Diagnosing Lower Extremity Arteriovenous Diseases Using Agentic LLMs

Zicen Liao[1] **Yunhao Sun**[2] **Matthew Purver**[1,3]
[1]Queen Mary University of London
[2]The First Affiliated Hospital of Army Medical University, PLA
[3]Jožef Stefan Institute
z.liao@qmul.ac.uk sunyh25@alumni.sysu.edu.cn m.purver@qmul.ac.uk

Abstract

Large language models (LLMs) achieve strong results on medical benchmarks, but real outpatient diagnosis is iterative and requires stable, guideline-consistent reasoning rather than one-shot predictions. We study this problem in lower-extremity arteriovenous (LEA) diseases and introduce LEA-Dialog, a multi-turn diagnostic dialogue dataset covering six common diseases, with stage annotations for each turn and guideline-grounded probability trends. Building on this dataset, we further propose a process-aligned agentic framework for real-world patient consultations. In this framework, the patient interacts with a Doctor agent, while a Teacher agent audits guideline adherence and reasoning quality, provides corrective feedback to refine the Doctor agent's behaviour, and ultimately produces a structured diagnostic report. An Evaluator agent then derives both turn-level and final rewards from the diagnostic conversations between the patient and the Doctor agent for PPO-based policy optimisation, while carefully selected high-quality diagnostic conversations are further used for supervised fine-tuning. For training, the patient is also simulated by a Patient agent that emulates real patient responses. In addition to final diagnostic accuracy, we evaluate unsupported claims, trend correctness, diagnostic-tree similarity, and repeated-run stability with a new metric, Rate-n. Experiments on both online and offline LLMs show that process-aligned multi-turn diagnosis reduces reasoning drift, improves stability, and yields the largest gains for smaller offline models.

1 Introduction

Large language models (LLMs) have achieved strong performance on a wide range of medical benchmarks, including medical question answering, report generation, and simulated diagnosis. However, high benchmark accuracy alone does not make a system clinically usable. In outpatient practice, diagnosis is rarely a one-shot prediction: physicians iteratively summarise the current patient state, refine the differential diagnosis, ask targeted follow-up questions, and update their beliefs as new evidence emerges (Vessey and Galletta, 1991). In this setting, clinicians care not only about final correctness but also about stable and guideline-consistent reasoning (Yang et al., 2024; Rosenbacke et al., 2024).

This gap is especially important for multi-turn diagnostic dialogue. A model may produce the correct final diagnosis while still showing undesirable behaviour during the interaction, such as drifting between incompatible hypotheses, asking poorly targeted questions, or generating unsupported intermediate claims. Such process-level failures reduce trust and make the model's reasoning difficult to inspect in clinical workflows. Nevertheless, most existing evaluations of medical LLMs still focus primarily on final diagnostic accuracy, with limited attention to turn-level reasoning quality, diagnostic trajectory consistency, or repeated-run stability (Liu et al., 2024; Schmidgall et al., 2024; Fan et al., 2025; Wang et al., 2024; Lyu et al., 2024; Xu et al., 2024). In our setting, even strong online models can show noticeable variation across repeated runs on the same case, suggesting that stability should be treated as a first-class target in medical dialogue evaluation.

To study this problem, we focus on lower-extremity vascular diseases, where diagnosis often depends on multi-round clarification of symptoms, progressive narrowing of the differential diagnosis, and dynamic updates of disease likelihoods across consultation turns. Guided by clinical reasoning practice, we formulate outpatient diagnosis as a process-aligned hypothetico-deductive loop consisting of four stages at each turn: Summary, Inference, Question Planning, and Diagnosis. Based on this formulation, we introduce LEA-Dialog, a multi-turn diagnostic dialogue dataset covering

Proceedings of the 25th Workshop on Biomedical Language Processing (BioNLP 2026), pages 250–267

six common lower-extremity arterial and venous diseases. Each dialogue is annotated with turn-wise reasoning and guideline-grounded probability trends, enabling evaluation of both final outcomes and intermediate reasoning behaviour.

Building on this dataset, we propose a process-aligned agentic framework for multi-turn diagnosis. A Doctor agent conducts the consultation, a Patient agent simulates the case trajectory, a Teacher agent audits record consistency, guideline adherence, and inquiry quality to construct a structured report, and an Evaluator agent converts these signals into turn-level and final rewards for optimisation. We also introduce Rate-n, a repeated-run stability metric that measures whether a model follows the same diagnostic trajectory across the first n turns. Specifically, a trajectory is considered consistent if its trend direction matches the majority trend across runs at each turn. Experiments on both online and offline LLMs show that process alignment reduces reasoning drift, improves diagnostic stability, and yields larger gains for smaller offline models.

Our contributions are as follows:

1. We introduce LEA-Dialog, a multi-turn diagnostic dialogue dataset covering six common diseases, annotated with reasoning processes for each turn and a carefully curated diagnostic handbook.

2. We propose a process-aligned agentic diagnostic framework with Doctor, Patient, Teacher, and Evaluator roles, together with a turn-level optimisation strategy for stabilising multi-turn reasoning.

3. We show that, under our framework, o3 achieves 4.6% improvement in average diagnostic accuracy and 66.7% reduction in misdiagnosis rate.

4. We further show that Qwen3-14B-T gains 68.3% in average diagnostic accuracy under the proposed framework, surpassing the original o3 model by 0.9%.

5. In addition to diagnostic accuracy, we demonstrate substantial gains in reasoning stability: after applying our framework, Qwen3-14B-T exceeds o3 in output stability by 5% from the second turn onward and by 25% from the third turn onward.

2 Related Work

2.1 Medical LLMs and diagnostic dialogue

Large language models have shown strong performance on medical question answering and benchmark-style evaluation, including systems such as AMIE (Tu et al., 2025), Med-PaLM2 (Singhal et al., 2025), and recent medical benchmarks in English and Chinese (Liu et al., 2024; Schmidgall et al., 2024; Fan et al., 2025; Wang et al., 2024; Lyu et al., 2024; Xu et al., 2024). More recent work has also begun to move from static question answering toward interactive diagnosis and simulated consultation, such as multi-agent clinical environments and conversational diagnostic systems (Schmidgall et al., 2024; Fan et al., 2025; Tu et al., 2025; Liu et al., 2025). However, most prior work still evaluates models primarily by final-answer quality, with limited focus on intermediate reasoning consistency.

2.2 Process-aligned and multi-step medical reasoning

A related line of work attempts to align model behaviour with the structure of clinical reasoning. In particular, Reasoning Like a Doctor (Xu et al., 2024) formulates medical dialogue generation through diagnostic reasoning process alignment, and MSDiagnosis introduces EMR-based multi-step diagnosis annotations (Hou et al., 2024). These studies support the view that medical dialogue should be modelled as a staged reasoning process rather than a single response generation problem. Our work follows this direction, but focuses more explicitly on the iterative nature of outpatient diagnosis and on modelling the diagnostic process itself rather than only the final prediction. Specifically, we treat diagnosis as a multi-round reasoning loop in which the model summarises accumulated evidence, updates its differential diagnoses, determines what information is still missing, and asks targeted follow-up questions to reduce uncertainty before making a final judgment. We further model turn-level diagnostic trajectory evolution, requiring disease hypotheses to change in a clinically meaningful manner across rounds. To support this, we introduce guideline-grounded supervision that specifies how diagnostic probabilities should be updated as new symptoms, history, and examination findings emerge, enabling process-level training and evaluation beyond endpoint accuracy. The comparison of evaluation dimensions between our model and other approaches is shown in Table 1.

2.3 Trustworthiness, guideline adherence, and stability

Accuracy alone is not sufficient for evaluating medical AI systems. Prior work has also highlighted the

Model	Focus	Stability
AMIE	Diagnostic dialogue	Not central
Med-PaLM2	QA and safety	Not central
ChatDoctor	Response quality	Not central
DocCHA	Online diagnosis	Not central
Emulation	Process alignment	Not central
Ours	Turn-level trajectory	Rate-n

Table 1: Evaluation focus of representative studies: AMIE (Tu et al., 2025), Med-PaLM2 (Singhal et al., 2025), ChatDoctor (Li et al., 2023), DocCHA (Liu et al., 2025), and Emulation (Xu et al., 2024). Stability refers to repeated-run diagnostic trajectory stability; "Not central" indicates that it is not a primary evaluation target.

importance of fitting clinical workflows (Xu et al., 2024), providing interpretable reasoning, and supporting clinician trust. In outpatient settings, diagnosis is inherently iterative: clinicians summarise the case, update differential diagnoses, ask targeted follow-up questions, and gradually move toward a decision. Yet the stability of this process across repeated runs remains rarely examined in medical AI, as shown in Table 1. Our work addresses this gap by studying multi-turn LLM diagnosis from a process perspective, introducing a process-aligned framework, a dataset with turn-level trend supervision, and a repeated-run stability metric, Rate-n.

