Diagnosing and Repairing Local Reasoning Preference Collapse in Medical Chain-of-Thought Distillation

RuiQing Wang[1*], Xiaoyu Tan[2*], Xiyao Wang[1], Chen Jue[1], Xihe Qiu[1†]
[1]Shanghai University of Engineering Science, Shanghai, China
[2]Tencent YouTu Lab, Shanghai, China
ruiqingwang@sues.edu.cn txywilliam1993@outlook.com
aliwang0126@gmail.com jadeschen@sues.edu.cn qiuxihe1993@gmail.com

Abstract

Chain-of-thought (CoT) distillation is a common strategy for transferring reasoning abilities from large teacher models to smaller student models, but in medical question answering, a student may produce plausible rationales while locally drifting toward teacher-disapproved reasoning paths. Thus, final-answer supervision and uniform rationale imitation fail to identify where reasoning fails. We study this problem as local reasoning preference collapse in structured medical CoT distillation and propose **Answer-Calibrated Collapse-Aware Preference Distillation (ACPD)**. ACPD exposes the student's local continuation preferences via teacher-guided verbalized candidates and measures domain-level collapse with a Local Collapse Margin (LCM). It selectively repairs collapse-prone domains by contrasting the student's most preferred teacher-approved and teacher-disapproved continuations. To ensure local repair improves decisions rather than rationale form, ACPD introduces answer calibration that calibrates the full structured reasoning prefix against the correct final choice. Experiments on medical multiple-choice benchmarks show that ACPD improves answer accuracy and collapse-sensitive repair metrics over strong distillation baselines while maintaining competitive out-of-distribution generalization. Results show that medical CoT distillation must diagnose and repair local preferences misaligned with teacher-approved reasoning, not merely transfer rationales. Code will be released at https://github.com/ssh-adventurer123/ACPD.

1 Introduction

Medical chain-of-thought (CoT) distillation aims to transfer clinically meaningful reasoning traces from strong teacher models to smaller students, enabling compact models to produce both accurate answers and reliable intermediate reasoning

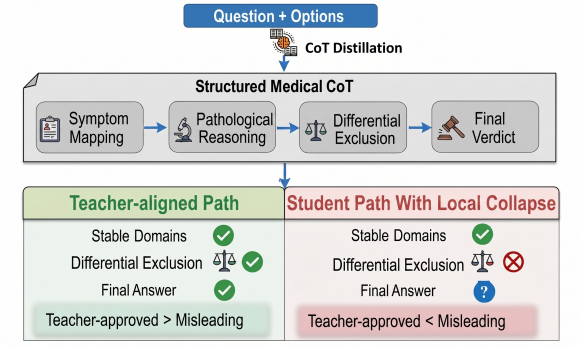


Figure 1: Motivating illustration of local reasoning preference collapse. A student may preserve the structured CoT format while locally preferring a misleading continuation over a teacher-approved one in a target domain such as Differential Exclusion.

(Wei et al., 2022; Kojima et al., 2022; Wang et al., 2023b; Ho et al., 2023; Hsieh et al., 2023). This setting is increasingly important as medical QA evaluation expands from exam-style benchmarks such as MedMCQA, MedQA, and HEAD-QA to physician-oriented evaluations of clinical reasoning (Pal et al., 2022; Jin et al., 2021; Vilares and Gómez-Rodríguez, 2019; Singhal et al., 2023, 2025). However, recent studies show that high answer-level performance can still mask brittle or clinically misaligned decision processes (Turpin et al., 2023; Lanham et al., 2023; He et al., 2023; Liu et al., 2024; Zhang et al., 2025c). As illustrated in Figure 1, this motivates a repair-oriented question: when a student produces structured medical reasoning, can we identify where its local reasoning preference departs from the teacher?

