, and Coverage improves from 0.3313 to 0.3546, showing that supervised initialization provides a stronger starting point. Standard OPD does not bring further gains, with both Doctor Acc. and GPT Acc. decreasing slightly. In contrast, DG-OPD achieves the best Doctor Acc. (60.72%) and GPT Acc. (68.24%), together with the best Coverage, Relevance, Adherence, and Coherence, while keeping the average number of consultation questions nearly unchanged. Results on CACI further evaluate the intervention capability of the Preceptor itself. SFT clearly improves both intervention decisions and corrective advice quality over the Base model. Standard OPD slightly improves Accuracy and Precision over SFT, but its Recall drops to 36.94%, leading to lower F1 and advice quality. DG-OPD achieves the best overall performance, reaching 66.11% Accuracy, 57.60% Precision, 82.08% Recall, and 67.70% F1, together with the highest overall advice quality score of 7.710. It also performs best on Alignment, Actionability, and Grounding, and overall outperforms the SFT-trained teacher.

5 Conclusion

In this work, we present MedPreceptor, a framework providing decoupled process supervision for multi-turn medical consultations. Operating without modifying the underlying Doctor LLM, MedPreceptor monitors the consultation trajectory and selectively intervenes only when clinical inquiry falters. To eliminate hindsight information leakage, we introduce a two-stage prefix-aligned annotation scheme and develop DG-OPD to jointly optimize intervention timing and actionable advice generation. In addition, we construct the expert-annotated CACI benchmark to evaluate intervention decisions and guidance quality. Evaluation shows that MedPreceptor reliably identifies consultation pitfalls, enhances clinical evidence acquisition, and improves diagnostic outcomes across diverse Doctor LLM architectures. These results demonstrate that decoupling process supervision from consultation execution provides an effective, modular, and reusable paradigm for advancing and supporting active clinical inquiry for interactive medical AI and potentially human clinicians.

AI USE STATEMENT

In this work, we used generative AI tools to assist with code development, manuscript writing, and dataset construction. In particular, generative AI was used to support data preprocessing, consultation trajectory generation, annotation assistance, and language polishing. All AI-assisted outputs, including generated data, annotations, code, and manuscript text, were reviewed and verified by the authors before use. The authors take full responsibility for the final content of this work, including all text, claims, experimental results, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies AI-assisted multi-turn medical consultation and is intended to improve information gathering during clinical interactions rather than replace professional medical judgment. Our experiments use de-identified clinical data and publicly available medical consultation benchmarks. The CACI benchmark is manually annotated by clinical medicine graduate students and does not involve direct interaction with patients. We take particular care to avoid exposing identifiable patient information during data processing and model evaluation. MedPreceptor provides process-level guidance to the Doctor model and does not directly provide diagnoses or treatment recommendations to patients. Nevertheless, errors such as unsupported assumptions or inappropriate intervention may still occur, and the system should therefore be considered a research prototype rather than a tool for unsupervised clinical deployment. CACI will be publicly released upon acceptance to support future research while following appropriate data privacy and release requirements.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the data construction, annotation procedure, model configuration, DG-OPD training algorithm, evaluation protocols, and hyperparameters in the main paper and appendix. Detailed training settings are provided in Appendix A. We also provide the prompts used for consultation generation, intervention annotation, and evaluation, together with representative CACI examples. The source code will be publicly released, and CACI will be released upon acceptance to facilitate reproduction and future research.

REFERENCES

Daniel Angus, Bernadette Watson, Andrew Smith, Cindy Gallois, and Janet Wiles. Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PLoS ONE*, 7(6):e38014, 2012. doi: 10.1371/journal.pone.0038014.

Baichuan-M2 Team. Baichuan-M2: Scaling medical capability with large verifier system. *arXiv preprint arXiv:2509.02208*, 2025.

Howard B. Beckman and Richard M. Frankel. The effect of physician behavior on the collection of data. *Annals of Internal Medicine*, 101(5):692–696, 1984. doi: 10.7326/0003-4819-101-5-692.

DeepSeek-AI. Deepseek-v4.1-flash: Pushing the limits of kv cache compression, 2026. URL https://arxiv.org/abs/2609.19969.

Jinrui Fang, Runhan Chen, Xu Yang, Jian Yu, Jiawei Xu, Ashwin Vinod, Wenqi Shi, Tianlong Chen, Heng Ji, ChengXiang Zhai, Ying Ding, and Yuji Zhang. Benchmarking multi-turn medical diagnosis: Hold, lure, and self-correction. *arXiv preprint arXiv:2604.04325*, 2026.

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

Yuqian Fu, Haohuan Huang, Kaiwen Jiang, Jiacai Liu, Zhuo Jiang, Yuanheng Zhu, and Dongbin Zhao. Revisiting on-policy distillation: Empirical failure modes and simple fixes. *arXiv preprint arXiv:2603.25562*, 2026.

Siyuan Gan, Yuhan Li, Xiran Wang, Linjian Meng, Boyan Wang, Zhen Zhao, Jing Huo, and Yang Gao. When teacher guidance misleads: Reward-aligned on-policy distillation. *arXiv preprint arXiv:2608.27960*, 2026.

GLM-5 Team. Glm-5: From vibe coding to agentic engineering, 2026. URL https://arxiv.org/abs/2602.15763.

Lecheng Gong, Weimin Fang, Ting Yang, Dongjie Tao, Chunxiao Guo, Peng Wei, Bo Xie, Jinqun Guan, Zixiao Chen, Fang Shi, Jinjie Gu, and Junwei Liu. MedDialogRubrics: A comprehensive benchmark and evaluation framework for multi-turn medical consultations in large language models. *arXiv preprint arXiv:2601.03023*, 2026.

Linlu Gong, Ante Wang, Yunghwei Lai, Weizhi Ma, and Yang Liu. The dialogue that heals: A comprehensive evaluation of doctor agents' inquiry capability. *arXiv preprint arXiv:2509.24958*, 2025.

Google. Gemini 3.8 flash. https://ai.google.dev/gemini-api/docs/models/gemini-3.8-flash, 2026. Google AI for Developers.

Hsin-Ling Hsu, Zizheng Wang, Donghua Zhang, Nai-Chia Chen, Jerry Wang, Jun-En Ding, Chia-Hsuan Hsu, Guoan Wang, Feng Liu, Fang-Ming Hung, Chenwei Wu, and Liyue Shen. MedAction: Towards active multi-turn clinical diagnostic LLMs. *arXiv preprint arXiv:2605.07305*, 2026.

Yuxin Hu, Danni Liu, Bo Liu, Yida Chen, Jiuxin Cao, and Yan Liu. PsyAdvisor: A plug-and-play strategy advice planner with proactive questioning in psychological conversations. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12205–12229, 2025. doi: 10.18653/v1/2025.acl-long.596.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Leo Anthony Celi, Roger Mark, and Steven Horng. MIMIC-IV-ED. *PhysioNet*, 2023a. doi: 10.13026/5ntk-km72. URL https://doi.org/10.13026/5ntk-km72. Version 2.2.