3 LEA-Dialog Dataset

3.1 Task Scope

LEA-Dialog focuses on six common lower-extremity arteriovenous (LEA) diseases: deep vein thrombosis (DVT), post-thrombotic syndrome (PTS), arteriosclerosis obliterans (ASO), arterial thrombosis (AT), varicose veins (VV), and thromboangiitis obliterans (Buerger disease). We choose this disease set because diagnosis in this domain is highly iterative, requiring repeated clarification of symptoms and progressive updating of disease likelihoods across consultation rounds.

3.2 Data Composition

The dataset contains three components:

1. Expert-written consultation records. We collect 120 consultation records manually compiled and refined by experts based on real outpatient visits.

2. Synthetic consultation records. To improve coverage and diagnostic difficulty, we further construct 480 synthetic consultation records using data augmentation based on expert-crafted seed cases.

3. Clinical guideline file. In addition to the dialogue data, we include a companion guideline document for lower-extremity vascular medicine, which specifies diagnostic criteria and structured probability-scoring rules for supervising disease-trend evolution across turns.

Overall, LEA-Dialog combines expert-authored high-quality cases with larger-scale augmented data, while keeping all dialogues grounded in a shared guideline source. The detailed case collection procedure is provided in Appendix A, and the handbook development plan and reference guidelines are presented in Appendix B.

3.3 Annotation Scheme

Each dialogue contains up to 15 turns, with most expert cases reaching a final diagnosis within 5-7 rounds. For each turn, the doctor-side reasoning is annotated with four process-aligned stages:

Summary: a structured summary of the current patient state;

Inference: revision of the differential diagnosis based on newly acquired evidence;

Question Planning: targeted follow-up questions for reducing diagnostic uncertainty;

Diagnosis: final diagnostic output, produced only when sufficient evidence has been collected.

In the final round, each case additionally includes a trend report, which records how the estimated probability of each candidate disease changes across turns and why. Unlike free-form confidence scores, these trends are grounded in the companion guideline file, making them suitable for both evaluation and reward design in later training.

Compared with prior multi-turn medical datasets, LEA-Dialog provides stage annotations for each turn, guideline-grounded probability trends, and an expert-developed diagnostic handbook, enabling evaluation of diagnostic trajectory quality and repeated-run stability. Specific examples from the dataset are provided in Appendix C.3, Appendix C.5, and Appendix C.6.

3.4 Data Construction

We construct LEA-Dialog in three stages. First, three researchers with vascular-surgery backgrounds expand de-identified and standardised medical records into outpatient-style diagnostic dialogues following the annotation scheme in Section 3.3 and the handbook workflow. For each of the six target diseases, 20 expert-written records are created, yielding 120 seed cases.

Second, we use o3 with one-shot prompting to generate 80 augmented dialogues per disease, resulting in 480 synthetic cases. Following expert advice, selected disease-defining features are removed from some prompts to create more difficult differential-diagnosis cases.

Third, all generated dialogues undergo quality control. Two researchers independently review each dialogue for information completeness, temporal coherence, premature diagnosis leakage, label consistency, and evidential scope. Disagreements are resolved by a senior vascular surgeon. Further details are provided in Appendix A.

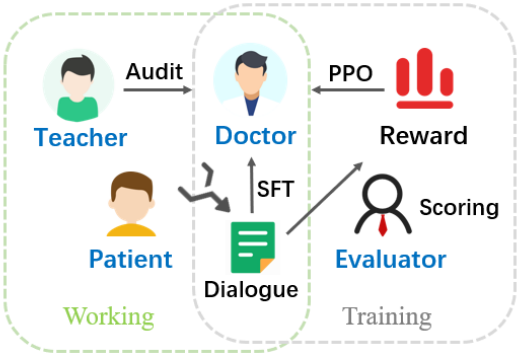


Figure 1: Overview of the proposed multi-agent diagnostic framework, illustrating the interaction between the Doctor, Patient, Teacher, and Evaluator agents.

4 Method

4.1 Process-Aligned Diagnostic Loop

We model outpatient diagnosis as a process-aligned, multi-turn hypothetico-deductive loop. At each consultation turn, the model performs four sequential reasoning stages: Summary, Inference, Question planning, and Diagnosis. The first three stages are repeated across turns to refine the differential diagnosis and reduce uncertainty, while the diagnosis stage is executed only when sufficient evidence has been collected.

More concretely, the summary stage updates a structured representation of the patient's current condition based on the dialogue history. The inference stage revises the differential diagnosis and the relative likelihood of candidate diseases. The question planning stage generates targeted follow-up questions intended to resolve remaining ambiguity. Finally, when the model determines that diagnostic uncertainty has been reduced sufficiently, it enters the diagnosis stage and produces the final diagnostic decision together with a disease-trend report. This staged formulation is also consistent with the annotation structure of LEA-Dialog, examples of which are provided in Appendix C.3, Appendix C.5, and Appendix C.6, and supports supervision for both intermediate reasoning and final outcomes.

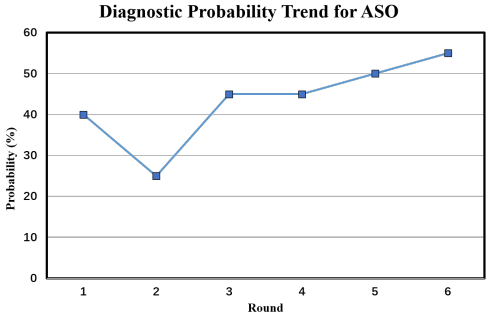
4.2 Multi-Agent Diagnostic Framework

Building on this diagnostic loop, we propose a multi-agent framework consisting of four roles: a Doctor agent, a Patient agent, a Teacher agent, and an Evaluator agent, shown in Figure 1. The Doctor agent is the trainable policy and is responsible for conducting the consultation. The Patient agent simulates the patient side of the dialogue by responding according to the case history. The Teacher agent acts as a clinical instructor who compares the Doctor's behaviour against the guideline file and provides structured feedback. The Evaluator agent converts the consultation trace into optimisation signals for training. The prompts and outputs example for each agent are provided in Appendix C.

The Doctor agent follows the four-stage reasoning structure at each turn. By explicitly separating summarisation, differential refinement, follow-up questioning, and diagnosis, the Doctor agent produces outputs that are easier to audit than a single free-form response. In the question-planning stage, particular emphasis is placed on clarifying ambiguous or potentially incomplete patient descriptions, reflecting the uncertainty commonly encountered in outpatient consultations.

The Patient agent plays the role of the patient and answers questions using the diagnostic history provided in the dataset. This design allows us to simulate multi-turn consultations in a controlled setting while keeping the dialogue grounded in case-specific evidence.

The Teacher agent evaluates the Doctor's reasoning process after each turn by cross-referencing it with the companion guideline file. It scores the appropriateness of the Doctor's actions, identifies omitted or mistimed questions, and provides strategic feedback such as revisiting a previous step or proceeding to the next stage. After the final diagnosis, the Teacher agent also generates a structured summary of the diagnostic trajectory, including the historical evolution of disease-probability assessments. This report provides a turn-level reference for subsequent training and evaluation.



- (1) 40% in Round 1 based only on sudden exertional pain;
- (2) Lowered to 25% in Round 2 due to lack of classic 5 Ps;
- (3) Raised to 45% in Round 3 with typical claudication (100m pain, relief with rest) and no venous signs;
- (4) Kept at 45% in Round 4 with sedentary lifestyle and cessation of anticoagulation but no strong trigger;
- (5) Raised to 50% in Round5 due to prior DVT with anticoagulation stop, suggesting a prothrombotic milieu;
- (6) Final 55% since still no swelling/warmth and arterial risk predominates.

Figure 2: Visualisation of diagnostic probability trends across rounds. The figure shows the probability assigned by the doctor agent to ASO at the end of each round, with the corresponding rationale. The full report includes probability trends and explanations for all candidate diseases.

