ProActIS: Boosting LLMs' Proactive Information Seeking in Dynamic Clinical Workflows with a Lightweight Exam Proposer

Anonymous ACL submission

Abstract

Large language models (LLMs) have shown strong promise as clinical reasoning agents, yet most benchmarks assume complete evidence availability before reasoning begins. However, real clinical practice unfolds through dynamic clinical workflows with progressive disclosure of patient evidence, requiring clinicians to decide what evidence to acquire next from partial and evolving patient states. This mismatch poses a key barrier to reliable and wide deployment of LLM-based clinical agents. To address this gap, we introduce **ProActIS**, a lightweight state-conditioned examination proposer for proactive evidence acquisition. Given an evolving patient state, ProActIS ranks standardized laboratory, microbiology, and radiology examinations, together with `STOP`, through structured patient-state decomposition and action-card encoding. We construct a leakage-controlled benchmark from MIMIC-IV, comprising 81,960 hospital admissions and 860,099 turn-level patient-state/examination pairs over a 129-examination action bank. On full-bank next-examination ranking, ProActIS achieves 0.602 Recall@1, 0.886 Recall@5, and 0.725 MRR, substantially outperforming frequency and LLM baselines. In LLM-assisted dynamic workflows, ProActIS improves evidence coverage and trajectory alignment, highlighting proactive evidence acquisition as a distinct bottleneck and a critical step toward reliable clinical language agents in real world.

1 Introduction

Clinical reasoning rarely begins with a complete view of the patient (Kassirer, 1983). A clinician first observes an initial presentation, forms tentative hypotheses, and then gathers additional evidence through laboratory tests, microbiology studies, and imaging (Elstein and Schwarz, 2002). This iterative process is central to diagnosis: each new finding can confirm or rule out possible explanations (National Academies of Sciences, Engineering, and Medicine, 2015). While LLMs have shown strong performance on medical question answering and patient-record reasoning, these evaluations usually measure reasoning over evidence that has already been collected, rather than the ability to decide which evidence should be collected next.

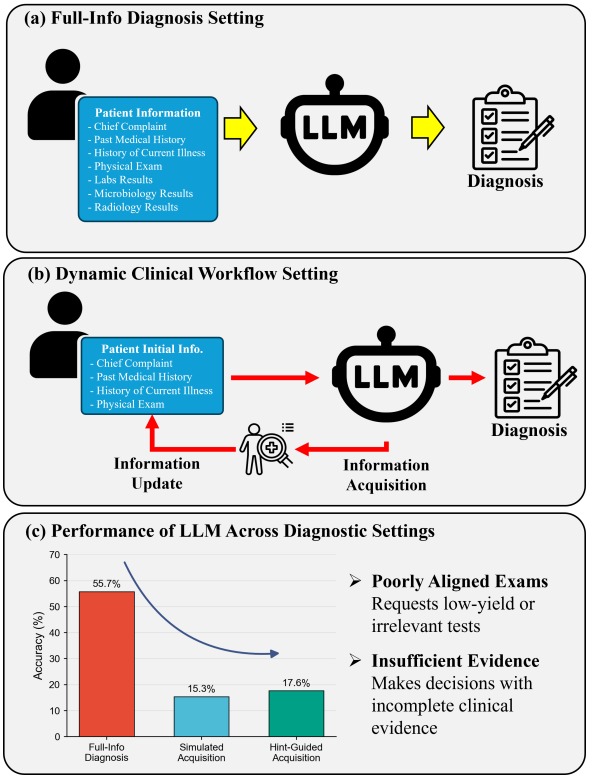

Figure 1: Static full-information diagnosis versus dynamic clinical evidence acquisition. Dynamic settings require the agent to request evidence before final diagnosis and expose a substantial drop in performance.

When LLMs are placed in this dynamic setting, the difficulty shifts from only interpreting available evidence to actively acquiring useful evidence. Specifically, the same diagnostic LLM performs much better when complete patient information is provided than when it must request examinations sequentially. This gap suggests that dynamic

failures are not only failures of final diagnostic reasoning, but also failures of proactive evidence acquisition. An agent may request clinically plausible examinations while still missing tests that are timely, non-redundant, and aligned with the current patient state.

Prior work leaves this middle step underexplored. Medical QA and patient-record benchmarks emphasize reasoning from provided evidence(Schmidgall et al., 2025; Fleming et al., 2023),, while interactive diagnosis methods often optimize dialogue behavior, disease prediction, or end-to-end task success(Fan et al., 2025). We instead focus on state-conditioned examination proposal before final diagnosis. We formulate this problem as turn-level ranking over a standardized examination action bank: the model observes the initial patient context at each turn, previous examination requests, and returned findings, then ranks candidate examinations together with the decision to stop evidence acquisition. This formulation isolates examination selection from downstream diagnosis generation while preserving the sequential and state-dependent nature of clinical evidence acquisition.

We introduce **ProActIS**, a lightweight state-conditioned examination proposer for LLM clinical agents. ProActIS encodes the evolving patient state and standardized action cards, then scores candidate examinations with a structured state-action ranking model. It handles STOP within the same action interface and uses trajectory-aware training and evaluation to accommodate ambiguity in valid examination order. Across offline next-examination alignment and assisted dynamic workflows, ProActIS improves evidence acquisition over frequent-exam and LLM baselines. As a standalone ranker, it better identifies state-aligned next examinations from the full action bank; as plug-in support for LLM agents, it improves the coverage and alignment of dynamically requested examinations. These results show that proactive information seeking is a distinct capability for clinical language agents.

Our contributions are threefold:

1. We identify proactive clinical information seeking as a distinct bottleneck for LLM-based diagnostic agents, separating evidence acquisition from final diagnostic reasoning.

2. We build a leakage-controlled benchmark with 82K examination trajectories and 860K patient-state/action pairs over laboratory, microbiology, and radiology actions.

3. We propose ProActIS, a lightweight state-conditioned examination proposer, and perform extensive experiments to show it as both a full-bank ranker and plug-in support for improving LLM-based dynamic diagnostic workflows.

2 Related Work

Medical QA and fixed-context clinical reasoning. A growing body of research on medical LLMs and clinical agents has relied on question answering and examination-style benchmarks, including MedQA(Jin et al., 2021), PubMedQA(Jin et al., 2019), MultiMedQA(Bani-Harouni et al., 2025), etc., to evaluate clinical and biomedical reasoning. These benchmarks assess medical knowledge, biomedical evidence interpretation, and reasoning over provided records, passages, or patient-specific contexts. However, they generally adopt a fixed-context formulation in which the evidence needed to answer the question is assumed to be available at inference time. Consequently, they evaluate a model's ability to reason over given information rather than its ability to determine which additional evidence should be acquired. ProActIS addresses this complementary setting by studying proactive examination selection under partial and evolving patient information.