Existing CoT distillation methods improve small-model reasoning by imitating teacher rationales, strengthening rationale-answer coupling, or decomposing long traces into more manageable supervision units (Magister et al., 2023; Li et al., 2023; Wang et al., 2023a; Chen et al., 2024; Wadhwa et al., 2024; Chen et al., 2025b,a; Shen et al., 2025). Yet these methods usually treat the rationale as a

*Equal contribution.
†Corresponding author.

Findings of the Association for Computational Linguistics: EMNLP 2026, pages 20390–20412
October 24-29, 2026 ©2026 Association for Computational Linguistics

uniformly useful training target. In medical reasoning, failures are often localized: a student may correctly identify symptoms but fail to exclude a distractor diagnosis, or follow the right pathological concept but collapse when comparing similar options. Uniform imitation can therefore allocate supervision to already stable reasoning regions while under-addressing the specific domain where the student's local preference diverges from the teacher.

We formalize this failure as teacher-relative local reasoning preference collapse. Instead of relying on model self-verification, which can be unreliable on reasoning tasks (Stechly et al., 2025), we use an external teacher-guided signal to diagnose local misalignment inside structured medical reasoning. For each reasoning domain, we construct teacher-approved and teacher-disapproved candidate continuations and define a Local Collapse Margin (LCM), measuring whether the student assigns more probability mass to teacher-approved continuations than to teacher-disapproved alternatives. This connects to process-supervision and verifier-based reasoning methods (Lightman et al., 2024; Lu et al., 2024; Wang et al., 2024; Setlur et al., 2025; Zhang et al., 2025b; Jiang et al., 2025), but differs in granularity: LCM localizes collapse inside semantically meaningful medical reasoning domains rather than scoring an unstructured sequence of steps.

We propose **Answer-Calibrated Collapse-Aware Preference Distillation (ACPD)**, which converts LCM diagnosis into targeted repair pressure. ACPD retains domain-wise CoT scaffold learning, applies LCM-guided preference repair to collapse-prone domains, and uses answer calibration to align repaired reasoning with the final medical choice. Experiments on MedMCQA show that ACPD improves overall accuracy and yields stronger collapse-sensitive repair behavior compared with strong CoT distillation baselines. Further slice-level and qualitative analyses suggest that ACPD particularly benefits collapse-sensitive examples, including failures around differential exclusion.

Our contributions are:

- We identify teacher-relative local reasoning preference collapse as a failure mode in structured medical CoT distillation.
- We introduce Local Collapse Margin, a domain-level diagnostic comparing student preference over teacher-approved and teacher-disapproved continuations.
- We propose ACPD, an answer-calibrated preference-distillation framework that targets LCM-diagnosed collapse-prone domains.
- We evaluate ACPD with answer accuracy, collapse-sensitive slices, repair rates, and qualitative cases, showing that localized preference repair improves both final prediction and reasoning-path recovery.

2 Related Work

Our work is most closely related to three lines of research: CoT distillation for small language models, process-level and preference-based reasoning supervision, and medical reasoning evaluation. We organize prior work by technical role rather than chronology, focusing on how ACPD differs from methods that transfer rationales, supervise reasoning processes, or evaluate medical QA performance.

2.1 Chain-of-Thought Distillation

CoT distillation transfers teacher-generated reasoning traces to smaller models, building on the observation that explicit rationales can improve multi-step reasoning (Wei et al., 2022; Kojima et al., 2022; Wang et al., 2023b). Early work showed that teacher rationales can provide effective supervision for compact students, especially when reasoning traces are aligned with target answers (Ho et al., 2023; Hsieh et al., 2023; Magister et al., 2023; Shridhar et al., 2023; Li et al., 2025). Subsequent methods improve reasoning transfer through symbolic rationale distillation, self-consistent distillation, stronger rationale-answer coupling, step-wise attention transfer, chunk-wise reasoning compression, and latent CoT representations (Li et al., 2023; Wang et al., 2023a; Chen et al., 2024; Wadhwa et al., 2024; Chen et al., 2025b,a; Shen et al., 2025). These methods refine how rationales are represented or transferred, but they generally do not diagnose which local reasoning domain is fragile for a given student. In contrast, ACPD identifies teacher-relative local preference collapse with LCM and applies targeted repair pressure to LCM-diagnosed collapse-prone domains.