4.3 Teacher-Guided Diagnostic Trajectory Analysis Report

A key component of our framework is the structured diagnostic trace produced by the Teacher agent. Rather than treating the consultation as an unstructured sequence of responses, we represent it as a partial diagnostic trend grounded in clinical guideline nodes shown in Figure 2. Each consultation turn contributes to this trend by adding or revising evidence, narrowing the differential diagnosis, and updating probability trends over candidate diseases. The final report, therefore, captures not only the final diagnosis but also whether the model followed an appropriate diagnostic trajectory.

This representation serves two purposes. First, it improves interpretability by making intermediate decisions easier to inspect. Second, it enables turn-level supervision beyond final-answer accuracy. In particular, we use the Teacher-generated report to assess whether the Doctor agent followed the handbook guidance, whether probability trends changed in clinically plausible directions, and whether unsupported claims were introduced during reasoning.

4.4 Turn-Level optimisation

To optimise the Doctor agent for both diagnostic correctness and process stability, we define the total reward as a weighted combination of turn-level rewards and a final reward:

$$G = \sum_{t=1}^{T} R_t^{\text{inst}} + R^{\text{final}}$$

where R_t^{inst} is the immediate reward at turn t,R^{final} is the final reward for the completed consultation. Intuitively, the turn-level reward encourages the Doctor agent to follow a clinically coherent diagnostic trajectory by rewarding appropriate guideline-node expansion and correct probability-trend updates while penalising unsupported claims. The final reward evaluates the overall diagnostic quality, including final diagnosis correctness and trajectory-level similarity to the reference diagnostic tree. Full details of the reward design are deferred to Appendix D.

5 Experimental Setup

5.1 Models

We evaluate both online and offline LLMs. For the Doctor agent, we use two online models, GPT-4o and o3, and two offline models, Qwen3-8B and Qwen3-14B. For the two offline models, we further evaluate their framework-trained versions after supervised fine-tuning and PPO, denoted as Qwen3-8B-T and Qwen3-14B-T, respectively. This setup allows us to test whether the proposed framework improves not only strong proprietary models but also smaller deployable open-weight models, and whether process-aligned optimisation can reduce the gap between offline and online systems. Unless otherwise specified, GPT-4o is used for the Patient, Teacher, and Evaluator roles.

5.2 Training and Comparison Settings

The Doctor agent is optimised using a hybrid PPO+SFT strategy; full details of the training framework are provided in Appendix E. During training, the Doctor interacts with the Patient agent in a multi-round consultation setting. Based on the interaction history, the Teacher agent audits guideline-node coverage, probability-trend updates, and unsupported claims, and provides structured process-level feedback. The Evaluator agent then converts these signals, together with the final diagnostic outcome, into training rewards. To maintain

language fluency and structural consistency, we interleave reinforcement learning with supervised fine-tuning, performing one SFT round after every five PPO rounds.

We compare each framework-enhanced Doctor model with its corresponding base model under direct prompting. We further conduct ablation studies on three key components of the framework: the companion guideline file, the Teacher agent, and the turn-level optimisation procedure, to isolate the contributions of process alignment, structured feedback, and reward-based optimisation.

5.3 Evaluation Metrics

We evaluate the framework from four perspectives: final diagnostic performance, Teacher reliability, repeated-run stability, and process-level error patterns.

For final diagnostic performance, we report final accuracy. For each case, the model outputs a probability distribution over the candidate diseases. The disease with the highest final probability is taken as the prediction and compared with the reference clinical diagnosis. A prediction is counted as correct if the predicted and reference labels refer to the same disease category; this matching is manually verified to avoid surface-form variation. Formally,

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}[\hat{y}_i = y_i],$$

where N is the number of evaluated cases, $\hat{y}_i$ is the manually verified top-probability diagnosis, and y_i is the reference diagnosis.

For Teacher reliability, we evaluate three binary process-level error types: record-inconsistent summary, reasoning non-adherence, and questioning non-adherence. These correspond to contradiction or omission in the summary, deviation from handbook-defined diagnostic logic, and failure to ask guideline-relevant or discriminative follow-up questions. We compare the Teacher's judgements with expert annotations using confusion matrices. The validation sample contains 102 dialogue histories, with 17 instances per disease, each independently assessed by three clinical experts.

For repeated-run stability, we report Rate-n. For each test case, the model is run multiple times, and a run is counted as stable up to turn n if the trend direction of the correct disease matches the majority trend direction at each of the first n turns.

Model	Base	Multi Only	HB Only	Multi+HB
Qwen3-8B	19.8	27.2	62.1	89.7
Qwen3-8B-T	23.2	27.3	65.4	90.3
Qwen3-14B	23.5	31.0	69.3	92.3
Qwen3-14B-T	25.7	31.3	71.1	94.0
GPT-4o	87.6	89.3	90.9	96.0
o3	93.1	93.3	96.9	97.7

Table 2: Performance comparison (Acc %) under different settings on the held-out test set ($N = 60$, 10 cases per disease). Qwen3-8B-T and Qwen3-14B-T denote the models after PPO+SFT training. Multi refers to the multi-step reasoning scheme, while HB refers to the use of the handbook to provide diagnostic guidance and supplementary knowledge to the model.

Finally, we analyse misdiagnosed trajectories to identify the first clinically meaningful deviation and its dominant error type.

6 Results

6.1 Main Results

Table 2 summarises the main diagnostic results. Overall, the proposed framework with the handbook improves performance across both online and offline models, with especially strong gains for the smaller offline models. In particular, framework-enhanced Qwen3-14B-T with handbook improves average diagnostic accuracy by 68.3% and surpasses the original o3 model by 0.9%, showing that process-aligned optimisation and refined diagnosis handbook can substantially narrow the gap between deployable offline models and strong online models. For o3, the framework still yields measurable benefits, improving average diagnostic accuracy by 4.6% and reducing the misdiagnosis rate from 6.9% to 2.3%. The handbook is a primary source of the observed performance gains. In the handbook ablation study, access to the handbook substantially improves diagnostic accuracy for offline models and also yields a smaller but consistent improvement for online models. Notably, this benefit becomes larger after applying the multi-step framework, suggesting that the proposed multi-stage diagnostic procedure is more effective at incorporating external clinical knowledge. In contrast, when the underlying diagnostic logic is unchanged, reinforcement learning alone provides only limited gains in accuracy. These results indicate that, for final diagnostic performance, the effective use of external clinical knowledge is more critical than policy optimisation alone. This finding highlights the importance of knowledge utilisation,

rather than optimisation alone, in process-aligned diagnostic systems.

6.2 Teacher Reliability and Ablation

We next provide a more detailed analysis of the Teacher agent's performance. Specifically, we first evaluate whether the Teacher can correctly identify unsupported claims produced by the Doctor agent. Naturally, given the three stage labels in each diagnostic round—Summary, Reasoning, and Question Planning—we categorize such unsupported behaviours into three corresponding types: Record-inconsistent summary, where the Doctor's summary is inconsistent with the patient record or omits key clinical information; Reasoning non-adherence, where the Doctor's reasoning deviates from the guideline; and Questioning non-adherence, where the Doctor's planned follow-up questions do not conform to the guideline requirements.

The confusion matrix for Questioning non-adherence is shown in Figure 3, while the corresponding results for the other two criteria are provided in Appendix F. The confusion matrices across the three unsupported-claim criteria show that the Teacher's behaviour varies by criterion. For record-inconsistent summaries and reasoning guideline non-adherence, the Teacher is relatively conservative, producing few false positives but missing a non-trivial portion of rare errors. For questioning guideline non-adherence, it is more sensitive, identifying more problematic questioning steps at the cost of increased false positives and additional re-questioning. These results suggest that the Teacher does not apply a uniform decision rule, but instead exhibits a criterion-dependent safety–efficiency trade-off.

Teacher ablation further shows that removing the Teacher yields only a modest change in overall diagnostic accuracy, but substantially weakens record consistency, guideline questioning adherence, and guideline reasoning adherence. This finding indicates that the Teacher's main contribution is not simply improving final accuracy, but regularising the diagnostic process toward handbook-consistent and more trustworthy behaviour.