Contrastive learning for clinical representation and recommendation. Contrastive learning has been effective for longitudinal and heterogeneous patient modeling. EBCL(Jeong et al., 2024) shows that event-centered contrastive objectives can learn patient embeddings sensitive to temporal clinical progression, while ConMEHR(Yin et al., 2025) aligns structured and unstructured EHR signals in a shared latent space. ARCI(Hadizadeh Moghaddam et al., 2024) and DGCL(Li et al., 2023) apply contrastive learning to sequential prescription or medication recommendation, where models learn not only patient states but also their compatibility with candidate clinical actions. However, their action spaces mainly concern medications rather than diagnostic evidence acquisition. ProActIS addresses this gap by aligning evolving patient states with standardized examination concepts, making patient-state-to-exam matching the core learning problem.

Sequential diagnosis and clinical decision making. Recent work has also studied diagnosis as an interactive or sequential decision process. AMIE(Tu et al., 2025) evaluates diagnostic dialogue, while LA-CDM(Bani-Harouni et al., 2025) and DiagAgent (Qiu et al., 2025) model multi-step clinical decision making through examination requests, result observation, and diagnosis in constrained or simulated environments. These approaches highlight the importance of evidence acquisition, but often optimize full dialogue behavior, simulated environment policies, or end-to-end diagnostic success. As a result, the compatibility between a specific evolving patient state and each concrete examination action is usually implicit in the agent policy or constrained by a limited action space. ProActIS focuses on this explicit turn-level matching problem, enabling direct evaluation of examination selection and lightweight integration with LLM-based clinical agents.

3 Dataset Construction

We construct the dataset from MIMIC-IV(Johnson et al., 2020) discharge summaries and structured laboratory, microbiology, and radiology records. Each case corresponds to one hospital admission. To approximate information available near admission, the model-visible Patient Medical Record is limited to chief complaint, age at admit, race, past medical history, history of present illness, and physical exam. Diagnosis fields are used only for cohort filtering and reference analysis.

We reduce diagnosis leakage by removing discharge-oriented and result-revealing sections, including discharge diagnoses, hospital course, discharge medications, and reported laboratory or imaging results. Cases are further excluded when the final diagnosis is explicitly mentioned in visible patient fields, using string matching, clinical synonym matching, and LLM-assisted review.

Reference trajectories are derived from timestamped laboratory, microbiology, and radiology records within a 24-hour window around the case index time. Raw item-level records were normalized into order-level examination actions using LLMs guided by the LOINC terminology(Regenstrief Institute, Inc., 2025); the resulting action set was manually reviewed and frozen across data splits.

The final action bank contains 129 standardized examination actions: 31 laboratory, 64 microbiology, and 34 radiology actions. We add a terminal STOP action, yielding 130 actions. Each trajectory is converted into turn-level examples where the state contains the initial patient context and prior examination-result history, and the target is the next reference action. The final dataset contains 81,960cases from 60,039 patients and 860,099 turn-level pairs, split at the patient level into train, validation, and test sets with a 0.90/0.05/0.05 ratio. These trajectories are used as observed-practice references rather than gold-standard clinical policies.

Table 1: Dataset statistics after leakage filtering, action normalization, and turn-level conversion.

Statistic	Count
Patients	60,039
Cases / notes	81,960
Patient-state/action pairs	860,099
Lab action pairs	528,535
Microbiology action pairs	132,716
Radiology action pairs	116,888
STOP action pairs	81,960
Total actions, incl. STOP	130

4 Methodology

ProActIS is a lightweight state-conditioned exam proposer that ranks standardized examination actions or STOP from patient state. It implements by a contrastive dual-encoder model(Radford et al., 2021) with structured patient-state channels and action-card representations, trained to prefer the reference next action over competing legal candidates.

4.1 Problem Definition

For each patient, we define the turn-level state as

$$S_t = (S_0, p_t),$$

where S_0 is the initial static patient context, including chief complaint, demographics, medical history, current presentation, and physical examination, and p_t is the ordered prefix of previously requested examinations and returned results.

The model selects from the standardized action bank $\mathcal{A}$ constructed in Section 3, which contains laboratory, microbiology, radiology, and STOP actions. At turn t, ProActIS ranks the legal candidate set $\mathcal{C}_t \subseteq \mathcal{A}$, where completed non-STOP examinations are removed and STOP remains available until termination.

Each instance is supervised by the immediate next reference action $e^*_{t+1} \in \mathcal{C}_t$. ProActIS is there-

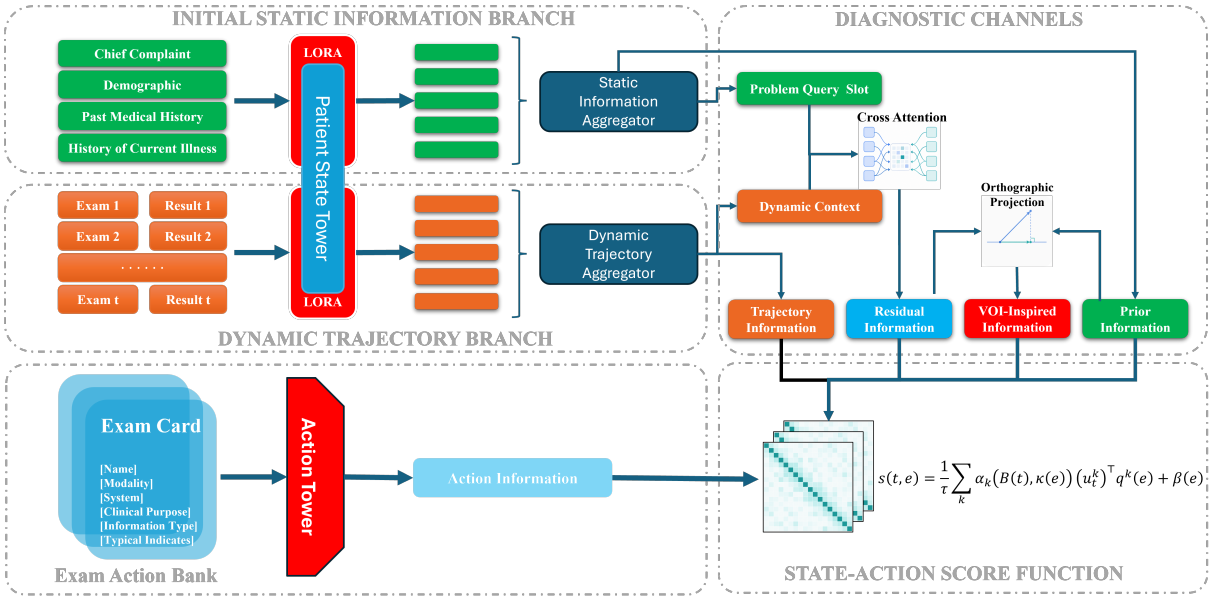

Figure 2: ProActIS architecture: structured state encoding and action-card encoding for unified exam/STOP ranking.

fore trained as a state-action ranker for proactive information seeking: it outputs a ranked list of standardized actions and STOP, but does not interpret examination results or produce the final diagnosis.