Alistair Johnson, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV-Note: Deidentified free-text clinical notes. *PhysioNet*, 2023b. doi: 10.13026/1n74-ne17. URL https://doi.org/10.13026/1n74-ne17. Version 2.2.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Brian Gow, Benjamin Moody, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV. *PhysioNet*, 2024. doi: 10.13026/kpb9-mt58. URL https://doi.org/10.13026/kpb9-mt58. Version 3.1.

Shreya Johri, Jaehwan Jeong, Benjamin A. Tran, Daniel I. Schlessinger, Shannon Wongvibulsin, Leandra A. Barnes, Hong-Yu Zhou, Zhuo Ran Cai, Eliezer M. Van Allen, David Kim, Roxana Daneshjou, and Pranav Rajpurkar. An evaluation framework for clinical use of large language models in patient interaction tasks. *Nature Medicine*, 31:77–86, 2025. doi: 10.1038/s41591-024-03328-5.

Yunghwei Lai, Kaiming Liu, Ziyue Wang, Weizhi Ma, and Yang Liu. Doctor-R1: Mastering clinical inquiry with experiential agentic reinforcement learning. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=vQGHTyL0Jw.

Tian Lan, Wenwei Zhang, Chengqi Lyu, Shuaibin Li, Chen Xu, Heyan Huang, Dahua Lin, Xian-Ling Mao, and Kai Chen. Training language models to critique with multi-agent feedback. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 1474–1501, 2025. doi: 10.18653/v1/2025.findings-emnlp.78.

Zicheng Lin, Zhibin Gou, Tian Liang, Ruilin Luo, Haowei Liu, and Yujiu Yang. CriticBench: Benchmarking LLMs for critique-correct reasoning. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 1552–1587, 2024. doi: 10.18653/v1/2024.findings-acl.91.

OpenAI. Gpt-5.6: Frontier intelligence that scales with your ambition. https://openai.com/index/gpt-5-6/, July 2026. Accessed: 2026-09-22.

Chuhan Qiao, Jianghua Huang, Daxing Zhao, Ziding Liu, Yanjun Shen, Bing Cheng, Wei Lin, and Kai Wu. MedConsultBench: A full-cycle, fine-grained, process-aware benchmark for medical consultation agents. *arXiv preprint arXiv:2601.12661*, 2026.

Samuel Schmidgall, Rojin Ziaei, Carl Harris, Ji Woong Kim, Eduardo Pontes Reis, Jeffrey Jopling, and Michael Moor. AgentClinic: A multimodal benchmark for tool-using clinical AI agents. *npj Digital Medicine*, 2026. doi: 10.1038/s41746-026-02674-7.

Hardeep Singh, Traber Davis Giardina, Ashley N. D. Meyer, Samuel N. Forjuoh, Michael D. Reis, and Eric J. Thomas. Types and origins of diagnostic errors in primary care settings. *JAMA Internal Medicine*, 173(6):418–425, 2013. doi: 10.1001/jamainternmed.2013.2777.

Tao Tu, Mike Schaekermann, Anil Palepu, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Yong Cheng, Elahe Vedadi, Nenad Tomasev, Shekoofeh Azizi, Karan Singhal, Le Hou, Albert Webson, Kavita Kulkarni, S. Sara Mahdavi, Christopher Semturs, Juraj Gottweis, Joelle Barral, Katherine Chou, Greg S. Corrado, Yossi Matias, Alan Karthikesalingam, and Vivek Natarajan. Towards conversational diagnostic artificial intelligence. *Nature*, 642:442–450, 2025. doi: 10.1038/s41586-025-08866-7.

Xiaoyu Wang, Yue Zhao, Qingqing Gu, Zhonglin Jiang, Yong Chen, and Luo Ji. Convert language model into a value-based strategic planner. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pp. 1444–1456, 2025. doi: 10.18653/v1/2025.acl-industry.102.

Takashi Watari, Ashwin Gupta, Yu Amano, and Yasuharu Tokuda. Japanese internists' most memorable diagnostic error cases: A self-reflection survey. *Internal Medicine*, 63(2):221–229, 2024. doi: 10.2169/internalmedicine.1494-22.

xAI. Grok 4.3. `https://docs.x.ai/developers/models/grok-4.3`, 2026. xAI Model Documentation.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Tianshu Yu, Chao Xiang, Mingchuan Yang, Pei Ke, Bosi Wen, Cunxiang Wang, Jiale Cheng, Li Zhang, Xinyu Mu, Chuxiong Sun, and Minlie Huang. Training language model to critique for better refinement. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 26760–26804, 2025. doi: 10.18653/v1/2025.findings-acl.1373.

Guangtao Zeng, Wenmian Yang, Zeqian Ju, Yue Yang, Sicheng Wang, Ruisi Zhang, Meng Zhou, Jiaqi Zeng, Xiangyu Dong, Ruoyu Zhang, Hongchao Fang, Penghui Zhu, Shu Chen, and Pengtao Xie. MedDialog: Large-scale medical dialogue datasets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 9241–9250. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.743. URL `https://aclanthology.org/2020.emnlp-main.743/`.

A Training Details

A.1 Training Case Construction

Case cohort selection. We construct the training cohort from MIMIC-IV, MIMIC-IV-ED, and MIMIC-IV-Note. Emergency department encounters are linked with admission records and discharge summaries to obtain patient demographics, chief complaints, diagnoses, vital signs, medication records, history of present illness (HPI), and past medical history (PMH). We first identify

Table 6: Clinical-category distribution of the 15,000 cases used for training data construction.

Clinical Category	Cases	Proportion
Gastrointestinal	4,356	29.04%
Cardiovascular	3,687	24.58%
Respiratory	3,510	23.40%
Infectious	1,922	12.81%
General internal medicine	682	4.55%
Renal/urologic	484	3.23%
Hematologic/oncologic	213	1.42%
Endocrine/metabolic	60	0.40%
Rheumatologic/musculoskeletal	54	0.36%
Dermatologic/soft-tissue	32	0.21%
Total	**15,000**	**100.00%**

31,991 candidate cases from broad internal-medicine categories, including cardiovascular, respiratory, infectious, gastrointestinal, renal/urologic, endocrine/metabolic, rheumatologic/musculoskeletal, hematologic/oncologic, dermatologic/soft-tissue, and general internal medicine cases. Cases dominated by neurological, trauma/orthopedic, obstetric/gynecologic, psychiatric/substance-use, ophthalmologic/ENT, or postoperative surgical conditions are excluded.

To ensure sufficient information for multi-turn consultation, we retain only admitted cases with a known arrival mode, a valid discharge diagnosis, at least 100 words in the HPI, at least 8 words in the PMH, and sufficient internal-medicine evidence in the clinical record, resulting in 20,889 high-quality candidates. From these candidates, we select 10,000 cases using category-aware sampling to construct the SFT training data. Cases are ranked within each category according to record completeness and clinical information richness, while category quotas are allocated based on the square root of the number of eligible cases to preserve representation of smaller categories. We additionally select a separate set of 5,000 cases for generating the on-policy training data used in DG-OPD.