6.3 Stability Analysis

In our experiments, we observed that when the model repeatedly diagnoses the same patient, the diagnostic trajectories tend to remain highly similar across successful runs, as illustrated by Test 1–3 in Figure 4, whereas failed runs (Test 4–5) often

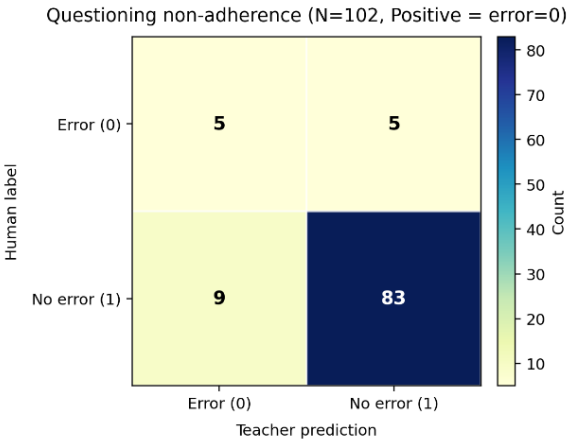


Figure 3: Confusion matrices for questioning non-adherence. In total, we sampled 102 dialogue histories, with 17 instances per disease. Each history was independently assessed by three experts for the three types of unsupported behaviours. Histories without unsupported behaviour were labelled as Positive (1), whereas those containing unsupported behaviour were labelled as Negative (0).

follow different trajectories. This divergence arises because the model cannot perfectly adhere to the guidelines in every run. Under the assumption that the diagnostic guideline is of sufficiently high quality, the stability of the model's outputs becomes an important determinant of diagnostic accuracy. Motivated by this observation, we introduce Rate-n to evaluate output stability.

Rate-n measures whether, across 10 repeated runs on the same case, the model maintains a similar diagnostic trajectory for the correct disease within the first n rounds. Here, a *similar trajectory* refers to the majority trend direction observed across runs, that is, whether the probability of the correct disease increases or decreases at each turn. Using Figure 4 as an example, among 10 repeated runs, the model correctly identified the possibility of ASO in only 5 runs at Round 1 (namely, Test1–5), yielding a Rate-1 of 50%. At Round 2, Test1, Test2, Test3, and Test5 shared the same trend for ASO (a decrease), which constitutes the majority pattern, yielding a Rate-2 of 40%. By the same logic, Rate-3 is 30%.

It is worth noting that a decrease in the probability trend of the correct disease at a particular round does not necessarily indicate an error in either the model or the guideline. Rather, it may reflect that the newly collected evidence at that stage also supports other competing diagnoses, which is common in real clinical reasoning.

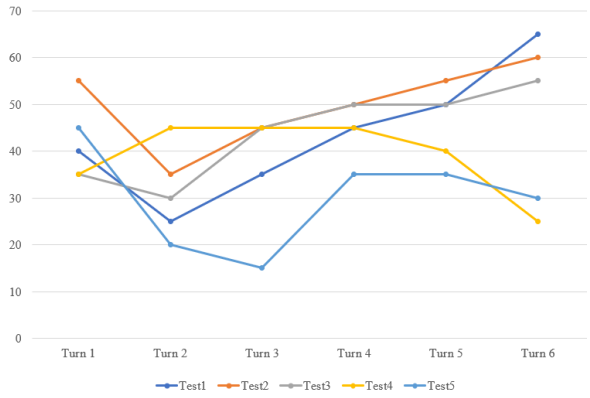


Figure 4: Evolution of the model's inferred probability for ASO across sequential rounds in an ASO clinical case (%). Among them, Tests 1–3 correspond to successful diagnoses and show similar diagnostic trends, whereas Tests 4 and 5 correspond to failed diagnoses and follow trajectories that differ from the others.

Model	Rate-1	Rate-2	Rate-3
Qwen3-8B	91.7	76.7	58.3
Qwen3-8B-T	91.7	81.7	61.7
Qwen3-14B	93.3	80.0	61.7
Qwen3-14B-T	95.0	83.3	66.7
GPT-4o	96.7	76.7	45.0
o3	98.3	78.3	41.7

Table 3: Stable diagnostic rates across sequential turns for six distinct disease categories

As shown in Table 3, across models, large online models show a sharper decline in Rate-n after the first turn, whereas small offline models exhibit a smaller drop. The PPO-trained offline models are the most stable overall. In particular, Qwen3-14B-T outperforms o3 by 5% from Turn 2 onward and by 25% from Turn 3 onward, showing that process-aligned RL can substantially reduce reasoning drift in multi-turn diagnosis.

This result shows that stronger base capability does not necessarily translate into more stable diagnostic trajectories. In our setting, reinforcement learning (PPO) acts as a regulariser that keeps the model more tightly aligned with the structured knowledge source, leading to more consistent reasoning across repeated trials.

6.4 Error Pattern Analysis

We further conduct a retrospective error-pattern analysis on 60 misdiagnosed trajectories to identify both the earliest clinically meaningful deviation and its primary error type. We annotate three error categories:
(1) Summary error, where the model distorts the patient summary or omits key symptoms;
(2) Reasoning error, where the model fails to follow the handbook-defined differential diagnostic logic;
(3) Question-planning error, where the model fails to ask the most discriminative follow-up question at a critical decision point.

As shown in Table 4, the dominant failure mode is question-planning error, which accounts for 53.3% of all cases. This indicates that the main bottleneck is often not missing medical knowledge, but failure to ask the most discriminative next question at critical branching points. Reasoning errors are the second most frequent category (40.0%), while summary errors are less common but tend to occur earlier and distort downstream reasoning.

Errors also emerge early: 71.7% of trajectories show their first meaningful deviation within the first two turns. Summary errors tend to concentrate in Turn 1, question-planning errors in Turn 2, and reasoning errors more often from Turn 2 onward. These findings suggest that misdiagnosis rarely results from a single final-turn mistake, but instead develops progressively from earlier process-level deviations.

Error Type	Number	Rate
Question Planning Error	32	53.3%
Reasoning Error	24	40.0%
Summary Error	4	6.7%
Total	60	100.0%

Table 4: Distribution of process-level error types (N=60)

7 Conclusion

We introduced LEA-Dialog and a process-aligned agentic framework for multi-turn diagnosis of lower-extremity arteriovenous diseases. The results show that evaluating only final accuracy is insufficient: guideline-grounded reasoning, process-level errors, and repeated-run stability are also important for assessing diagnostic dialogue systems. Structured clinical guidance and process-level optimisation improve performance and reduce reasoning drift, particularly for smaller offline models. Future work will broaden disease coverage and include more real clinical interactions.

8 Limitations

This study has several limitations. LEA-Dialog focuses on a narrow but clinically coherent group of six lower-extremity arteriovenous diseases. This enables controlled analysis of diagnostic trajectories among closely related vascular conditions, but limits claims about generalisation to broader medical domains. In addition, part of the dataset is synthetically expanded from expert-written seed cases, which may introduce distributional bias and make the dialogues more regular than real outpatient interactions. We therefore view LEA-Dialog as a controlled, domain-specific testbed for process-aligned diagnostic reasoning rather than a fully representative corpus of clinical consultations. The simulated multi-agent setup and the use of GPT-4o for the Patient, Teacher, and Evaluator roles may further introduce an LLM echo-chamber effect. We mitigate this risk by grounding patient responses in case records, using a guideline-based handbook, and validating a subset of Teacher judgements against expert annotations. Due to clinical data governance constraints, the original expert-derived records cannot be fully released. We plan to release the synthetic dialogues, annotation schema, prompts, handbook structure, and evaluation scripts where permitted, subject to institutional approval. Future work should broaden disease coverage, include more real clinical interactions, and validate the framework through independent clinician-in-the-loop studies.

Acknowledgements

This work was partly supported by EPSRC via Responsible AI UK (grant number EP/Y009800/1, project KP0016 AdSoLve) and the Slovenian Research Agency (ARRS) under core research programme Knowledge Technologies (P2-0103).

References

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

Ruihui Hou, Shencheng Chen, Yongqi Fan, Lifeng Zhu, Jing Sun, Jingping Liu, and Tong Ruan. 2024. MS-Diagnosis: An EMR-based dataset for clinical multi-step diagnosis. *arXiv preprint arXiv:2408.10039*.

Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. 2023. ChatDoctor: A medical chat model fine-tuned on a large language model Meta AI (LLaMA) using medical domain knowledge. *Cureus*, 15(6).