4.2 Model Architecture

ProActIS uses a contrastive dual-encoder architecture only with 0.3B parameters. The patient encoder maps the turn-level state $S_t = (S_0, p_t)$ into structured state representations, while the action encoder maps each candidate examination $e \in \mathcal{C}_t$ into action representations.

The patient state is represented by four learned diagnostic channels:

$$u_t = \{u_t^{\text{prior}}, u_t^{\text{traj}}, u_t^{\text{res}}, u_t^{\text{voi}}\}.$$

The prior channel encodes baseline examination plausibility from the static patient context S_0. The trajectory channel encodes the ordered examination-result prefix p_t. The residual channel is obtained through static-dynamic interaction, allowing latent problem representations from S_0 to be updated by observed examination evidence. The VOI-inspired channel captures remaining information need after accounting for the evidence-conditioned state. These channels serve as architectural inductive biases rather than fully disentangled clinical variables.

Each standardized examination is represented by a frozen action card constructed in Section 3. The card contains structured natural-language fields describing the examination identity, modality, body system, clinical purpose, and typical information type, without patient-specific information. The action encoder first produces a shared action representation and then maps it into the same four channel spaces:

$$q^{\text{prior}}(e), \quad q^{\text{traj}}(e), \quad q^{\text{res}}(e), \quad q^{\text{voi}}(e).$$

This design allows each candidate action to be evaluated along the same four decision factors as the patient state: initial plausibility, trajectory consistency, residual uncertainty, and remaining information value.

4.3 State-Action Score Function

The scoring function maps the structured patient and action representations to a contrastive ranking score. Because examination selection is a state-dependent utility ranking problem rather than simple semantic matching, ProActIS uses unnormalized channel-wise inner products instead of L2-normalized cosine similarity.

For each channel $k \in \{\text{prior}, \text{traj}, \text{res}, \text{voi}\}$, ProActIS computes

$$z_k(t,e) = (u_t^k)^\top q^k(e),$$

where each logit measures the compatibility between candidate examination e and one diagnostic factor of the current state. The final score is a calibrated mixture of the four channel logits:

$$s(t,e) = \frac{1}{\tau}\sum_k \alpha_k(B(t),\kappa(e)) z_k(t,e) + \beta(e).$$

Here, $\tau > 0$ is a learned temperature, $\beta(e)$ is a learned action bias, $\kappa(e)$ denotes the action kind, and $B(t)$ is a prefix-derived workflow bucket. The channel weights are normalized as

$$\alpha_k(B(t), \kappa(e)) = \frac{\exp \Gamma_{B(t),\kappa(e),k}}{\sum_{k'} \exp \Gamma_{B(t),\kappa(e),k'}},$$

where Γ is a learned calibration table indexed by workflow bucket, action kind, and channel.

This calibration allows different diagnostic factors to dominate under different workflow stages and action types. For example, STOP may use a different channel mixture from examination actions because it represents information sufficiency rather than additional evidence acquisition. The same score $s(t, e)$ is used for training and inference, and all legal examinations and STOP are compared in one unified score space.

4.4 STOP-Aware Termination Modeling

STOP differs from laboratory, microbiology, and radiology actions because it terminates evidence acquisition rather than requesting additional information. Premature STOP may end the interaction before useful evidence is collected, whereas underranking STOP may lead to unnecessary examinations. ProActIS therefore models STOP both as a ranked candidate in the unified action space and as an auxiliary supervision signal for information sufficiency.

STOP is included in the standardized action bank $\mathcal{A}$, remains legal in $\mathcal{C}_t$ before termination, and receives a compatibility score from the same scoring function:

$$s(t, \mathrm{STOP}).$$

Thus, termination competes directly with further examination acquisition; the model does not first make a binary stop/continue decision and then choose an examination.

During training, we add an auxiliary STOP objective with label

$$y_t^{\mathrm{STOP}} = \mathbf{1}[e_{t+1}^* = \mathrm{STOP}],$$

which encourages the shared patient representation to encode evidence sufficiency. This objective shapes the residual and VOI-inspired channels but does not replace the state-action ranker. At inference time, ProActIS does not apply a separate STOP threshold or use the auxiliary STOP probability to override the ranked list. STOP is selected only through the unified ranking score, so the deployed model remains a single ranker over examinations and STOP.

4.5 Future-Aware Negative Construction

Reference examination trajectories provide ordered supervision, but the immediate next examination is not always the only trajectory- consistent action under a partially observed state. If all non-target actions are treated as equally negative, training may suppress examinations that appear later in the same reference trajectory. ProActIS therefore keeps exact next-action supervision strict while reducing negative pressure on later trajectory-consistent actions.

For each turn t, the exact positive is the immediate next reference action e_{t+1}^*. We define the future-valid set as legal actions that appear later in the same trajectory:

$$\mathcal{F}_t = \{e_j : j > t + 1\} \cap \mathcal{C}_t.$$

Future-valid actions are not treated as additional positives; the model is still trained to rank e_{t+1}^* highest. However, they are protected from strong negative pressure because they remain consistent with the reference trajectory.

For hard-negative learning, ProActIS mines negatives only from invalid legal candidates:

$$\mathcal{I}_t = \mathcal{C}_t \setminus \left(\{e_{t+1}^*\} \cup \mathcal{F}_t\right).$$

This future-aware construction focuses hard-negative learning on actions that are legal but neither the immediate next action nor a later reference action. Future-aware labels are used only during training and offline evaluation; they are not available at inference time.

4.6 Training Objective and Inference

The primary training signal is exact next-action ranking over the legal candidate set. For each turn t, the immediate next reference action e_{t+1}^* is the only positive target, and the model is trained with full-bank cross entropy over $\mathcal{C}_t$. This preserves strict supervision from the recorded trajectory, including direct competition between examination actions and STOP.