Separation from CACI and de-duplication. We construct the training data and CACI from disjoint source partitions before consultation simulation and annotation. We enforce separation at three levels. At the patient level, all records belonging to the same patient are assigned exclusively to either the training partition or CACI. At the case level, no encounter or clinical case is shared across the two partitions. At the dialogue level, consultation trajectories are de-duplicated within and across partitions, and no dialogue or consultation prefix used in CACI appears in either the SFT or on-policy training data. Consequently, CACI provides a held-out evaluation set with no patient-, case-, or dialogue-level overlap with training.

Case preprocessing and consultation trajectory generation. We first convert the selected MIMIC cases into a unified MAQuE-style format for multi-turn consultation simulation. Specifically, we use an LLM to extract and normalize clinically relevant information from the original medical records into structured patient profiles, including demographics, medical and family history, presenting symptoms, and available diagnostic findings. To prevent information leakage, objective test results are separated from subjective symptoms, and fields that may directly reveal the final diagnosis are excluded from the information visible during consultation.

Each structured patient profile is then converted into a MAQuE-style case containing patient information, atomized clinical facts that can be progressively revealed during the dialogue, a reference diagnosis, and five diagnostic choices. We further use an LLM to normalize the reference diagnosis and generate clinically plausible distractors, followed by automatic validation of the resulting choices. Finally, we use Doctor-R1, Doctor-Agent-RL, and Baichuan-M2-32B as Doctor models, with Qwen3-32B serving as the Patient Simulator, to generate multi-turn consultation trajectories over the selected cases. Each trajectory contains at most 16 consultation turns. The resulting consultation trajectories are subsequently used to construct the dataset.

Details about two-stage intervention annotation. We adopt a two-stage annotation procedure using GPT-5.5 as the annotator in both stages to construct intervention supervision while avoiding

hindsight leakage. In Stage A, GPT-5.5 reviews the complete consultation trajectory together with the final diagnosis and reference answer to identify candidate intervention turns where a locally observable and correctable consultation issue has already emerged. In Stage B, each candidate is re-evaluated by GPT-5.5 using only the information available up to that turn, including the visible case information and dialogue prefix. Future dialogue, the final diagnosis, the reference answer, and the Stage-A rationale are hidden. A candidate is retained as an `INTERVENE` state only if a concrete and actionable problem can be justified from the prefix alone; otherwise, it is relabeled as `NO_INTERVENE`.

Stage A: retrospective candidate localization. The Stage-A annotator is instructed as follows:

Stage A Annotation Prompt

You are a Stage-A retrospective annotator for medical consultation trajectories. You are given the complete dialogue, the final diagnosis, and the reference answer. Use them only to understand whether the consultation contains correctable process issues.
For each trajectory, determine its trainability and select at most {`max_candidate_turns`} candidate turns at which the Doctor could appropriately receive a reminder. Candidate turns should be ordered chronologically. Each identified issue must already be observable from the dialogue by the end of that turn and must be correctable by the Doctor in the next turn. Do not infer an issue solely from an incorrect final diagnosis. A trajectory with a correct final diagnosis may still contain trainable consultation issues.
Prioritize the following problems: missing key history of present illness, drifting away from the chief complaint, repeated or overly generic questioning, failure to ask important associated symptoms or discriminative negative findings, premature diagnostic closure or treatment recommendation, and failure to follow up on clues already provided by the patient. Questions that reasonably contribute to information gathering should not be marked as intervention points. For trajectories ending in an incorrect diagnosis, identify up to two early issues supported by locally available evidence whenever possible; do not invent issues when the evidence is insufficient.
The `issue_type` must be one of: `missing_key_information`, `misdirected_inquiry`, `low_value_questioning`, `repeated_questioning`, `inefficient_information_gathering`, `premature_closure`, `late_end_consultation`, `protocol_violation`, or `patient_inconsistent_wsth_case`. The `evidence` field should describe evidence already visible by that turn. The `suggested_advice` field should provide a brief next-step reminder to the Doctor. It must not reveal the final diagnosis, the correct option, future information, or treatment recommendations. All free-text fields should be written in Chinese.
The input is a JSON array in which each item contains an `item_id` and a `trajectory`. Return a strict JSON array with one output for every input `item_id`. Each output should contain: `trainability` (`trainable`, `partially_trainable`, or `not_trainable`), `non_trainable_reasons`, and `candidate_intervention_turns`. Each candidate turn should include the turn index, issue type, whether the issue is Doctor-correctable, supporting evidence, and suggested advice. If `trainability` is `not_trainable`, `candidate_intervention_turns` must be empty. Do not output any additional text.

Stage B: prefix-only intervention relabeling. Only Stage-A candidates assigned to one of the five target issue categories—*repeated questioning*, *inefficient inquiry*, *misdirected focus*, *omission of key clinical clues*, and *premature diagnostic closure*—are passed to Stage B. Each candidate is then re-evaluated using only the case information and dialogue available up to that turn, without access to future dialogue, the reference answer, the final diagnosis, or the Stage-A rationale.

Stage B Annotation Prompt

You are a Stage-B online advisor for a medical consultation. You only observe the dialogue up to and including the candidate turn and the case information visible at that time. You do not have access to future dialogue, the reference answer, the final diagnosis, or the retrospective explanation produced in Stage A.
Determine whether there is a clear and Doctor-correctable consultation issue at the current point. Output `NO_INTERVENE` for minor imperfections, questions that still provide useful information, or situations in which the available evidence is insufficient to support a specific recommendation. Output `INTERVENE` when the Doctor has repeated information, drifted away from the chief complaint, missed an important clue already visible in the current context, or prematurely moved to diagnosis or treatment, and when a concrete next-step recommendation can be provided.

The recommendation may only refer to facts already present in `case_visible_info` and `dialogue_prefix`. Do not use hidden symptoms, future test results, the reference answer, diagnostic choices, or treatment recommendations. State already visible facts directly, and suggest further questioning only for information that remains unknown. The `evidence` and `suggested_advice` fields should be concise and written in Chinese. The input is a JSON array in which each item contains `item_id`, `candidate_turn`, `case_visible_info`, and `dialogue_prefix`. The output must be a strict JSON array, with each item following the format:

```
{
  "item_id": "original item_id",
  "annotation": {
    "decision": "INTERVENE or NO_INTERVENE",
    "doctor_correctable": true,
    "confidence": 0.8,
    "evidence": "Chinese evidence",
    "suggested_advice": "Chinese recommendation"
  }
}
```

For `NO_INTERVENE`, set `doctor_correctable=false` and `suggested_advice=""`. Every input `item_id` must appear exactly once in the output. Do not output any additional text.

Annotation example. Table 7 presents a representative example of our two-stage intervention annotation process. Stage A first identifies Turn 5 as a candidate intervention point based on the complete consultation trajectory. Stage B then re-evaluates this candidate using only the case information and dialogue prefix available up to that turn. In this example, the Doctor shifts to urinary symptoms despite unresolved anemia-related clues already present in the consultation, and Stage B therefore confirms the candidate as an `INTERVENE` state.