Mianxin Liu, Weiguo Hu, Jinru Ding, Jie Xu, Xiaoyang Li, Lifeng Zhu, Zhian Bai, Xiaoming Shi, Benyou Wang, Haitao Song, Pengfei Liu, Xiaofan Zhang, Shanshan Wang, Kang Li, Haofen Wang, Tong Ruan, Xuanjing Huang, Xin Sun, and Shaoting Zhang. 2024. MedBench: A comprehensive, standardised, and reliable benchmarking system for evaluating Chinese medical large language models. *Big Data Mining and Analytics*, 7(4):1116–1128.

Xinyi Liu, Dachun Sun, Yi Fung, Dilek Hakkani-Tur, and Tarek F Abdelzaher. 2025. DocCHA: Towards LLM-augmented interactive online diagnosis system. In *Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 609–619.

Shiwei Lyu, Chenfei Chi, Hongbo Cai, Lei Shi, Xiaoyan Yang, Lei Liu, Xiang Chen, Deng Zhao, Zhiqiang Zhang, Xianguo Lyu, Ming Zhang, Fangzhou Li, Xiaowei Ma, Yue Shen, Jinjie Gu, Wei Xue, and Yiran Huang. 2024. RJUA-QA: A comprehensive QA dataset for urology.

Rikard Rosenbacke, Åsa Melhus, Martin McKee, and David Stuckler. 2024. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: Systematic review. *JMIR AI*, 3:e53207.

Samuel Schmidgall, Rojin Ziaei, Carl Harris, Eduardo Reis, Jeffrey Jopling, and Michael Moor. 2024. AgentClinic: A multimodal agent benchmark to evaluate AI in simulated clinical environments. *arXiv preprint arXiv:2405.07960*.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R. Pfohl, Heather Cole-Lewis, Darlene Neal, Qazi Mamunur Rashid, Mike Schaekermann, Amy Wang, Dev Dash, Jonathan H. Chen, Nigam H. Shah, Sami Lachgar, Philip Andrew Mansfield, and 16 others. 2025. *Nature Medicine*, 31(3):943–950.

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, and 7 others. 2025. Towards conversational diagnostic artificial intelligence. *Nature*, 642(8067):442–450.

Iris Vessey and Dennis Galletta. 1991. Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1):63–84.

Xidong Wang, Guiming Chen, Song Dingjie, Zhang Zhiyi, Zhihong Chen, Qingying Xiao, Junying Chen,

Feng Jiang, Jianquan Li, Xiang Wan, Benyou Wang, and Haizhou Li. 2024. CMB: A comprehensive medical benchmark in Chinese. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6184–6205, Mexico City, Mexico.

Kaishuai Xu, Yi Cheng, Wenjun Hou, Qiaoyu Tan, and Wenjie Li. 2024. Reasoning like a doctor: Improving medical dialogue systems via diagnostic reasoning process alignment. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6796–6814.

Bufang Yang, Siyang Jiang, Lilin Xu, Kaiwei Liu, Hai Li, Guoliang Xing, Hongkai Chen, Xiaofan Jiang, and Zhenyu Yan. 2024. DrHouse: An LLM-empowered diagnostic reasoning system through harnessing outcomes from sensor data and expert knowledge. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(4):1–29.

A Detailed Case Collection Procedure

A.1 Study Design and Data Sources

This study is a single-center retrospective study. Cases were collected from routine clinical care at a tertiary hospital, including the electronic medical record system, laboratory information system, imaging reports, surgical records, and discharge summaries. The study population included patients with DVT, PTS, lower-extremity varicose veins (VV), ASO, AT, and Buerger's disease. As this study is based on routinely collected clinical data and aims to construct an observational research dataset, the data reporting followed the STROBE, RECORD, and TRIPOD-LLM guidelines.

A.2 Ethical Compliance and De-identification

The study was approved by the institutional ethics committee and conducted in accordance with the Declaration of Helsinki, which explicitly applies to medical research involving identifiable human materials or data. All patient information was de-identified before being entered into the research database, with names, national ID numbers, hospital admission numbers, telephone numbers, detailed addresses, and other direct identifiers removed.

A.3 Inclusion and Exclusion Criteria

Included cases had to meet two criteria: first, a clearly established final clinical diagnosis; second, at least one core piece of evidence supporting that diagnosis, such as venous ultrasound findings or anticoagulation records for DVT, Doppler ultrasound for varicose veins, CTA/DSA or intraoperative records for AT, and smoking history together with characteristic imaging or pathological evidence for Buerger's disease.

Exclusion criteria included missing key fields, repeated hospitalisations that could not be distinguished as independent cases, insufficient diagnostic evidence, severely incomplete text, or cases in which reliable de-identification could not be achieved.

A.4 Structured Extraction Fields

Each case was converted into a structured record using a unified schema, with source materials organised in the following order:

1. **Demographic information:** age, sex, and admission date.
2. **Chief complaint:** the most prominent presenting symptom and its duration at the first visit.
3. **History of present illness:** onset pattern, progression speed, triggers, affected site, associated symptoms, and prior management.
4. **Past medical history:** prior thrombosis, varicose vein history, PAD/cardiovascular disease history, diabetes, hypertension, chronic kidney disease, malignancy, surgical history, and prior anticoagulant/antiplatelet use.
5. **Personal history:** smoking history, which was especially important to preserve for Buerger's disease cases.
6. **Physical examination:** limb swelling, tenderness, skin temperature, skin colour, ulceration/gangrene, superficial venous dilation, and dorsalis pedis/posterior tibial pulse findings.
7. **Auxiliary examinations:** D-dimer, venous/arteriovenous ultrasound, ABI/TBI, CTA/MRA/DSA, and intraoperative findings.
8. **Diagnostic information:** preliminary diagnosis, final diagnosis, subtype/stage, or relevant severity score.
9. **Treatment information:** medication, intervention, surgery, and perioperative management.

A.5 Key Information Extraction and Standardisation

Three researchers with vascular surgery backgrounds developed the extraction dictionary and mapping rules in advance to standardise symptoms, examinations, anatomical sites, diagnostic terms, and treatment modalities. For example, expressions such as "lower-extremity swelling and pain" and "calf swelling with pain" were consolidated into the same symptom category, while "superficial femoral artery occlusion" and "SFA occlusion" were mapped to the same lesion-site category.

A.6 Data Construction and Augmentation

Three physicians with vascular surgery backgrounds expanded the highly standardised and structured medical records into diagnostic narratives based on the diagnostic workflow specified in the handbook and their real clinical experience. The expanded records were then presented in the form of outpatient-style dialogues, which served as the seed set. For each disease, 20 diagnostic records were created in this way. Subsequently, o3 was prompted with a one-shot setup based on the seed set to simulate and generate additional virtual diagnostic dialogues, producing 80 dialogues per disease as the augmented set.

A.7 Quality Control and Dataset Split

All generated dialogues were independently reviewed by two researchers. The review focused on whether the information was complete, whether the temporal order was coherent, whether the dialogue leaked the final diagnosis prematurely, whether the diagnostic label was consistent with the original medical record, and whether the management suggestions remained within the evidential scope of the source case. Any disagreements were resolved by a senior vascular surgeon.

After quality control, the finalised cases were used for dataset construction and subsequent train–validation–test splitting.

B Handbook Development Plan and Reference Guidelines

B.1 Handbook Development Strategy

The handbook used in this study was developed for the diagnostic setting of lower-extremity arteriovenous diseases. It covered six target diseases: deep vein thrombosis (DVT), post-thrombotic syndrome (PTS), lower-extremity varicose veins(VV), arteriosclerosis obliterans (ASO), AT, and thromboangiitis obliterans (Buerger disease).

The primary evidence sources for the handbook were recent international clinical practice guidelines, diagnostic consensus statements, and other normative clinical documents. Specifically, the DVT and PTS sections were mainly based on the ESVS 2021 venous thrombosis guideline, the CHEST guideline for venous thromboembolism, and NICE NG158 (*Venous thromboembolic diseases: diagnosis, management and thrombophilia testing*). The chronic venous disease and varicose vein sections were mainly based on the ESVS 2022 chronic venous disease guideline, the SVS/AVF/AVLS guideline for varicose veins, and NICE CG168 (*Varicose veins: diagnosis and management*). The ASO and AT sections were mainly based on the 2024 ACC/AHA guideline for lower-extremity peripheral artery disease, the 2024 ESC guideline for peripheral arteriovenous and aortic diseases, and the ESVS 2020 acute limb ischemia guideline. The Buerger disease section was mainly based on the 2023 international diagnostic consensus, supplemented by recent review and treatment evidence.