To reduce excessive negative pressure on later trajectory-consistent examinations, ProActIS blends the exact cross-entropy loss with a future-relaxed cross entropy:

$$\mathcal{L}_{\mathrm{ranker}} = (1 - \lambda_{\mathrm{F}})\mathcal{L}_{\mathrm{CE}} + \lambda_{\mathrm{F}}\mathcal{L}_{\mathrm{FCE}}.$$

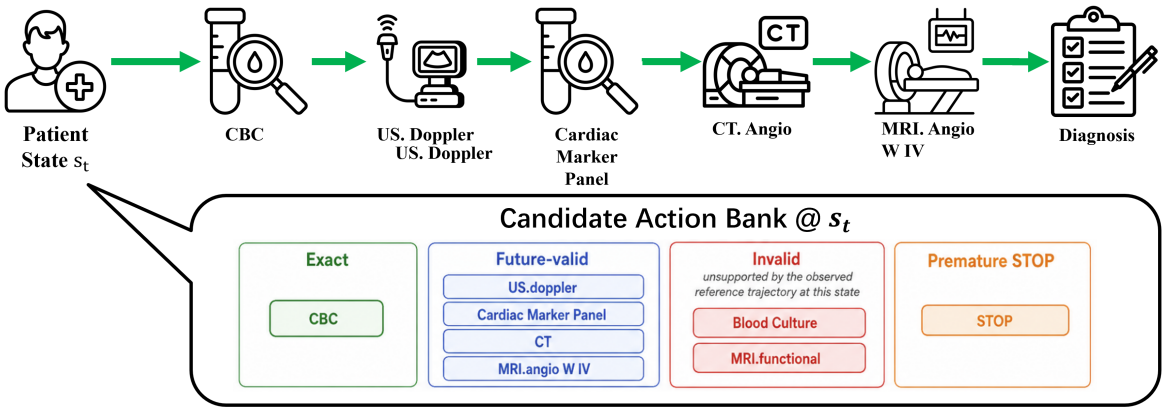

Figure 3: Architecture overview of ProActIS. The state tower encodes static patient information and prior examination-result history into structured diagnostic channels, while the action tower encodes standardized examination cards for unified ranking over examinations and STOP.

Here, $\mathcal{L}_{\mathrm{FCE}}$ discounts non-STOP future-valid examinations in the denominator while keeping e^*_{t+1} as the only positive target. STOP is not discounted, preventing later terminal actions from becoming easier to rank before the reference stopping point.

The full objective combines this ranker loss with the future-filtered hard-negative loss and the auxiliary STOP sufficiency loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{ranker}} + \lambda_{\mathrm{HN}}\mathcal{L}_{\mathrm{HN}} + \lambda_{\mathrm{STOP}}\mathcal{L}_{\mathrm{STOP}}.$$

When class balancing is enabled, per-target weights multiply the corresponding per-instance ranker loss to reduce domination by frequent examination categories.

At inference time, ProActIS receives only S_0, the observed prefix p_t, and the legal candidate set $\mathcal{C}_t$. It does not observe future trajectory labels, exact targets, or final diagnoses. The model computes $s(t, e)$ for every legal candidate and returns a ranked list. And STOP is selected only if it receives the highest unified ranking score.

5 Experiments and Results

5.1 Experimental Setup and Metric

We evaluate ProActIS in two settings. The primary setting is offline turn-level examination proposal, where each instance contains the current patient context, previous examination-result history, and the legal candidate set. The model ranks all candidates, including STOP, against the immediate next reference action. This setting directly measures state-conditioned proactive examination selection. We also evaluate ProActIS as plug-in support for an LLM-based diagnostic workflow, where the LLM receives either no proposer support or ProActIS Top-1/Top-K suggestions before requesting the next examination. This tests whether the proposer can improve dynamic evidence acquisition without producing the final diagnosis itself.

The main evaluation target is examination selection quality. For offline ranking, we report Recall@K, MRR, and exact-or-future@1 over turn-level states, with STOP included as a candidate and as the target at terminal turns. Recall@K and MRR measure strict agreement with the immediate next reference action, but diagnostic trajectories are not always uniquely ordered: an examination predicted one step early may still be clinically meaningful if it appears later in the same reference trajectory. We therefore use exact-or-future@1 to separate clearly unsupported top predictions from temporally misordered but trajectory-consistent examinations. For dynamic workflows, we report hit rate, case-level recall, Jaccard overlap, ordered trajectory similarity, and matched reference examinations. Diagnosis accuracy is reported only as contextual evidence for downstream diagnostic quality.

5.2 Evidence Acquisition in Dynamic Workflows

We evaluate DeepSeek V4 Pro(DeepSeek-AI, 2026) under a full-information control setting and several dynamic acquisition settings. The full-information setting provides the complete patient record before diagnosis, whereas dynamic settings start from initial patient information and require sequential examination requests. Simulated and Hint-Guided Acquisition test whether interaction feedback alone can support effective evidence seek-

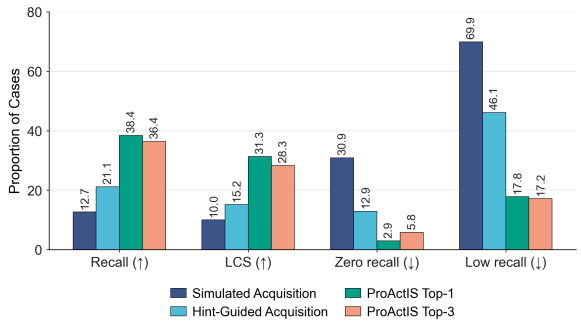

Figure 4: Case-level evidence recovery across dynamic workflow settings. ProActIS improves recall and ordered trajectory similarity while substantially reducing zero-recall and low-recall cases.

ing. ProActIS-assisted settings provide Top-1 or Top-3 state-conditioned proposals, while the LLM retains final control over both examination requests and diagnosis.

Table 2 shows that unguided dynamic acquisition is the main bottleneck. Although DeepSeek V4 Pro achieves 55.7% accuracy with complete information, performance drops to 15.3% under Simulated Acquisition and 17.6% under Hint-Guided Acquisition, with both settings requesting substantially fewer examinations than the reference trajectories. Hint guidance improves Jaccard overlap from 12.67% to 21.15%, but lacks positive guidance on the next examination.