A.2 Model Configuration and Training Framework

Our training pipeline is implemented with PyTorch, verl, and FSDP, with SGLang used for asynchronous generation during on-policy rollout. Training is conducted on NVIDIA RTX A6000 GPUs with 48 GB memory per GPU. We use Qwen3.5-4B as the Student Advisor and Qwen3.5-27B as the Teacher. Both models are first supervised fine-tuned on the constructed Advisor dataset using LoRA with rank 32 and alpha 16 over all linear layers, where the target output is a structured tool call containing the intervention decision and corresponding advice-related fields. After SFT, the LoRA adapters are merged into the corresponding base models. The Qwen3.5-4B checkpoint is then further optimized using our DG-OPD, while the Qwen3.5-27B checkpoint remains frozen as the Teacher. For SFT, we use AdamW with a learning rate of 1×10^{-5} and a maximum sequence length of 2,048 tokens for three epochs, computing cross-entropy loss only over the target Advisor response. DG-OPD is initialized from the SFT-trained Qwen3.5-4B checkpoint and trained for five epochs with a learning rate of 5×10^{-7}. One on-policy rollout is generated for each input, with a maximum prompt length of 4,096 tokens and a maximum response length of 250 tokens. DG-OPD is applied to all semantic output fields, including the intervention decision, supporting evidence, and corrective advice. For samples with a correct intervention decision, we retain the full token-level Teacher signal; for samples with an incorrect decision, we retain only its negative part. We use bfloat16 precision, gradient checkpointing, and FSDP sharding to reduce memory consumption. GPT-5.5 is used only for Stage A/B data annotation.

For the standard OPD baseline, we use exactly the same SFT initialization, on-policy training examples, frozen Teacher, Teacher-side conditioning information, semantic-token mask, rollout count, sequence-length limits, optimizer, learning rate, and five-epoch training budget as DG-OPD. The two methods differ only in how the token-level Teacher signal is processed: standard OPD retains $\hat{A}_{t,j}$ for every semantic token regardless of decision correctness, whereas DG-OPD replaces it with $\widetilde{A}_{t,j}$ and applies the negative-only truncation $\min(\hat{A}_{t,j}, 0)$ when the intervention decision is incorrect. Thus, the OPD–DG-OPD comparison isolates the contribution of decision-guided negative-signal truncation.

Table 7: Example of two-stage intervention annotation. The candidate intervention point is identified retrospectively in Stage A and re-evaluated in Stage B using only information available up to Turn 5.

Field	Content
Visible case information	
Chief complaint	Progressively worsening fatigue for one week. The patient feels normal at rest but becomes easily fatigued when walking and reports that the symptoms feel similar to a previous episode of anemia.
Relevant history	65-year-old male with a history of urinary tract abnormalities, bilateral hydronephrosis with nephrostomy tubes, acute renal failure, chronic urinary retention, and prior abdominal surgery.
Dialogue prefix	
Turn 1	**Doctor:** Did your fatigue start suddenly, or has it gradually worsened? **Patient:** It has gradually worsened.
Turn 2	**Doctor:** How long have you been experiencing this fatigue? **Patient:** About one week.
Turn 3	**Doctor:** Have you noticed any particular discomfort while walking, such as numbness, swelling, or pain in your legs? **Patient:** I become easily fatigued when walking.
Turn 4	**Doctor:** Have you experienced dizziness, palpitations, or shortness of breath? **Patient:** I have experienced palpitations.
Turn 5	**Doctor:** Have you experienced urinary frequency, urgency, or painful urination? **Patient:** No.
Intervention annotation	
Stage-A category	**Missing key information**
Stage-B decision	`INTERVENE` (confidence: 0.74)
Evidence	The patient presents with exertional fatigue, explicitly reports that the symptoms resemble a previous episode of anemia, and has already reported palpitations. At Turn 5, the Doctor shifts to urinary symptoms before adequately exploring anemia-related symptoms or potential sources of bleeding.
Suggested advice	The current presentation resembles the patient's previous anemia and is accompanied by palpitations. Further assess anemia- and bleeding-related clues, including dyspnea, dizziness, melena or hematochezia, hematuria, recent bleeding, dietary history, and the cause of prior anemia.

A.3 Detailed Training Procedure of DG-OPD

For completeness, Algorithm 1 provides the detailed training procedure of DG-OPD. At each iteration, the current Student policy first generates on-policy Advisor outputs. The frozen Teacher then scores the Student-generated semantic tokens conditioned on the consultation context, the gold intervention decision, and the supporting evidence, without access to the reference advice. The semantic tokens include the intervention decision, supporting evidence, and corrective advice. The resulting token-level distillation signal is selectively modified according to whether the Student makes the correct intervention decision: full Teacher guidance is retained for correct decisions, whereas only negative signals are preserved for incorrect ones. The modified signal is then used in the standard clipped OPD objective.

B Evaluation Details

B.1 MAQuE

MAQuE is a benchmark for evaluating the overall quality of multi-turn medical consultations (Gong et al., 2025). It considers not only whether the final diagnosis is correct, but also whether the Doctor effectively collects clinically relevant information and maintains good dialogue quality and patient-facing communication. In this work, we use MAQuE to evaluate whether MedPreceptor improves the end-to-end consultation performance of different Doctor LLMs, including information coverage, question relevance, dialogue quality, and communication with the patient. **Evaluation setting.** We

Algorithm 1 Decision-Guided Selective On-Policy Distillation

Require: Teacher policy π_{te}, Student policy π_θ
Require: Training dataset $\mathcal{D} = \{(\mathcal{H}_t, d_t^\star, e_t^\star)\}$
Require: OPD clip coefficient ϵ, learning rate η

for each training iteration **do**
  Set behavior policy $\pi_{\theta_{\mathrm{old}}} \leftarrow \pi_\theta$
  Sample a prompt batch $\mathcal{B} \subset \mathcal{D}$
  Initialize rollout buffer $\mathcal{R} \leftarrow \emptyset$
  for each $(\mathcal{H}_t, d_t^\star, e_t^\star) \in \mathcal{B}$ **do**
    Sample an on-policy Advisor output

$$\hat{\mathbf{s}}_t = [\hat{d}_t; \hat{e}_t; \hat{a}_t] \sim \pi_{\theta_{\mathrm{old}}}(\cdot \mid \mathcal{H}_t)$$

    if $\hat{\mathbf{s}}_t$ cannot be parsed as a valid structured output **then**
      continue
    end if
    Let $\mathcal{J}_t$ denote token positions of the semantic fields (intervention decision, supporting evidence, and corrective advice)
    for each $j \in \mathcal{J}_t$ **do**
      Compute standard OPD token signal

$$\hat{A}_{t,j} = \log \pi_{\mathrm{te}}\left(\hat{s}_{t,j} \mid \mathcal{H}_t, d_t^\star, e_t^\star, \hat{s}_{t,<j}\right) - \log \pi_{\theta_{\mathrm{old}}}\left(\hat{s}_{t,j} \mid \mathcal{H}_t, \hat{s}_{t,<j}\right)$$

      if $\hat{d}_t = d_t^\star$ **then**
        $\widetilde{A}_{t,j} \leftarrow \hat{A}_{t,j}$
      else
        $\widetilde{A}_{t,j} \leftarrow \min(\hat{A}_{t,j}, 0)$
      end if
    end for
    Store

$$(\mathcal{H}_t, \hat{\mathbf{s}}_t, d_t^\star, \{\widetilde{A}_{t,j}\}_{j \in \mathcal{J}_t}) \rightarrow \mathcal{R}$$

  end for
  for each mini-batch gradient step **do**
    Sample a mini-batch $\mathcal{B}_{\mathrm{mini}} \subset \mathcal{R}$
    Compute the importance ratio for each semantic token $j \in \mathcal{J}_t$:

$$r_{t,j}(\theta) = \frac{\pi_\theta(\hat{s}_{t,j} \mid \mathcal{H}_t, \hat{s}_{t,<j})}{\pi_{\theta_{\mathrm{old}}}(\hat{s}_{t,j} \mid \mathcal{H}_t, \hat{s}_{t,<j})}.$$