2.2 Process Supervision and Preference-Based Reasoning

Recent work has moved beyond final-answer supervision by using intermediate reasoning signals

to train verifiers, process reward models, or search-guided reasoning systems (Lightman et al., 2024; Wang et al., 2024; Lu et al., 2024; Zhang et al., 2024; Setlur et al., 2025; Zhang et al., 2025b). Related studies further explore process-supervised pretraining, error-aware reward modeling, self-verification limitations, and medical process reward models (Jiang et al., 2025; Pala et al., 2025; Stechly et al., 2025; Yun et al., 2025). In parallel, preference optimization methods show that pairwise preference signals can align language models without explicit scalar reward modeling (Rafailov et al., 2023; Gheshlaghi Azar et al., 2024). These approaches establish the value of intermediate and preference-based supervision, but they typically operate over generic steps, full trajectories, or broad preference pairs. ACPD shares the teacher-constructed pairwise preference supervision of teacher-generated preference distillation, but uses student-specific LCM diagnosis to determine where repair is applied: only collapsed or fragile reasoning domains receive additional domain-level preference supervision.

2.3 Medical Reasoning and Evaluation

Clinical-reasoning research has proposed multiple theories and models, each accounting for only part of this complex process (Yazdani and Hoseini Abardeh, 2019). Medical LLM research has developed strong clinical QA systems and increasingly realistic evaluation protocols (Singhal et al., 2023, 2025; He et al., 2023; Liu et al., 2024; Zhang et al., 2025c). Recent work also improves clinical diagnostic reasoning with Toulmin-structured argumentation and curriculum training (Zhan et al., 2026). Other work improves medical language models through test-time adaptation, clinical alignment, knowledge aggregation, few-shot reasoning, or synthetic medical reasoning data (Shi et al., 2024; Liao et al., 2024; Nachane et al., 2024; Sun et al., 2025). These studies highlight the importance of medical reasoning quality, but most evaluation remains answer-centric or relies on global explanation assessment. Our work studies medical CoT distillation through a repair lens: using structured reasoning domains, we measure whether a student locally prefers teacher-approved continuations over teacher-disapproved alternatives and then train the model to repair those localized failures.

3 Method

3.1 Overview

We propose **Answer-Calibrated Collapse-Aware Preference Distillation (ACPD)**, a framework for diagnosing and repairing local reasoning preference collapse in structured medical CoT distillation. Given a medical multiple-choice question x, the teacher provides a structured rationale with four domains: *Symptom Mapping*, *Pathological Reasoning*, *Differential Exclusion*, and *Final Verdict*; the student is trained to generate both the structured rationale and the final answer. As shown in Figure 2, ACPD first exposes the student's local continuation preferences under teacher-completed domain prefixes, then uses teacher scores to compute a Local Collapse Margin (LCM) and identify domains where the student's preference shifts toward teacher-disapproved continuations.

After diagnosis, ACPD applies collapse-aware domain preference repair to increase the likelihood of teacher-approved continuations over teacher-disapproved ones, while answer calibration keeps the repaired reasoning trajectory aligned with the final medical choice. The following subsections describe the framework in order: Section 3.2 introduces teacher-guided candidate construction, Section 3.3 defines LCM-based collapse diagnosis, Section 3.4 presents domain-level preference repair, and Section 3.5 gives the final answer-calibrated training objective.

3.2 Teacher-Guided Verbalized Candidate Construction

We use teacher-guided verbalized candidate construction to expose the student's local continuation preference inside each structured reasoning domain. This procedure is inspired by verbalized sampling, which prompts a model to explicitly verbalize multiple candidate responses and their associated probabilities (Zhang et al., 2025a). Unlike its original use for improving open-ended generation diversity, we use verbalized candidates as a local preference probe under teacher-completed medical reasoning prefixes.