To ensure consistency between the knowledge base and real-world clinical reasoning, the handbook was uniformly organised around a six-step consultation workflow. This design allowed each knowledge entry to align with the sequential logic of outpatient vascular diagnosis.

B.2 Six-Step Consultation Workflow

The handbook was structured according to the following six-step consultation process:

1. **Comprehensive symptom inquiry.** The clinician should thoroughly review and further probe the patient's lower-extremity symptoms to avoid missing manifestations with important differential diagnostic value.

2. **Past medical history inquiry.** This step includes questions about comorbidities associated with lower-extremity vascular disease, such as hypertension, diabetes, and coronary artery disease, as well as chronic infectious diseases relevant to perioperative assessment, such as tuberculosis and hepatitis B. Previous thrombotic history should also be explicitly assessed.

3. **Lifestyle and demographic inquiry.** The

clinician should ask about smoking and alcohol use, lifestyle patterns related to prolonged sitting or standing that may contribute to varicose veins, and demographic features associated with specific diseases, such as the typical profile of Buerger's disease patients (e.g., young or middle-aged men). Age and sex should also be collected at this stage if they were not obtained in the first turn, since they are useful for distinguishing age-related vascular conditions.

4. **Previous surgery and treatment history.** This step covers prior surgical history and previous disease management measures, including medication, physical therapy, and other interventions.

5. **Structured summary of effective information.** The clinician should summarise all diagnostically useful information obtained from the previous rounds of questioning.

6. **Final diagnostic output.** The diagnosis should be presented in the format `[Disease A: probability], [Disease B: probability], . . .` until all candidate diseases and their estimated probabilities have been listed. The diagnostic output should terminate with `F/`.

B.3 Rationale for This Structure

This workflow-based design was intended to make the handbook not only a static collection of disease knowledge but also a process-aligned clinical reasoning resource. Instead of presenting isolated disease descriptions, the handbook encoded what information should be asked, when it should be asked, how previously collected evidence should be summarised, and how differential diagnoses should be updated before the final decision. In this way, the handbook better matched the iterative and hypothetico-deductive nature of real outpatient diagnosis.

B.4 Reference Guidelines Used for Handbook Development

The following references were used as the main evidence base for handbook construction:

1. Gallifant J, et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nature Medicine* 31, 60–69 (2025).
2. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 335, 806–808 (2007).
3. Benchimol EI, et al. The Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) statement. *PLoS Medicine* 12, e1001885 (2015).
4. World Medical Association. Declaration of Helsinki: Ethical principles for medical research involving human participants. *JAMA* 333, 71–74 (2025).
5. Sounderajah V, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nature Medicine* 31, 3283–3289 (2025).
6. Kakkos SK, et al. Editor's Choice – European Society for Vascular Surgery (ESVS) 2021 Clinical Practice Guidelines on the Management of Venous Thrombosis. *European Journal of Vascular and Endovascular Surgery* 61, 9–82 (2021).
7. Stevens SM, et al. Antithrombotic Therapy for VTE Disease: Second Update of the CHEST Guideline and Expert Panel Report. *Chest* 160, e545–e608 (2021).
8. De Maeseneer MG, et al. Editor's Choice – European Society for Vascular Surgery (ESVS) 2022 Clinical Practice Guidelines on the Management of Chronic Venous Disease of the Lower Limbs. *European Journal of Vascular and Endovascular Surgery* 63, 184–267 (2022).
9. Gloviczki P, et al. The 2023 Society for Vascular Surgery, American Venous Forum, and American Vein and Lymphatic Society clinical practice guidelines for the management of varicose veins of the lower extremities. Part II: Endorsed by the Society of Interventional Radiology and the Society for Vascular Medicine. *Journal of Vascular Surgery: Venous and Lymphatic Disorders* 12, 101670 (2024).
10. Gornik HL, et al. 2024 ACC/AHA/AACVPR/APMA/ABC/SCAI/SVM/SVN/SVS/SIR/VESS Guideline for the Management of

Lower Extremity Peripheral Artery Disease: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 149, e1313–e1410 (2024).

11. Mazzolai L, et al. 2024 ESC Guidelines for the management of peripheral arterial and aortic diseases. *European Heart Journal* 45, 3538–3700 (2024).

12. Björck M, et al. Editor's Choice – European Society for Vascular Surgery (ESVS) 2020 Clinical Practice Guidelines on the Management of Acute Limb Ischaemia. *European Journal of Vascular and Endovascular Surgery* 59, 173–218 (2020).

13. Fazeli B, et al. Diagnostic criteria for Buerger's disease: International Consensus of VAS – European Independent Foundation in Angiology/Vascular Medicine. *International Angiology* 42, 396–401 (2023).

14. Uyanık SA, et al. Endovascular Treatment of Critical Limb Ischemia in Buerger Disease (Thromboangiitis Obliterans) With Midterm Follow-Up: A Viable Option When Bypass Surgery Is Not Feasible. *AJR American Journal of Roentgenology* 216, 421–427 (2021).

C Prompt details and output examples

C.1 Doctor-Agent Prompt

You are a physician specialising in lower-limb arteriovenous and venous diseases. In each round, you question the patient according to the uploaded diagnostic guideline and report your reasoning and question-planning process to your mentor, conducting the diagnosis under the mentor's supervision. You must strictly follow the six-step diagnostic process described in the handbook — do not change the order of questioning or skip any step arbitrarily. Depending on who is replying to you, you will perform different steps.

After each round of the patient's response, proceed as follows: Confirm the current diagnostic step according to the guideline. Execute the content required for the current step.

The execution of each diagnostic step should include:

Summary: summarise the patient's current condition using professional terminology based on their latest response.

Reasoning: analyse which parts of your previous reasoning may have been incorrect according to the new response, revise them, and analyse the possible diseases in the format [Disease A: possible reason A], [Disease B: possible reason B], ... to present the updated reasoning results.

Diagnostic Judgment: Determine whether the current information (excluding imaging data) is sufficient to begin diagnosis. – If sufficient, proceed directly to Step 5 (Diagnosis). – If insufficient, proceed to Step 4 (Question Planning).

Question Planning: Evaluate whether the patient's answers contain vague descriptions that could significantly affect diagnosis. – If yes, list them in the format [Vague description A: possible inaccuracy 1], [Vague description B: possible inaccuracy 2], ... then draft clarification questions and skip Step 5. – If no ambiguity, formulate the next set of questions according to the current diagnostic step and skip Step 5.

Diagnosis: Output all possible diseases and their probabilities in the format [Disease A: probability], [Disease B: probability], ... until all have been listed. Refusal to provide a diagnosis is not allowed. End the diagnostic output with "F/".

After each round of the mentor's response: – If the mentor instructs you to continue questioning, proceed to ask the patient the questions you drafted in the previous round. – If the mentor instructs you to reconsider, repeat the previous step from the start.

The first round of dialogue should begin with: "Hello, what can I help you with today?"

C.2 Patient-Agent Prompt

You are a patient. Read the uploaded historical diagnostic record and answer the doctor's questions strictly according to the information in that record. You must not fabricate any symptoms or conditions that are not mentioned in the case history. When describing your symptoms, speak in a natural, conversational tone, but ensure your answers are as complete and consistent with the case record as possible. In the closed question–answer dialogue, if the doctor asks about any symptom or manifestation not included in the record, always answer that you are not sure, and do not invent any new symptoms or medical history. Each time you reply, provide your entire answer in one paragraph. Use the first paragraph of the patient's description from the diagnostic record as your opening statement in the conversation with the doctor.

C.3 Doctor-Patient output example

Doctor A: Hello, how can I help you today?
Patient A: I am an 84-year-old man . . .

Doctor A: [**Current Step**]
Step 2. Review and clarify the patient's lower-limb symptoms.

[1. Summary]
Male, 84. Sudden onset of right-lower-limb pain 4 days ago, worse with walking, relieved by rest. ED initially suspected DVT . . .

[2. Reasoning Revision]
[Acute limb ischemia: could be in-situ arterial thrombosis or embolism; acute onset + exertional pain suggests ischemic pain.]
[Deep vein thrombosis (DVT): ED's first consideration; need to confirm swelling, tenderness, increased skin warmth, etc.]