    Compute the clipped DG-OPD loss:

$$\mathcal{L}_{\text{DG-OPD}}(\theta) = -\frac{1}{|\mathcal{B}_{\mathrm{mini}}|} \sum_{t \in \mathcal{B}_{\mathrm{mini}}} \frac{1}{|\mathcal{J}_t|} \sum_{j \in \mathcal{J}_t} \min\left(r_{t,j}(\theta)\widetilde{A}_{t,j}, \mathrm{clip}(r_{t,j}(\theta), 1-\epsilon, 1+\epsilon)\widetilde{A}_{t,j}\right).$$

    Update $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{\text{DG-OPD}}(\theta)$
  end for
end for

follow the official MAQuE evaluation protocol with incremental diagnosis disabled. For **GPT-5.6 Acc.**, the complete Doctor–Patient dialogue, the diagnostic question, and five candidate diagnoses are provided to an independent GPT-5.6-Sol model, which is required to select a single diagnosis. The prediction is then matched against the ground-truth answer. This metric reflects whether the information collected during the consultation is sufficient to support the correct diagnosis. We additionally report **Doctor Acc.**. After the consultation ends, the original Doctor model makes a final diagnosis based on its own consultation trajectory. This complements GPT-5.6 Acc., since a trajectory may contain sufficient diagnostic evidence while the Doctor itself may still fail to correctly integrate or use that information. Reporting both metrics allows us to examine whether MedPrecep-

tor improves both the quality of information gathered during consultation and the Doctor's own final diagnostic performance.

Consultation quality is evaluated using GLM-5.3-Flash. **Coverage** measures how well the collected information covers the atomized clinical facts of the case, while **Relevance** measures whether the Doctor's questions focus on clinically relevant information. **Adherence** and **Coherence** are scored over the complete consultation trajectory, whereas **Clarity** and **Empathy** are evaluated based on the Doctor's questions. The original scores for these dimensions range from 1 to 5 and are normalized to $[0, 1]$. We apply the same evaluation protocol to Baichuan-M2-32B (Baichuan-M2 Team, 2025), Doctor-R1 (Lai et al., 2026), DeepSeek-V4.1-Flash (DeepSeek-AI, 2026), GPT-5.6-Sol (OpenAI, 2026), and GLM-5.3-Flash GLM-5 Team (2026).

B.2 MEDDIALOGRUBRICS

MedDialogRubrics is a benchmark designed to evaluate **information-gathering ability** in multi-turn medical consultations. It contains 4,700 cases and 63,828 case-specific inquiry rubrics across 21 clinical specialties. Each case is associated with a set of key clinical information that should be elicited during consultation, such as symptom characteristics, associated symptoms, medical history, and other information relevant to differential diagnosis. Unlike MAQuE, MedDialogRubrics does not evaluate final diagnostic accuracy. Instead, it directly measures how well a Doctor covers important case-specific information within a limited number of consultation turns. We therefore use this benchmark to evaluate whether MedPreceptor can reduce missed clinical information and improve the efficiency of information gathering.

We evaluate seven Doctor LLMs: Baichuan-M2-32B (Baichuan-M2 Team, 2025), Doctor-R1 (Lai et al., 2026), DeepSeek-V4.1-Flash (DeepSeek-AI, 2026), GPT-5.6-Sol (OpenAI, 2026), GLM-5.3-Flash GLM-5 Team (2026), Gemini-3.8-Flash (Google, 2026), and Grok-4.3 (xAI, 2026).. The patient is simulated by Qwen3-32B, which has access to the full case information and answers the Doctor's questions accordingly. Each consultation is limited to at most 12 turns. When MedPreceptor is used, it first decides whether to intervene after the first three consultation turns and then makes another decision every three turns. If it outputs `INTERVENE`, its suggestion is provided to the Doctor in the next turn.

After the consultation, GLM-5.3-Flash is used as the rubric judge. Each gold rubric is independently classified as `Satisfied` or `Not Satisfied`. A rubric is counted as satisfied only when the relevant information is actively elicited by the Doctor or directly revealed in response to the Doctor's question. We report four metrics. **WRS / Macro Coverage** first computes the rubric coverage for each case and then averages it across cases. **Micro Rubric Coverage** measures the fraction of satisfied rubrics over all rubrics across the full evaluation set. **Rubrics per Doctor Turn** measures the number of satisfied rubrics obtained per consultation turn and reflects information-gathering efficiency. **Paired Better / Tie / Worse** compares rubric coverage on the same cases with and without MedPreceptor and reports the proportion of cases that improve, remain unchanged, or decrease.

B.3 CLINICALLY ANNOTATED INTERVENTION EVALUATION

We further construct a **Clinically Annotated Consultation Intervention Benchmark** to directly assess whether MedPreceptor can correctly determine **when to intervene** and **what advice to provide**. The cases and multi-turn consultation trajectories are generated using the same case preprocessing and consultation simulation procedure described in Section A.1, but exclusively from the held-out CACI partition. Thus, "the same procedure" refers to the construction pipeline rather than reuse of any training patient, case, or dialogue. Unlike the training data, the intervention annotations in this benchmark are provided by **three medical graduate students** rather than generated by an LLM. Each annotator first reviews every consultation state independently, using only the case information and dialogue prefix visible at that point, and assigns an `INTERVENE` or `NO_INTERVENE` label. For an `INTERVENE` label, the annotator also identifies the concrete process-level issue and provides supporting evidence and corrective advice. After independent annotation, states with disagreement on the intervention decision or the underlying rationale are jointly re-examined using the same prefix-visible information and reconciled to a consensus label and reference rationale.

Table 8: A human-annotated INTERVENE example.