For a target domain d, let c_d denote the preceding reasoning context, consisting of the question, answer options, and the teacher-completed domains before d. We prompt the student to generate K alternative continuations for domain d:

$$\mathcal{C}_d = \{(s_{d,k}, p_{d,k})\}_{k=1}^{K}, \quad (1)$$

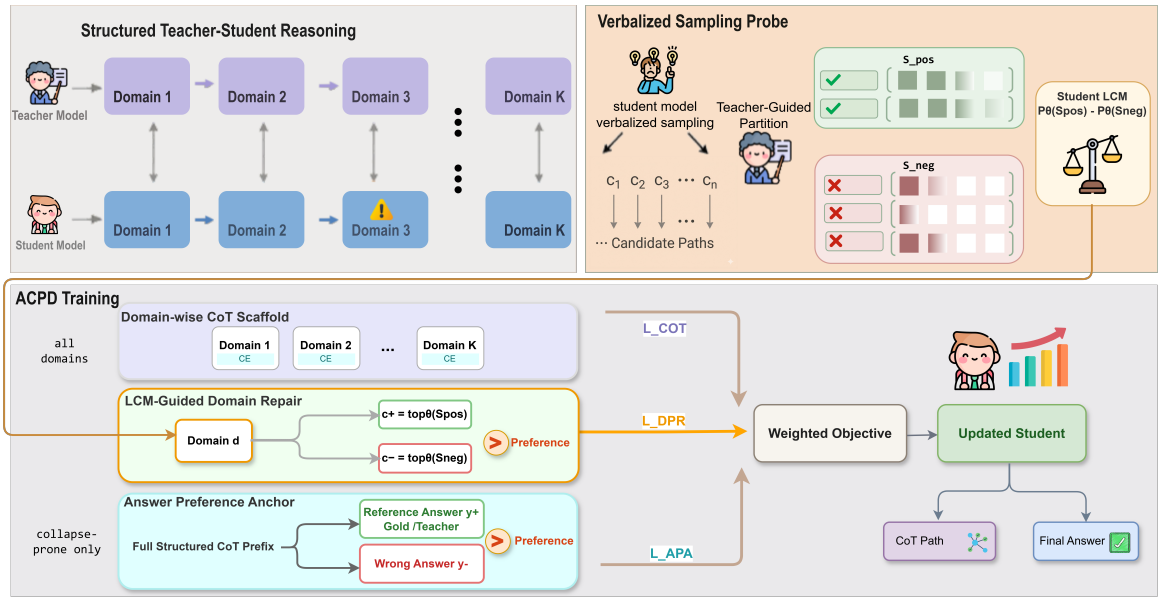


Figure 2: Overview of Answer-Calibrated Collapse-Aware Preference Distillation (ACPD). ACPD first uses teacher-guided verbalized candidate construction to expose the student's local continuation preferences, then diagnoses collapse-prone domains with Local Collapse Margin (LCM), and finally applies collapse-aware domain preference repair with answer calibration.

where $s_{d,k}$ is a verbalized candidate continuation and $p_{d,k}$ is the student's self-assigned probability for that continuation. We normalize the probabilities within each domain so that $\sum_k p_{d,k} = 1$.

This construction is used as a local preference probe rather than as a calibrated uncertainty estimator. The probabilities are interpreted only as relative preference masses over the generated candidate continuations. In our implementation, we profile three target domains: *Pathological Reasoning*, *Differential Exclusion*, and *Final Verdict*. *Symptom Mapping* is used as part of the structured reasoning scaffold and as an upstream prefix, but it is not directly profiled for collapse.

3.3 Local Collapse Margin Diagnosis

We diagnose local collapse by comparing the student's probability mass on teacher-approved and teacher-disapproved continuations. Each candidate $s_{d,k}$ is evaluated by the teacher under the same prefix c_d using mean token log-likelihood. We then partition the candidate set into teacher-approved continuations S^d_{pos} and teacher-disapproved continuations S^d_{neg} according to the teacher score threshold.