ProActIS markedly improves evidence acquisition. Top-1 support raises micro hit rate to 64.55%, Jaccard overlap to 38.40%, and average rounds to 5.93, closer to the 7.97-round reference. Top-3 achieves the strongest local matching performance, with 71.68% micro and 71.61% macro hit rates. In both settings, diagnosis accuracy approaches the all-information control, indicating that better examination proposals recover much of the evidence needed for diagnosis.

Figure 4 further shows that ProActIS reduces case-level failures. Unguided settings leave many cases with zero or low recall, whereas ProActIS Top-1 achieves the best case-level recall, ordered trajectory similarity, and matched examinations, and Top-3 maintains low failure rates with stronger local hit rates. Overall, ProActIS makes dynamic workflows more accurate at each step and more complete across the full trajectory.

5.3 Turn-Level Next-Exam Alignment

We next evaluate whether each method can select the immediate next reference examination at the turn level. To make the comparison fair, all LLM baselines (Qwen Team, 2026; Sellergren et al., 2025; Baichuan-M2 Team et al., 2025; DeepSeek-AI, 2026) are given the standardized action bank and asked to choose within the same action space. This setting therefore tests next-action selection under a shared interface, rather than comparing constrained ranking with unconstrained free-form generation.

The top-frequency baseline performs poorly, showing that next-examination selection is not a global popularity problem. The LLM baselines perform better than frequency ranking, but their exact next-action recall remains low even with the standardized action bank provided. ProActIS substantially outperforms all baselines, reaching 0.6028 Recall@1, 0.8868 Recall@5, and 0.7258 MRR. Its Exact-or-future@1 of 0.8245 shows that its gains are not limited to strict immediate-action matching, but also improve trajectory-relative validity.

5.4 Component Ablation Study

We conduct a cumulative component ablation to understand which design choices are responsible for the gains observed in turn-level next-exam ranking. Table 4 progressively adds the major modeling and training components introduced in Section 4.

The largest early gain comes from replacing the single-vector state encoder with the four-channel state representation. This suggests that next-examination selection requires more than a compact summary of the patient state. Hard-negative learning further improves ranking by forcing the model to distinguish the next reference action from plausible but incorrect legal alternatives. STOP-aware termination modeling mainly changes termination behavior, while structured channel calibration provides the largest late-stage ranking improvement. The final future-aware hard-negative refinement gives the best overall performance by avoiding excessive penalties on later trajectory-consistent examinations while still penalizing unsupported candidates.

5.5 Channel Ablation Study

Table 5 reports leave-one-channel-out ablations for the full model. Removing any channel reduces performance, indicating that the four-channel state representation is not merely redundant capacity.

Removing the prior channel causes the smallest but still consistent drop, while removing the trajectory and residual channels causes larger declines.

Table 2: Dynamic LLM workflow performance. Percent-valued columns are shown without percent signs.

Setting	Avg. GT	Avg. req.	Micro hit	Macro hit	Jaccard	Diag. Acc.	Judge
All information	–	–	–	–	–	55.7	0.599
Initial + Simulated Acquisition	7.97	3.17	39.90	37.94	12.67	15.3	0.334
Initial + Hint-Guided Acquisition	7.97	2.11	42.16	46.45	21.15	17.6	0.342
Initial + ProActIS Top-1	7.97	5.93	64.55	64.65	38.40	55.2	0.595
Initial + ProActIS Top-3	7.97	5.05	71.68	71.61	36.37	56.5	0.597

Table 3: Offline full-bank next-exam ranking. E/F@1 denotes exact-or-future validity at rank 1.

Model	E/F@1	R@1	R@3	R@5	MRR
Frequent	0.1864	0.0954	0.2817	0.3776	0.2038
MedGemma-27B	0.4244	0.1329	0.3036	0.3866	0.2263
Baichuan-M2 32B	0.4123	0.1051	0.2753	0.3979	0.2046
Qwen3.5 27B	0.4642	0.1411	0.3459	0.4885	0.2598
DeepSeek V4 Pro	0.4411	0.1378	0.3450	0.4891	0.2582
ProActIS	**0.8245**	**0.6028**	**0.8169**	**0.8868**	**0.7258**

Table 4: Cumulative component ablation.

Ablation	R@1	R@5	MRR	STOP@1
Single-vector	0.3092	0.6183	0.4489	0.8948
Four-channel	0.5019	0.8073	0.6355	0.6327
+ Hard neg.	0.5427	0.8654	0.6817	0.7440
+ STOP-aware	0.5441	0.8647	0.6830	0.9093
+ Channel calib.	0.5907	0.8781	0.7152	0.9543
+ Future-aware	**0.6028**	**0.8868**	**0.7258**	**0.9621**

Table 5: Leave-one-channel-out ablation.

Removed	R@1	R@3	R@5	MRR
None	**0.6028**	**0.8169**	**0.8868**	**0.7229**
Prior	0.5647	0.7848	0.8612	0.6912
Trajectory	0.5316	0.7556	0.8452	0.6652
Residual	0.5294	0.7521	0.8423	0.6619
VOI-inspired	0.4957	0.7661	0.8585	0.6490

The VOI-inspired channel has the largest average drop, especially in R@1 and MRR. This supports the view that next-exam proposal is fundamentally an information-seeking task: the model must estimate which examination is likely to reduce the current evidence gap, not merely whether an examination is clinically related to the case.

5.6 Case Study

One common reason for poor LLM performance in dynamic diagnosis is premature closure, where once early evidence supports the chief complaint, the model tends to stop collecting further information and make a final diagnosis. A representative case involved a 78-year-old woman with PAD/PVD, CKD, atrial fibrillation on Coumadin, anemia, cirrhosis, and heart failure, presenting with right lower-extremity ischemia, redness, severe pain, and ulceration.

Without ProActIS, the LLM rapidly anchored on the chief complaint. In both the simulated acquisition and guided-hint settings, it requested lower-extremity Doppler ultrasonography followed by CT imaging, and reached a final diagnosis after only two rounds. By contrast, ProActIS-assisted workflows acquired a broader spectrum of clinical evidence within a comparable number of turns. With either Top-1 or Top-3 support, the LLM requested six examinations and matched five reference examinations. Importantly, these matched examinations covered not only vascular evaluation, but also hematologic, coagulation, metabolic, hepatic function, and imaging evidence, suggesting that ProActIS promotes more comprehensive and clinically grounded evidence acquisition prior to diagnosis.

This case shows that unguided LLMs can stop after obtaining only chief-complaint-related evidence, causing failure in dynamic diagnosis. ProActIS improves the trajectory by guiding the LLM toward a broader set of reference-aligned examinations before final diagnosis.