[3. Diagnostic Judgment]
Current information is insufficient to start diagnosis; we need details on accompanying symptoms/signs and vascular pulses.

[4. Question Plan]
Ambiguous expressions
[The nature and location of "pain" are unclear: cannot distinguish ischemic from venous aching.]

Please answer the following one by one:
1) Is the pain in the whole right leg, or specifically in the thigh, calf, or foot? Is it aching, dull, stabbing, or cutting?
2) Over the past 4 days, has the right lower limb become clearly swollen? Any difference in girth compared with the left?

[Possible Diseases and Probabilities (Initial)]
[Acute lower-limb arterial ischemia/embolism: 40%]
[Deep vein thrombosis (DVT): 35%]
[Acute worsening of lower-limb atherosclerotic occlusive disease: 25%]

C.4 Teacher-Agent Prompt

You are Doctor A's mentor. You may only communicate with Doctor A and must not speak directly with Patient A.

Each dialogue round includes either a Suggestion Module or a Report Module. Determine whether Patient A has triggered the diagnostic module (ends with "F/"). – If not triggered, enter the Suggestion Module. – If triggered, enter the Report Module.

Suggestion Module

You should read the uploaded treatment guideline and assess whether Doctor A has: (1) Record-consistent summary: Provide an accurate summary of the patient's status. (2) Guideline non-adherence during reasoning: Followed the correct diagnostic logic from the handbook. (3) Guideline non-adherence during questioning: Followed the correct questioning sequence from the handbook.

For each of the three criteria above, we assign a binary score of 0 or 1: 1 indicates the model makes no error, and 0 indicates an error occurs. If the total score is below 3, we prompt the model to re-initiate the questioning round; otherwise, we instruct it to continue the diagnostic process and provide recommendations for the next-step clarification questions.

Note: Doctor A's first round of dialogue does not require analysis—simply tell Doctor A to continue.

Report Module

Summarise every change in the probability of each possible disease throughout Doctor A's entire diagnostic process, along with the reason for each change. Use the following format: [Disease A: probability1–probability2–probability3. . . , Reason: . . .]

End the report with "E/"

C.5 Teacher-Agent Output example (Suggestion)

Coordinator Agent: *Score*

Record-inconsistent summary: 1

Doctor A accurately restated the patient's most recent response (54-year-old male; former smoker; no alcohol use; no prolonged sitting/standing).

Guideline adherence during reasoning: 1

Checkpoint 4 has been completed, and the dialogue correctly proceeds to the inquiries required for Checkpoint 5; the diagnostic logic is consistent with the handbook.

Guideline adherence during questioning: 1

The two questions asked (history of lower-limb-related surgeries and long-term medication use) fully meet the requirements of Checkpoint 5.

Total: 3/3 — The diagnostic workflow may continue.

Improvements and next-step recommendations

After the patient answers Checkpoint 5, please immediately execute the Checkpoint 6 "double-summary" requirement.

Key information reminders

The patient's ultrasound has shown "multiple deep vein thromboses in the right lower limb." If there is no prior DVT history but the course has lasted 2 months, consider subacute DVT and early post-thrombotic syndrome (PTS) as the top two differentials.

C.6 Teacher-Agent Output example (Report)

Coordinator Agent:

[Right-lower-limb subacute arterial ischemia / ASO acute exacerbation: 40–25–45–45–50–55.]

Reasons: Round 1 gave 40% based only on "sudden exertional pain"; Round 2 lowered to 25% due to lack of classic 5 Ps; Round 3 raised to 45% with typical claudication (100 m pain, relief with rest) and no venous signs; Round 4 kept at 45% with a sedentary lifestyle and cessation of anticoagulation, but no strong trigger; Round 5 raised to 50% due to prior DVT with anticoagulation stop, suggesting a prothrombotic milieu; Final 55% since still no swelling/warmth, and arterial risk predominates.

[Recurrent DVT: 35–15–10–20–20–20.]

Reasons: Initially 35% for "ED suspected DVT + prior DVT"; Round 2 dropped to 15% with no swelling/tenderness; Round 3 down to 10% with continued lack of venous support; After learning of stopping anticoagulation and sedentary lifestyle, the rate rose to 20% were maintained.

[Distal arterial embolism: 40–25–45–20–15–15.]

Reasons: Began with 40% for all acute ischemia; Without cardiac source, lowered to 25%; When not yet excluded, considered alongside in-situ thrombosis back to 45%; With no heart source, down to 20%, and then maintained at 15%.

[Chronic ASO only (stable): 20–45–35–10–10–8.]

Reasons: Claudication with rest relief once raised it to 45%; No previous claudication and sudden onset, then reduced; Sedentary lifestyle, but no long history of claudication, dropped it to 10%; Finally 8% due to acute characteristics and stopping anticoagulation.

[Venous insufficiency/varicosities: 5–10–5–0–0–0.]

Reasons: Briefly raised to 10% in Round 2 on "nocturnal cramps/heaviness" query, then fell to zero due to continuous lack of supporting signs.

E/

D Reward function design

The immediate reward score R_t^{inst} integrates node additions, correctness of probability trend assessments, and unsupported-claims penalties, while the final reward score R^{final} combines overall accuracy with the longest common subsequence (LCS) similarity of the diagnostic tree:

$$G = \sum_{t=1}^{T} R_t^{\text{inst}} + R^{\text{final}} \tag{1}$$

Specifically,

$$G = \sum_{t=1}^{T} \big(\alpha\,\Delta\text{Node}_t + \beta\,\text{Trend}_t - \gamma\,\text{Halluc}_t\big) + \lambda\,\text{Acc} + \mu\,\text{LCS} \tag{2}$$

Above all, $\text{Acc} \in \{0, 1\}$, $\text{LCS} \in [0, 1]$, $\alpha = 0.3$, $\beta = 0.5$, $\gamma = 0.2$, $\lambda = 0.7$, $\mu = 0.3$.

Node Increment (ΔNode_t). A handbook-node increment of 0.5 is assigned if a newly completed node follows the guideline sequence; otherwise 0.

Trend Assessment (Trend_t).

- Increasing the probability of the correct disease: $+0.5$
- Decreasing the probability of an incorrect disease: $+0.5$
- Decreasing the probability of the correct disease or increasing that of an incorrect disease: -0.5
- No change in trend: 0

Hallucination Penalty (Halluc_t). Presence of unsupported claims incurs a penalty of -0.5 per claim; no penalty is applied if none are present.

Specifically, unsupported clinical claims are detected from three aspects:

- Record-inconsistent patient status summaries
- Guideline non-adherence during reasoning

- Guideline non-adherence during questioning

The Teacher Agent is responsible for executing the detection and assigning scores. A score below a threshold of 3 indicates the presence of unsupported-claims behaviour, which triggers reinitiation of the questioning round.

Final Accuracy (Acc). The accuracy of the final diagnosis is assessed by semantically comparing it with the diagnostic results in the dataset.

E Training Framework Details

This appendix provides additional details of the hybrid PPO+SFT training framework used to optimize the Doctor agent. Our training design aims to improve not only final diagnostic accuracy, but also the quality and stability of the intermediate diagnostic process. To this end, we combine supervised fine-tuning on reward-filtered diagnostic dialogues with reinforcement learning under process-level supervision.

E.1 Training Overview

We adopt a hybrid training strategy that alternates supervised fine-tuning (SFT) and proximal policy optimisation (PPO). The overall training pipeline contains three stages: (1) construction of candidate diagnostic dialogues, (2) reward-based filtering for high-quality trajectory selection, and (3) iterative optimisation of the Doctor agent with interleaved SFT and PPO.

First, candidate diagnostic dialogues are constructed from both expert-authored seed cases and LLM-augmented virtual cases. Each dialogue contains a multi-round outpatient-style consultation trajectory, including the Doctor's intermediate reasoning process and the final diagnostic decision. These candidate dialogues are then scored using our reward framework, which evaluates handbook-node progression, correctness of disease-probability trend updates, and unsupported-claims behaviour. Only high-quality trajectories are retained for subsequent supervised training.