Field	Content
Specialty	Respiratory Medicine
Chief complaint	Cough, sputum production, and sore throat for 8 days, with worsening dyspnea on exertion. The visible medical history includes chronic obstructive pulmonary disease (COPD).
Dialogue prefix	**Doctor (Turn 1):** When did the cough and sputum production begin? What color and consistency is the sputum? **Patient:** The cough and sputum production started about 8 days ago. I am not really sure about the color or consistency of the sputum. **Doctor (Turn 2):** Before this episode, did you have any cold symptoms, get chilled, or get caught in the rain? Have you had a fever? **Patient:** There was no obvious cold, chilling, or rain exposure before this started. My temperature at triage was 97.3, so I did not have a fever. **Doctor (Turn 3):** You mentioned that your shortness of breath gets worse with activity. Do you feel short of breath when walking on level ground or climbing stairs? Are you also short of breath at rest? **Patient:** I do feel short of breath when walking on level ground or climbing stairs. I am not very short of breath at rest. **Doctor (Turn 4):** When you cough in the morning or during the day, have you noticed whether the sputum is yellow-green, white and foamy, or blood-streaked? **Patient:** I am not really sure about the color of the sputum.
Human label	`INTERVENE` at Turn 4
Human supporting evidence	The Doctor repeatedly asks about sputum characteristics even though the patient has already stated that this information is unavailable. More informative questions remain unexplored, particularly whether the current dyspnea, cough, and sputum production have worsened from the patient's usual COPD baseline and whether there have been recent changes in treatment.
Human advice	Given the visible history of COPD, focus next on whether the current dyspnea, cough, and sputum production are clearly worse than usual, and ask about regular inhaler or oxygen use and recent medications.

Because clinical inquiry can follow multiple valid paths, annotators are instructed not to treat a difference from their preferred question order, diagnostic hypothesis, or inquiry strategy as a process error. An `INTERVENE` label requires a specific, locally observable, and correctable problem—such as repeated questioning, failure to follow up an already revealed clue, or omission of time-sensitive key information—that is likely to impede subsequent information gathering. Alternative but clinically reasonable questions are labeled `NO_INTERVENE`. Annotation consistency is therefore assessed at the level of the binary intervention decision and the presence of an identifiable process-level issue, rather than exact agreement on the wording or ordering of the suggested advice. The final reference advice represents one acceptable correction, not a unique gold consultation path.

Clinically annotation examples. Tables 8 and 9 present representative examples from the Human-Annotated Consultation Intervention Benchmark, illustrating INTERVENE and NO-INTERVENE cases, respectively.

Evaluation setting. The Clinically annotated evaluation consists of two parts. We evaluate MedPreceptor from two perspectives: when to intervene and how to intervene. For intervention timing, we evaluate all 1,664 human-annotated consultation prefixes, including 720 intervene and 944 not-intervene examples, and report accuracy, precision, recall, and F1, treating intervention-required states as the positive class. For advice quality, we evaluate only the 720 prefixes where intervention is required. To isolate advice quality from intervention timing, the model is explicitly told that intervention is needed, and only its generated advice is evaluated. This setting measures whether MedPreceptor can provide correct, useful, and well-grounded guidance once the need for intervention has been established.

Table 9: A human-annotated NO-INTERVENE example.

Field	Content
Specialty	Respiratory Medicine
Chief complaint	Retrosternal chest pain beginning this morning, non-radiating, accompanied by dyspnea and chills. The visible medical history includes pulmonary embolism, left lower-extremity deep vein thrombosis, inferior vena cava filter placement, and warfarin treatment.
Dialogue prefix	**Doctor (Turn 1):** What does the chest pain feel like? Is it a pressure-like tightness, a sharp pain, or a cutting pain? **Patient:** It feels like a pressure on my chest. **Doctor (Turn 2):** Is this pressure constantly present, or does it come and go? **Patient:** The pressure is there continuously. **Doctor (Turn 3):** Does the pain get worse when you take a deep breath, cough, or press on your chest? **Patient:** I am not really sure. **Doctor (Turn 4):** Do you currently have any cough or sputum production? **Patient:** No, I do not have any cough or sputum production. **Doctor (Turn 5):** Did you check your temperature at home? Do you have a fever now? **Patient:** I am not really sure. **Doctor (Turn 6):** Did the chest pain and shortness of breath start suddenly, or did they gradually worsen? Have you recently had swelling or pain in one leg? **Patient:** The chest pain and shortness of breath started suddenly this morning. I am not really sure about any leg swelling or pain.
Human label	`NO_INTERVENE`
Human supporting evidence	The patient has chest pain, dyspnea, and a history of pulmonary embolism and deep vein thrombosis. The Doctor is appropriately asking about sudden onset and unilateral leg symptoms as part of the evaluation for possible pulmonary embolism. At this point, there is no clear and correctable inquiry error that requires intervention.
Human advice	None.

Advice quality is evaluated by GPT-5.6-Sol as an LLM judge. The judge is given the currently visible case information, the dialogue prefix, the human reference decision, supporting evidence, and suggested advice, together with the output of the evaluated model. It is not allowed to use future dialogue, hidden case information, or the final diagnosis. Each model is evaluated on all 720 prefixes where intervention is required.

LLM-as-Judge criteria. For advice quality, we use GPT-5.6-Sol to score each generated suggestion along five dimensions. Each dimension is scored from 0 to 2, giving a total score from 0 to 10. The detailed criteria are shown in Table 10.

Hard flags. The judge additionally checks for six severe error types: hallucinated facts, contradictions with the visible context, repetition of already answered information, disclosure of the final diagnosis, directly questioning the patient instead of advising the Doctor, and explicit references to dialogue turn numbers. Hallucinated or contradictory content sets **Grounding** to 0 and caps the total score at 4. Repeating already answered information sets **Efficiency / non-redundancy** to 0. Direct patient questioning, revealing the final diagnosis or answer, or referring to specific turn numbers sets **MedPreceptor style / safety** to 0.

B.4 Qualitative Examples on CACI

To provide a direct comparison of intervention behaviors, we present three representative cases from CACI. For each case, all models receive the same consultation prefix, and we compare their

Table 10: LLM-as-Judge criteria for evaluating advice quality.

Dimension	2 points	1 point	0 points
Problem alignment	Accurately addresses the current consultation problem and its clinical priority.	Partially relevant, but too broad or insufficiently focused.	Misaligned with the current problem or suggests the wrong direction.
Clinical actionability	Provides clear guidance for the Doctor's next inquiry step.	Provides some useful direction, but remains vague.	Too generic, not actionable, or unable to guide the next inquiry step.
Grounding	Fully grounded in the visible case information and dialogue prefix.	Contains minor unsupported assumptions or imprecise statements.	Hallucinates facts, contradicts visible information, or relies on unavailable information.
Efficiency / non-redundancy	Focuses on high-value missing information without repeating known information.	Contains some low-value or inefficient content.	Mainly repeats known information or suggests low-yield generic questions.
MedPreceptor style / safety	Concise, directed to the Doctor, and framed as non-diagnostic guidance.	Somewhat verbose or imperfectly phrased.	Directly questions the patient, reveals the final diagnosis or answer, or contains unsafe guidance.

generated advice with the human-annotated advice. The first two cases show generally effective interventions, while the final case illustrates a remaining failure mode of MedPreceptor.

Case 1: Repeated Questioning (Rheumatology).

> **Doctor (Turn 1):** When did your right hip pain start?
> **Patient:** I suddenly felt a sharp pain in my right hip when I stood up from a chair.
> **Doctor (Turn 2):** Is your right hip pain continuous or intermittent?
> **Patient:** Continuous.
> **Doctor (Turn 3):** Where exactly is your right hip pain located?
> **Patient:** It is right at the right hip joint.
> **Doctor (Turn 4):** Is your right hip pain continuous?
> **Patient:** Yes, it is continuous.