6 Conclusion

This work shows that effective clinical language agents require not only strong final-answer reasoning, but also reliable evidence acquisition during the diagnostic process. By framing examination proposal as a state-conditioned ranking problem and evaluating it with both offline alignment and assisted dynamic workflows, ProActIS demonstrates that lightweight, explicit support for selecting the next examination can substantially improve evidence coverage and workflow alignment. Our findings suggest that proactive examination proposal is a practical and underexplored component of LLM-based clinical agents, and point toward future systems that integrate reasoning, evidence acquisition, and termination decisions in more faithful dynamic diagnostic workflows.

Limitations

This work has several limitations. First, ProActIS does not model physical examination actions. This is mainly a dataset limitation: in MIMIC-IV, physical examination information appears in clinical notes without a standardized action nomenclature, and the descriptions are often subjective, heterogeneous, and difficult to normalize into order-level actions comparable to laboratory, microbiology, or radiology examinations.

Second, the reference trajectories used in this work should not be interpreted as gold-standard clinical pathways. They reflect recorded examination trends under each patient's observed state and institutional practice patterns, but public datasets do not provide physician-validated optimal dynamic diagnostic trajectories for this setting.

Third, the proposed benchmark and model outputs have not yet been validated by physicians, so trajectory alignment should be treated as a proxy evaluation rather than clinical endorsement. Nevertheless, the benchmark is useful for measuring whether agents can recover observed examination-acquisition patterns under controlled leakage constraints, which is a necessary step before physician-validated prospective evaluation.

Fourth, ProActIS is an examination proposer only: it recommends candidate exams or STOP, but it does not interpret returned results, synthesize diagnoses, or make final clinical decisions. Finally, end-to-end diagnostic outcomes depend on the downstream LLM, the simulator or feedback mechanism, and the prompting protocol, so improvements in examination proposal may not transfer uniformly across all clinical-agent systems.

References

Baichuan-M2 Team, Chengfeng Dou, Chong Liu, and 1 others. 2025. Baichuan-m2: Scaling medical capability with large verifier system. *Preprint*, arXiv:2509.02208.

David Bani-Harouni, Chantal Pellegrini, Ege Özsoy, Nassir Navab, and Matthias Keicher. 2025. Language agents for hypothesis-driven clinical decision making with reinforcement learning. *arXiv preprint arXiv:2506.13474*.

DeepSeek-AI. 2026. Deepseek-v4: Towards highly efficient million-token context intelligence.

Arthur S. Elstein and Alan Schwarz. 2002. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *BMJ*, 324(7339):729–732.

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

Scott L. Fleming, Alejandro Lozano, William J. Haberkorn, Jenelle A. Jindal, Eduardo P. Reis, Rahul Thapa, Louis Blankemeier, Julian Z. Genkins, Ethan Steinberg, Ashwin Nayak, Birju S. Patel, Chia-Chun Chiang, Alison Callahan, Zepeng Huo, Sergios Gatidis, Scott J. Adams, Oluseyi Fayanju, Shreya J. Shah, Thomas Savage, and 11 others. 2023. Medalign: A clinician-generated dataset for instruction following with electronic medical records. *Preprint*, arXiv:2308.14089.

Google DeepMind. 2026. Gemini 3.1: Expanding the frontier of advanced reasoning and multimodal intelligence. `https://deepmind.google/models/gemini/pro/`. Accessed: 2026-05-26.

Arya Hadizadeh Moghaddam, Mohsen Nayebi Kerdabadi, Mei Liu, and Zijun Yao. 2024. Contrastive learning on medical intents for sequential prescription recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, CIKM '24, page 748–757. ACM.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Hyewon Jeong, Nassim Oufattole, Matthew Mcdermott, Aparna Balagopalan, Bryan Jangeesingh, Marzyeh Ghassemi, and Collin Stultz. 2024. Event-based contrastive learning for medical time series. *Preprint*, arXiv:2312.10308.

Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. 2021. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 11(14):6421.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. 2019. Pubmedqa: A dataset for biomedical research question answering. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 2567–2577.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. 2020. Mimic-iv. *PhysioNet. Available online at: https://physionet. org/content/mimiciv/1.0/(accessed August 23, 2021)*.

Jerome P. Kassirer. 1983. Teaching clinical medicine by iterative hypothesis testing: Let's preach what we practice. *New England Journal of Medicine*, 309(15):921–923.

Xingwang Li, Yijia Zhang, Xiaobo Li, Hao Wei, and Mingyu Lu. 2023. Dgcl: distance-wise and graph contrastive learning for medication recommendation. *Journal of Biomedical Informatics*, 139:104301.

National Academies of Sciences, Engineering, and Medicine. 2015. *Improving Diagnosis in Health Care*. The National Academies Press, Washington, DC.

OpenAI. 2026. Openai gpt-5 system card. *Preprint*, arXiv:2601.03267.

Pengcheng Qiu, Chaoyi Wu, Junwei Liu, Qiaoyu Zheng, Yusheng Liao, Haowen Wang, Yun Yue, Qianrui Fan, Shuai Zhen, Jian Wang, Jinjie Gu, Yanfeng Wang, Ya Zhang, and Weidi Xie. 2025. Evolving diagnostic agents in a virtual clinical environment. *Preprint*, arXiv:2510.24654.

Qwen Team. 2026. Qwen3.5: Towards native multimodal agents.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Amanda Sastry, Girish Askell, Pamela Mishkin, Jack Clark, and 1 others. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR.

Regenstrief Institute, Inc. 2025. LOINC (Logical Observation Identifiers Names and Codes). Accessed: 2026-05-26.

Samuel Schmidgall, Rojin Ziaei, Carl William Harris, Ji Woong Kim, Eduardo Pontes Reis, Jeffrey K Jopling, and Michael Moor. 2025. Agentclinic: a multimodal agent benchmark to evaluate AI in simulated clinical environments.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, and 1 others. 2025. Medgemma technical report. *arXiv preprint arXiv:2507.05201*.

Tao Tu, Mike Schaekermann, Anil Palepu, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Yong Cheng, and 1 others. 2025. Towards conversational diagnostic artificial intelligence. *Nature*, 642(8067):442–450.

Michihiro Yasunaga, Jure Leskovec, and Percy Liang. 2022. Linkbert: Pretraining language models with document links. *Preprint*, arXiv:2203.15827.

Qing Yin, Linda Zhong, Yunya Song, Liang Bai, Zhihua Wang, Chen Li, Yida Xu, and Xian Yang. 2025. A decision support system in precision medicine: contrastive multimodal learning for patient stratification. *Annals of Operations Research*, 348(1):579–607.