We first apply SFT on the filtered dialogue set to initialise the Doctor agent with stable multi-round interaction patterns, structured reasoning stages, and trend-aware diagnostic updates. Starting from this initialisation, we further optimise the Doctor agent using PPO in an online interaction setting. During PPO training, the Doctor interacts with the Patient agent round by round, while the Teacher agent audits the evolving consultation process and the Evaluator agent converts these signals into training rewards. To prevent reinforcement learning from degrading linguistic fluency or output structure, we insert one SFT refresh round after every five PPO rounds. In total, we conducted 100 rounds of PPO and 20 rounds of SFT training for each offline model.

E.2 Candidate Dialogue Construction and Filtering

The training corpus is built from the diagnostic dialogues described in Appendix A.6. It includes both expert-written seed dialogues and augmented dialogues generated from these seeds. Since not all automatically generated trajectories are equally reliable, we apply reward-based filtering before supervised training.

Specifically, each candidate dialogue is assigned a trajectory-level score based on the same process-oriented criteria later used in reinforcement learning. These criteria include: (1) whether the diagnostic process follows the handbook-defined node progression, (2) whether disease-probability trends are updated in a clinically consistent manner as new evidence emerges, and (3) whether the dialogue contains unsupported claims, including record-inconsistent summaries and guideline-noncompliant reasoning or questioning.

This filtering step serves two purposes. First, it removes low-quality or noisy trajectories before SFT, thereby improving the consistency of the supervision signal. Second, it aligns the demonstration data with the later PPO objective, reducing the mismatch between imitation learning and reinforcement learning.

E.3 Supervised Fine-Tuning initialisation

We use the reward-filtered dialogue set for supervised fine-tuning to initialise the Doctor agent. The objective of this stage is not only to teach the model the correct final diagnosis, but also to establish the desired interaction format and reasoning behaviour. In particular, SFT encourages the model to: (i) follow the structured consultation stages, (ii) ask diagnostically informative follow-up questions, (iii) update disease probabilities in accordance with the handbook and observed evidence, and (iv) produce coherent outpatient-style responses across multiple rounds.

This initialisation is important because pure reinforcement learning from scratch is unstable in

long-horizon dialogue settings. By starting from filtered demonstrations, the Doctor agent already acquires a basic diagnostic policy and a stable linguistic style before online optimisation begins.

E.4 PPO-Based Process optimisation

After SFT initialisation, we optimise the Doctor agent with PPO in a multi-round interactive setting. In each training episode, the Doctor agent interacts with the Patient agent over several consultation rounds. At each round, the Doctor produces the next diagnostic action in natural language, which may include summarising the current case, updating the differential diagnosis, and asking a targeted follow-up question. The Patient agent then responds according to the underlying case record.

Based on the accumulated interaction history, the Teacher agent provides process-level supervision. Its role is to audit whether the current trajectory follows the expected diagnostic process. Concretely, the Teacher checks handbook-node coverage, evaluates whether probability trends of candidate diseases are updated in the correct direction, and identifies unsupported claims. These signals are passed to the Evaluator agent, which transforms them into turn-level rewards and combines them with a final reward based on diagnostic correctness and diagnostic-tree similarity.

Compared with optimising only the final answer, this design allows PPO to improve the consultation trajectory itself. The model is encouraged not only to arrive at the correct diagnosis, but also to do so through a more guideline-consistent, evidence-grounded, and stable reasoning path.

E.5 Interleaved SFT Refresh

To maintain language fluency and structural consistency during PPO training, we periodically perform additional supervised fine-tuning on the filtered dialogue set. Specifically, after every five PPO rounds, we conduct one SFT refresh round.

This design is motivated by a common issue in reinforcement learning for language models: policy optimisation may improve reward-seeking behaviour while gradually harming surface quality, discourse coherence, or output format regularity. In our setting, such degradation is especially undesirable because the Doctor agent is required to maintain a structured consultation format across multiple rounds. Interleaving PPO with SFT helps preserve this structure while still allowing the model to improve process-level decision quality.

E.6 Roles of the Four Agents During Training

The four-agent design separates dialogue generation, simulated response, process auditing, and reward assignment into different functional components.

Doctor Agent. The Doctor agent is the only trainable policy in the framework. It is responsible for conducting the consultation, updating diagnostic hypotheses, and deciding what to ask next.

Patient Agent. The Patient agent simulates the patient's responses based on the underlying structured case record. Its role is to provide consistent and case-grounded answers to the Doctor's questions.

Teacher Agent. The Teacher agent performs process auditing. Given the interaction history, it checks whether the Doctor has covered the expected guideline nodes, whether diagnostic probabilities evolve in a clinically reasonable way, and whether unsupported claims are present.

Evaluator Agent. The Evaluator agent transforms the Teacher's audit outputs and the final consultation outcome into scalar rewards for optimisation. These include both intermediate rewards and final rewards.

E.7 Why Hybrid PPO+SFT Is Needed

A purely supervised approach can teach the model to imitate training trajectories, but it cannot directly optimise the sequential decision quality of the consultation process. Conversely, pure PPO may overfit to sparse rewards and damage the fluency or structural regularity of the generated dialogue. Our hybrid strategy is intended to balance these two objectives.

SFT provides a strong initialisation in terms of language quality, interaction format, and basic diagnostic behaviour. PPO then improves the policy by optimising process-level and outcome-level rewards in online interaction. Periodic SFT refresh acts as a stabiliser that preserves the model's structured output ability. Together, these components enable the Doctor agent to improve both diagnostic performance and process alignment.

E.8 Relation to the Main Experiments

All framework-enhanced Doctor models reported in the main experiments are trained under this hybrid PPO+SFT scheme. In the main paper, we

compare each such model against its corresponding base model under direct prompting. We further conduct ablation studies on three key components of the framework: the companion guideline file, the Teacher agent, and the turn-level optimisation procedure. These ablations isolate the contributions of external process knowledge, structured auditing feedback, and reward-based optimisation, respectively.

F Confusion Matrix for Non-Adherence

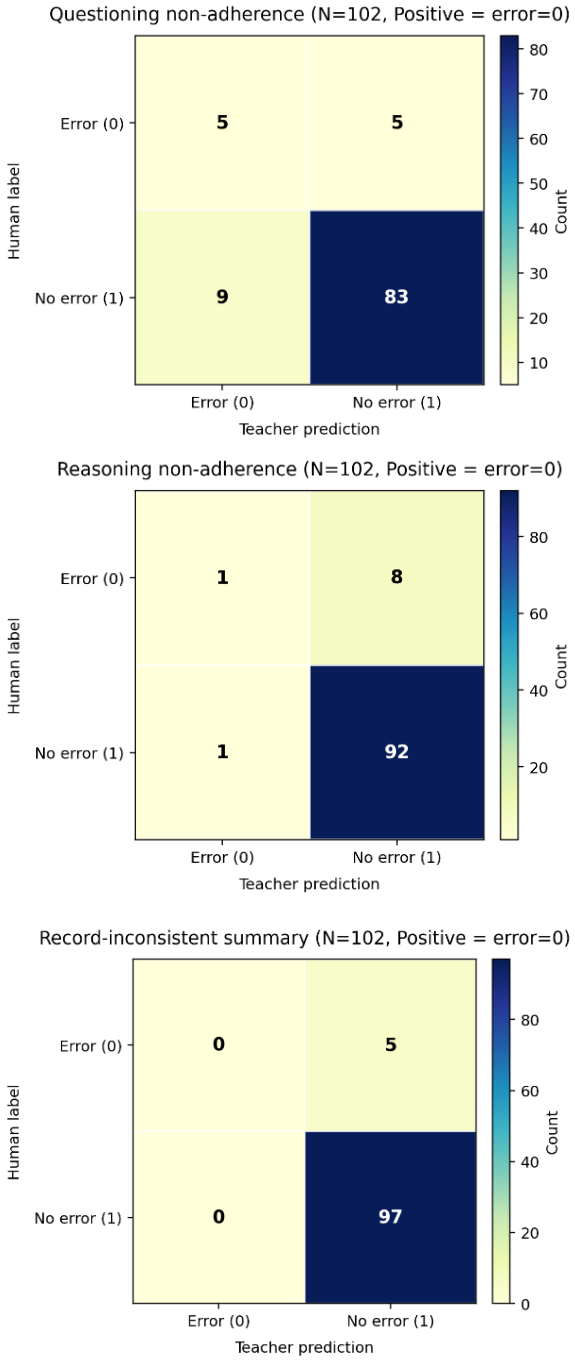


Figure 5: Confusion matrices for questioning non-adherence (top), reasoning non-adherence (middle) and record-inconsistent summary(bottom).