Human Advice. Avoid repeatedly asking whether the pain is continuous. Instead, ask about pain severity, ability to bear weight or walk, recent trauma or falls, and infection-related symptoms such as fever.

As shown in Table 11, MedPreceptor closely matches the human advice by stopping the repeated question and redirecting the consultation toward higher-value information.

Case 2: High-Priority Information Is Missed (Rheumatology).

> **Doctor (Turn 1):** Have you recently experienced marked fatigue, muscle soreness, or an irregular heartbeat?
> **Patient:** No, I have not experienced any of these symptoms.

Human Advice. First clarify how hyperkalemia was discovered, including whether it was identified during symptom evaluation, routine examination, or blood testing, and obtain the potassium level, renal-function results, and ECG findings.

Table 12 shows that MedPreceptor provides clinically useful guidance, although the priority of its suggested questions does not fully match the human reference.

Case 3: Unsupported Diagnostic Assumption (Rheumatology).

Table 11: Cross-model comparison for Case 1: repeated questioning.

Model	Advice	Evaluation
MedPreceptor (DG-OPD)	Avoid reconfirming pain continuity. Ask about pain severity, activity limitation, weight-bearing ability, fever, trauma history, and previous similar episodes.	10/10. Closely matches the human reference.
GPT-5.6-Sol	The continuity has already been established. Ask about pain severity, weight-bearing and activity limitation, trauma or falls, fever, and local redness or swelling.	10/10. Covers the main high-priority information.
DeepSeek-V4.1-Flash	Move to the trigger, pain characteristics and severity, aggravating or relieving factors, trauma, fever, limping or inability to bear weight, and previous similar episodes.	10/10. Correct and comprehensive, although somewhat verbose.
Baichuan-M2-32B	Ask about triggers, aggravating and relieving factors, fever, and degree of activity limitation, while avoiding further repetition.	10/10. Concise and clinically relevant.
Doctor-R1	Ask whether the hip pain is associated with fever, redness, swelling, or activity limitation to help distinguish infection or fracture.	8/10. Relevant, but misses several high-priority items.
Gemma-4-31B-IT	Stop repeating the question and ask about pain characteristics, severity, trauma history, fever, and activity limitation.	10/10. Closely aligned with the human reference.

Table 12: Cross-model comparison for Case 2: high-priority information is missed.

Model	Advice	Evaluation
MedPreceptor (DG-OPD)	Ask about recent changes in renal function, urine output, medications such as ACE inhibitors, ARBs, potassium-sparing diuretics or potassium supplements, dietary intake, vomiting, diarrhea, muscle symptoms, and relevant tests.	8/10. Useful, but does not prioritize the most immediate information.
GPT-5.6-Sol	Ask when hyperkalemia was discovered, whether it was confirmed on repeat testing, recent urine-output changes, kidney disease, dehydration or infection, muscle injury, medications, and potassium supplementation.	8/10. Appropriate, but underemphasizes potassium level and ECG findings.
DeepSeek-V4.1-Flash	Ask when hyperkalemia was identified, the potassium value, whether it was confirmed, renal-function abnormalities, urine-output changes, medications, and possible triggers.	7/10. Covers most relevant information but differs in prioritization.
Baichuan-M2-32B	Ask about onset, possible triggers, potassium measurements and trends, and associated symptoms such as palpitations or muscle weakness.	8/10. Relevant and reasonably focused.
Doctor-R1	Ask about renal dysfunction, potassium-sparing medications, and high-potassium food intake.	8/10. Relevant, but focuses on etiology before basic test information.
Gemma-4-31B-IT	Ask about onset, possible causes, renal dysfunction, and medication history.	8/10. Reasonable, but misses potassium level and ECG findings.

Doctor (Turn 1): When did your abdominal pain start?
Patient: I developed upper abdominal pain after returning to bed.

Doctor (Turn 2): Is the abdominal pain continuous or intermittent?
Patient: It started intermittently this morning.

Doctor (Turn 3): Do you have abdominal distension, nausea, or vomiting when the pain occurs?
Patient: I have nausea and vomiting.

Doctor (Turn 4): Is the vomit clear, or does it contain bile or food residue?
Patient: It contains food residue.

Table 13: Cross-model comparison for Case 3: unsupported diagnostic assumption.

Model	Advice	Evaluation
MedPreceptor (DG-OPD)	Focus on obstruction-related history, including cessation of bowel movements or flatus, progression of abdominal distension, pain location and progression, previous abdominal surgery, and hernia history.	4/10. Relies on an unsupported SBO assumption and misses the vomiting–pain relationship.
GPT-5.6-Sol	Ask about bowel movements and flatus, abdominal distension, vomiting frequency and volume, pain severity and progression, previous abdominal surgery, and similar episodes.	9/10. Useful, but does not prioritize the vomiting–pain relationship.
DeepSeek-V4.1-Flash	Ask whether vomiting relieves the pain, as well as bowel movements and flatus, abdominal distension, previous abdominal surgery, and similar episodes.	4/10. Covers the key question, but the full response introduces an unsupported SBO assumption.
Baichuan-M2-32B	Ask about the exact pain location, pain characteristics, aggravating and relieving factors, bowel movements and flatus, previous abdominal surgery, and similar episodes.	8/10. Reasonable, but misses the highest-priority vomiting–pain relationship.
Doctor-R1	Ask whether the Patient has stopped passing stool or flatus to help distinguish small-bowel obstruction from other causes of acute abdominal pain.	9/10. Concise and clinically useful without assuming the diagnosis.
Gemma-4-31B-IT	Ask about bowel movements, flatus, and the characteristics and severity of the abdominal pain.	4/10. The full response relies on an unsupported SBO assumption and misses the key relationship.

Doctor (Turn 5): Do you have fever or chills?
Patient: No fever or chills.

Human Advice. Since nausea, vomiting, and the characteristics of the vomit have already been established, further ask about the temporal relationship between vomiting and abdominal pain and whether the pain improves after vomiting, which may help distinguish possible causes of obstructive abdominal pain.

Table 13 illustrates a remaining limitation: MedPreceptor suggests clinically relevant questions, but its response relies on an unsupported SBO assumption and misses the highest-priority information in the human reference.

Overall, the first two cases show that MedPreceptor can provide useful advice that is largely aligned with human annotations. However, Case 3 shows that it can still rely on unsupported diagnostic assumptions or fail to prioritize the most informative next question.

C Cross-Model Analysis of Consultation Process Issues

C.1 Process-Level Issues in Human Clinical Practice

Previous clinical studies have shown that human physicians can also exhibit process-level problems similar to those considered in our work, including incomplete history taking, overly narrow inquiry, missed important information, and premature diagnostic closure.