A Declaration of Generative AI Usage

During the preparation of this work, Large Language Models (LLMs) were utilized solely as language-polishing tools to enhance clarity, readability, grammatical accuracy, and stylistic refinement. The AI technology was strictly confined to proofreading and did not contribute to the generation or conceptualization of new scientific content, data, or ideas. For enhanced visual communication, AI tools were employed strictly to generate individual icons within the flowcharts. The overall architecture and layout of the flowcharts were manually designed and constructed by the authors. All scientific data visualizations were generated exclusively from raw data, with no AI involvement in data representation. All AI-generated outputs were thoroughly reviewed, verified, and curated by the authors before integration into the final manuscript.

B Dataset Construction Details

Leakage Control We retained only admission-facing sections from each discharge summary as model-visible patient context. Discharge diagnosis, hospital course, discharge medications, and sections containing reported laboratory or imaging results were removed because they may reveal downstream diagnostic conclusions or examination outcomes. The Diagnosis field was retained only for cohort filtering and reference analysis. Cases were further screened for explicit diagnosis mentions in Past Medical History and History of Present Illness using string matching, clinical synonym matching, and DeepSeek-assisted binary leakage review before training.

Action Bank Construction Raw laboratory tests, microbiology tests, and radiology records were first collected within a 24-hour window centered on the case index time and ordered chronologically. Because MIMIC-IV records are often item-level rather than order-level, raw items and studies were normalized into order-level examination actions. We used LOINC codes, examination metadata, and clinical semantics to propose mappings from raw records to standardized actions. Candidate mappings were reviewed with GPT-5(OpenAI, 2026) and Gemini 3.1 Pro(Google DeepMind, 2026), manually inspected, and then frozen. The frozen mapping was applied deterministically across all splits.

The final action bank contains 129 standardized

examination actions, grouped into 31 laboratory, 64 microbiology, and 34 radiology actions. A terminal STOP action was added to represent the end of evidence acquisition. DeepSeek was used to assist action-bank and action-card generation. Each action card contains the examination name, bucket, parent ontology, modality, body system, clinical purpose, information type, and typical indications. These cards provide examination semantics to the action encoder rather than serving as one-hot labels. They are not clinically validated guidelines and contain no patient-specific information or case-specific labels.

Turn-Level Example Construction Each raw examination record was mapped to one standardized action. Admissions without mapped examinations were removed. Repeated standardized actions within the same case were collapsed to their first occurrence to avoid over-weighting repeated item-level measurements. Each case trajectory was then converted into turn-level examples: at each turn, the model state includes the initial patient context and the ordered history of previously selected examinations with returned results, while the target is the next reference action. STOP is appended as the final target for each case.

C Method Detail

Model and Training Details. We initialize the text encoder with BioLinkBERT(Yasunaga et al., 2022) to encode patient states, examination-result histories, and action-card descriptions. To reduce trainable parameters and stabilize optimization, we apply LoRA-based(Hu et al., 2022) parameter-efficient fine-tuning to the backbone encoder while training the task-specific projection layers, channel calibration parameters, action bias, STOP head, and ranking objectives. The model is trained with full-bank candidate scoring over the standardized action set, including laboratory, microbiology, radiology, and STOP actions. All main training runs are conducted on 8 NVIDIA H800 GPUs with distributed data-parallel training.

Compute Cost and Hyperparameters. Training for 10 epochs required approximately 60 GPU-hours on 8 GPUs; we used batch size 14 per GPU, bfloat16 with FlashAttention-2, embedding dimension 256, dropout 0.1, AdamW with base learning rate 2×10^{-5}, LoRA rank 8 with $\alpha = 16$ and dropout 0.05, temperature 1.0, max 8 static chunks of length 512 with overlap 64, max 32 events, max event/action length 512/256, hard-negative $K = 5$, rank margin 0.3, rank loss weight 0.05, STOP sufficiency loss weight 0.05, future loss weight 0.05 with denominator discount 0.25, class-balance strength 0.5 with weights clipped to $[0.5, 3.0]$, structured channel calibration and global action bias enabled, and seed 13.

Structured Ranker Details. ProActIS follows a contrastive dual-encoder design. The patient encoder maps the turn-level state $S_t = (S_0, p_t)$ into structured state representations, and the action encoder maps each candidate examination $e \in \mathcal{C}_t$ into action representations. The model then computes a state-action compatibility score $s(t, e)$ for each legal candidate.

The patient state is represented with four learned diagnostic channels:

$$u_t = \{u_t^{\text{prior}}, u_t^{\text{traj}}, u_t^{\text{res}}, u_t^{\text{voi}}\}.$$

These channels are architectural inductive biases rather than guaranteed disentangled clinical variables. The prior channel summarizes baseline examination plausibility from the static patient context. The trajectory channel summarizes the ordered examination-result prefix. The residual channel models the clinical state after observed examination results update the initial presentation. The VOI- inspired channel represents remaining information need; it is a learned representation, not a formal expected-utility calculation.

The static branch encodes the initial patient information S_0 into a global static representation h_{static} and static problem slots Z_{static}. The prior channel is obtained by projecting the global representation:

$$u_t^{\text{prior}} = W_{\text{prior}} h_{\text{static}}.$$

The dynamic branch encodes the ordered examination-result prefix p_t. Each examination-result pair (e_i, r_i) is encoded as an event representation, positional information is added to preserve order, and a sequence encoder produces an event memory M_t and pooled trajectory representation h_{traj}. The trajectory channel is

$$u_t^{\text{traj}} = W_{\text{traj}} h_{\text{traj}}.$$

For the first decision step, where $t = 0$, no examination-result event has been observed. We use a learned empty-prefix representation so that

trajectory-dependent components remain well-defined.

The residual channel is produced by static-dynamic cross-attention. ProActIS uses static problem slots as queries and trajectory event memory as keys and values:

$$A_t = \mathrm{MHA}(Q = Z_{\mathrm{static}}, K = M_t, V = M_t).$$

The attended evidence is combined with the original static slots through a gated update:

$$G_t = \sigma\left(W_g[Z_{\mathrm{static}}; A_t; Z_{\mathrm{static}} \odot A_t; Z_{\mathrm{static}} - A_t]\right),$$

$$\Delta_t = W_\Delta[Z_{\mathrm{static}}; A_t; Z_{\mathrm{static}} \odot A_t],$$

$$R_t = \mathrm{LayerNorm}\left(Z_{\mathrm{static}} + G_t \odot \Delta_t\right).$$

Pooling the updated slots gives the residual representation:

$$h_{\mathrm{res}} = \mathrm{Pool}(R_t), \qquad u_t^{\mathrm{res}} = W_{\mathrm{res}} h_{\mathrm{res}}.$$

The VOI-inspired channel uses an orthogonal residual between the static representation and the evidence-conditioned residual representation. The projection of h_{static} onto h_{res} is

$$\mathrm{Proj}_{h_{\mathrm{res}}}(h_{\mathrm{static}}) = \frac{h_{\mathrm{static}}^\top h_{\mathrm{res}}}{\|h_{\mathrm{res}}\|_2^2 + \epsilon} h_{\mathrm{res}}.$$

The orthogonal residual is

$$d_t = h_{\mathrm{static}} - \mathrm{Proj}_{h_{\mathrm{res}}}(h_{\mathrm{static}}).$$

The VOI-inspired channel is then computed as

$$u_t^{\mathrm{voi}} = \mathrm{LayerNorm}\left(W_{\mathrm{voi}} d_t + e_{\ell(t)}\right),$$

where $e_{\ell(t)}$ is a learned prefix-length embedding.

Each action card contains structured natural-language fields such as the standardized examination name, modality or bucket, parent ontology, body system, clinical purpose, information type, and typical indications. These cards are shared across training, validation, and test splits. They contain no patient-specific information and do not serve as clinical guidelines.

For each candidate examination e, the action encoder produces a shared representation $h_{\mathrm{act}}(e)$. Channel-specific projections then map this shared representation into four action vectors:

$$q^{\mathrm{prior}}(e), \quad q^{\mathrm{traj}}(e), \quad q^{\mathrm{res}}(e), \quad q^{\mathrm{voi}}(e).$$

Score Function and STOP Details. For each channel $k \in \{\mathrm{prior}, \mathrm{traj}, \mathrm{res}, \mathrm{voi}\}$, ProActIS computes

$$z_k(t, e) = (u_t^k)^\top q^k(e),$$

where each logit measures compatibility between candidate examination e and one diagnostic factor of the current state. The final score is a calibrated mixture of the four channel logits:

$$s(t, e) = \frac{1}{\tau} \sum_k \alpha_k(B(t), \kappa(e)) z_k(t, e) + \beta(e).$$

Here, $\tau > 0$ is a learned temperature, $\beta(e)$ is a learned action bias, $\kappa(e)$ denotes the action kind, and $B(t)$ is a prefix-derived workflow bucket. The channel weights are normalized as

$$\alpha_k(B(t), \kappa(e)) = \frac{\exp \Gamma_{B(t),\kappa(e),k}}{\sum_{k'} \exp \Gamma_{B(t),\kappa(e),k'}}.$$

At inference time, ProActIS ranks all legal candidates by the final state-action score:

$$\pi_t = \mathrm{argsort}_{e \in \mathcal{C}_t}\, s(t, e),$$

and predicts the top-ranked action:

$$\hat{e}_t = \pi_t(1).$$

All candidate examinations and STOP are compared in the same score space. Future trajectory labels, final diagnoses, and separate STOP thresholds are not used.

STOP differs from laboratory, microbiology, and radiology actions because it terminates evidence acquisition rather than requesting additional information. ProActIS therefore models STOP both as a ranked candidate in the unified action space and as an auxiliary supervision signal for information sufficiency.

The auxiliary STOP head receives the concatenation of the four state channels and prefix-length embedding:

$$g_t = [u_t^{\mathrm{prior}}; u_t^{\mathrm{traj}}; u_t^{\mathrm{res}}; u_t^{\mathrm{voi}}; e_{\ell(t)}].$$

Future-Aware Training Details. For each turn t, the exact positive is the immediate next reference action e_{t+1}^*. We define the future-valid set as legal actions that appear later in the same trajectory:

$$\mathcal{F}_t = \{e_j : j > t + 1\} \cap \mathcal{C}_t.$$

Future-valid actions are not treated as additional positives; the model is still trained to rank e_{t+1}^*

highest. However, they are protected from strong negative pressure because they remain consistent with the reference trajectory.

For hard-negative learning, ProActIS mines negatives only from invalid legal candidates:

$$\mathcal{I}_t = \mathcal{C}_t \setminus \left(\{e^*_{t+1}\} \cup \mathcal{F}_t\right).$$

Given the invalid candidate set $\mathcal{I}_t$, the future-filtered hard-negative set is

$$\mathcal{H}_t^K = \mathrm{TopK}_{e \in \mathcal{I}_t}\, s(t, e).$$

The exact full-bank cross-entropy loss is

$$\mathcal{L}_{\mathrm{CE}} = -\log \frac{\exp s(t, e^*_{t+1})}{\sum_{e \in \mathcal{C}_t} \exp s(t, e)}.$$

Let

$$\mathcal{F}_t^{\mathrm{exam}} = \mathcal{F}_t \setminus \{\mathrm{STOP}\}.$$

For discount coefficient $\mu \in [0, 1]$, the future-relaxed denominator is

$$Z_t^{\mathrm{future}} = \exp s(t, e^*_{t+1}) + \sum_{e \in \mathcal{C}_t \setminus (\{e^*_{t+1}\} \cup \mathcal{F}_t^{\mathrm{exam}})} \exp s(t, e) + \mu \sum_{e \in \mathcal{F}_t^{\mathrm{exam}}} \exp s(t, e).$$

The future-relaxed cross entropy is

$$\mathcal{L}_{\mathrm{FCE}} = -s(t, e^*_{t+1}) + \log Z_t^{\mathrm{future}}.$$

When $\mu = 1$, this recovers ordinary cross entropy; when $\mu = 0$, non-STOP future-valid examinations are removed from the relaxed denominator. STOP is not discounted, preventing later terminal actions from becoming easier to rank before the reference stopping point.

Given the invalid hard-negative set $\mathcal{H}_t^K$, ProActIS uses a smooth margin objective:

$$\mathcal{L}_{\mathrm{HN}} = \log \left(1 + \sum_{e^- \in \mathcal{H}_t^K} \exp\left(m - s(t, e^*_{t+1}) + s(t, e^-)\right)\right),$$

where m is the ranking margin.

The full objective combines the ranker loss with the future-filtered hard-negative loss and auxiliary STOP sufficiency loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{ranker}} + \lambda_{\mathrm{HN}} \mathcal{L}_{\mathrm{HN}} + \lambda_{\mathrm{STOP}} \mathcal{L}_{\mathrm{STOP}}.$$