Watari et al. (Watari et al., 2024) analyzed 687 diagnostic-error cases recalled by internists and found that 84.3% involved problems in data gathering or interpretation, including inadequate history taking in 23.4% of cases. Cognitive biases were also common, including confirmation bias, anchoring

bias, and premature closure. These findings suggest that missing important information, focusing too early on a particular hypothesis, and closing the diagnostic process prematurely are also common challenges in human clinical practice. Similar patterns have been reported in other settings. Singh et al. (Singh et al., 2013) found that most diagnostic errors in primary care involved breakdowns during the patient–practitioner encounter, with frequent problems related to history taking and diagnostic workup. Beckman and Frankel (Beckman & Frankel, 1984) showed that physicians often interrupted patients before they had fully expressed their concerns and redirected the conversation toward a specific issue, potentially leaving relevant information unexplored. Conversation-level analyses have also identified repetitive and poorly aligned information exchange in ineffective doctor–patient consultations (Angus et al., 2012).

Overall, these studies show that incomplete, misdirected, repetitive, or prematurely narrowed information gathering is not unique to Medical LLMs, but reflects broader challenges in clinical information acquisition and reasoning.

C.2 Related Findings in Medical LLM Benchmarks

Recent benchmarks for multi-turn medical consultation show that high final diagnostic accuracy does not necessarily imply an effective consultation process. Medical LLMs can still exhibit clear problems in how they gather information.

Specifically, **MedConsultBench** reports redundant or irrelevant questions, suboptimal inquiry order, and low information gain per turn, which are closely related to **repeated questioning**, **inefficient information gathering**, and **misdirected inquiry** (Qiao et al., 2026). **MedDialogRubrics** and **MAQuE** further show that Doctor LLMs often fail to cover important clinical information and may spend turns on low-value or weakly relevant questions, corresponding to **missing key information** and **inefficient information gathering** (Gong et al., 2026; 2025). In addition, **MINT** finds that some models commit to a diagnosis before sufficient evidence has been collected, which is closely related to **premature closure** (Fang et al., 2026).

Based on these recurring findings, we focus on five process-level consultation issues: *repeated questioning*, *inefficient inquiry*, *misdirected focus*, *omission of key clinical clues*, and *premature diagnostic closure*.

C.3 Cross-Model Analysis on MAQuE

We further examine whether the five process-level issues recur across different Doctor LLMs under the same consultation setting. We run five Doctor models on the same 200 MAQuE cases and identify representative intervention states. As shown in Table 14, different models exhibit distinct consultation problems, covering all five issue types considered in our analysis.

Table 14: Representative consultation process issues observed across different Doctor LLMs on MAQuE.

Dialogue Prefix	Issue Type and Evidence
Baichuan-M2-32B **Doctor (Turn 1):** Which eye feels irritated? **Patient:** My left eye. **Doctor (Turn 2):** When did the irritation in your left eye begin? **Patient:** It started immediately after I accidentally put EC liquid into my eye. **Doctor (Turn 3):** Did you have pain, tearing, sensitivity to light, or blurred vision right after the exposure? **Patient:** I immediately had pain and blurred vision. **Doctor (Turn 4):** Did you rinse your eye or use any medication after the exposure? **Patient:** I rinsed my eye after realizing the mistake. **Doctor (Turn 5):** Did the pain or blurred vision improve after rinsing? **Patient:** No, neither improved. **Doctor (Turn 6):** After rinsing, did you have eye redness, increased discharge, or persistent tearing? **Patient:** My left eye became red. **Doctor (Turn 7):** After rinsing, did you have increased discharge or persistent tearing from the left eye? **Patient:** My left eye became red.	**Repeated Questioning.** Turn 7 largely repeats the previous question about discharge and tearing without obtaining new clinically useful information.
Doctor-R1 **Doctor (Turn 1):** When did the swelling of your eyelids begin? **Patient:** I noticed it this morning when it was very hot outside. **Doctor (Turn 2):** Is the swelling on one side or both sides? **Patient:** Both sides. **Doctor (Turn 3):** Besides the eyelid swelling, do you have swelling elsewhere? **Patient:** Yes, I have swelling in my legs. **Doctor (Turn 4):** Have you noticed decreased urine output or darker urine? **Patient:** I'm not sure. **Doctor (Turn 5):** Have you ever had protein in your urine or foamy urine? **Patient:** Yes, I have had protein in my urine. **Doctor (Turn 6):** Do you have a history of high blood pressure or diabetes? **Patient:** I have a history of rheumatoid arthritis.	**Misdirected Inquiry.** After edema and proteinuria have been identified, the Doctor shifts to hypertension and diabetes before sufficiently exploring more direct renal clues, such as hematuria, weight change, or other urinary abnormalities.
DeepSeek-V4.1-Flash **Doctor (Turn 1):** Was the baby born at full term? **Patient:** No, the baby was born at 36 weeks. **Doctor (Turn 2):** Did the baby pass meconium after birth? **Patient:** I'm not sure. **Doctor (Turn 3):** Did the baby's abdomen become swollen or hard after vomiting? **Patient:** Yes, the abdomen became hard after vomiting. **Doctor (Turn 4):** Did you try feeding the baby after the vomiting? **Patient:** I'm not sure. **Doctor (Turn 5):** Has the baby had an abdominal X-ray? **Patient:** Yes. The doctor said the X-ray showed three gas shadows in the upper abdomen.	**Inefficient Information Gathering.** Key history about the progression of vomiting, bowel movements, and abdominal distension remains incomplete, but the Doctor moves early to imaging findings instead of completing the presenting history.
GLM-5.3-Flash **Doctor (Turn 1):** Where did your abdominal pain start? **Patient:** It started in the right lower abdomen. **Doctor (Turn 2):** Is the pain constant, or does it come and go like cramping? **Patient:** It comes and goes like cramping. **Doctor (Turn 3):** When was your last menstrual period? **Patient:** Around December last year. **Doctor (Turn 4):** Since your period stopped, have you had nausea, vomiting, or other early pregnancy symptoms? Have you taken a pregnancy test? **Patient:** I'm not sure.	**Missing Key Information.** The prolonged amenorrhea raises a pregnancy-related concern, but key obstetric information such as gestational age, contractions, vaginal bleeding or fluid loss, and fetal movement has not yet been collected.
GPT-5.6-Sol **Doctor (Turn 1):** When did the headache begin, and does it improve when you lie down? **Patient:** I have had headaches for ten years. They get worse when I stand up. **Doctor (Turn 2):** Have you previously had a lumbar puncture, spinal surgery, or an injury to your head or spine? **Patient:** I'm not sure. **Doctor (Turn 3):** How quickly does the headache improve after you lie down? **Patient:** It clearly improves when I lie down. **Doctor (Turn 4):** Is the headache accompanied by tinnitus, hearing changes, nausea, or double vision? **Patient:** I have migraines. **Doctor (Turn 5):** Have you had a brain or spine MRI, and did the imaging show a cerebrospinal fluid leak or any meningeal abnormality? **Patient:** I had a brain MRI and CT myelography. The imaging showed a ventral epidural cerebrospinal fluid collection extending from the mid-cervical to the mid-thoracic spine.	**Premature Closure.** After identifying the positional nature of the headache, the Doctor directly asks about imaging findings that support a specific diagnosis before fully completing the history and possible causes.