Given the normalized candidate probabilities, we define the positive and negative preference masses as:

$$P_d^+ = \sum_{s_{d,k} \in S^d_{\text{pos}}} p_{d,k}, \qquad P_d^- = \sum_{s_{d,k} \in S^d_{\text{neg}}} p_{d,k}. \tag{2}$$

The Local Collapse Margin for domain d is then:

$$\text{LCM}_d = P_d^+ - P_d^-. \tag{3}$$

A domain is considered collapsed when $\text{LCM}_d \leq 0$, meaning that the student assigns at least as much local preference mass to teacher-disapproved continuations as to teacher-approved ones. We also treat domains with $0 < \text{LCM}_d \leq \epsilon$ as fragile, where ϵ is a small margin threshold. The first collapsed domain along the structured reasoning order is used as the estimated collapse location. LCM is teacher-relative: it does not claim absolute medical correctness, but measures whether the student's local continuation preference agrees with the teacher's candidate ranking.

3.4 Collapse-Aware Preference Repair

We convert LCM diagnosis into targeted domain-level preference repair. For each collapse-prone profiled domain, ACPD constructs a representative preference pair from the teacher-approved and teacher-disapproved candidate sets. Specifically, we select the student-preferred representative in

each set:

$$s_d^+ = \arg \max_{s_{d,k} \in S_{\text{pos}}^d} p_{d,k},$$
$$s_d^- = \arg \max_{s_{d,k} \in S_{\text{neg}}^d} p_{d,k}. \quad (4)$$

Thus, s_d^+ is the teacher-approved continuation that the student already finds most plausible, while s_d^- is the teacher-disapproved continuation that most attracts the student. Both continuations are conditioned on the same prefix c_d, so the objective directly contrasts two alternative local reasoning paths for the same medical context.

For a continuation $s = (s_1, \ldots, s_{|s|})$, we score it using length-normalized continuation log-likelihood:

$$\ell_\theta(s \mid c) = \frac{1}{|s|} \sum_{t=1}^{|s|} \log P_\theta(s_t \mid c, s_{<t}). \quad (5)$$

For each preference pair, we define the local preference gap as:

$$\Delta_\theta(c_d, s^+, s^-) = \ell_\theta(s^+ \mid c_d) - \ell_\theta(s^- \mid c_d). \quad (6)$$

The domain-level repair loss is:

$$\mathcal{L}_{\text{dom}} = \frac{\sum_i \omega_i \left[-\log \sigma \left(\Delta_\theta^{(i)} - m_{\text{dom}} \right) \right]}{\sum_i \omega_i}, \quad (7)$$

where ω_i is the LCM-dependent repair weight and m_{dom} is a preference margin. Domains with $\text{LCM}_d \leq 0$ receive stronger repair weight than fragile domains with $0 < \text{LCM}_d \leq \epsilon$, while stable domains are not assigned domain preference pairs.

Unlike uniform rationale imitation, this repair loss does not apply equal pressure to every part of the reasoning trace. Stable domains remain primarily supervised by domain-wise scaffold imitation, while collapsed or fragile domains receive explicit preference pressure. This directly targets the local preference reversal measured by LCM, rather than treating all rationale segments as equally informative.

3.5 Answer-Calibrated Training Objective

ACPD combines domain-wise CoT scaffold learning, domain-level preference repair, and answer calibration. The scaffold loss $\mathcal{L}_{\text{sft}}$ is implemented as domain-wise CWT: each training view supervises one target domain or the final answer while masking non-target rationale segments. The domain repair loss $\mathcal{L}_{\text{dom}}$ repairs local preference reversals in collapse-prone domains. The answer calibration loss further encourages the repaired structured reasoning prefix to remain consistent with the final medical choice.

For a collapse-prone example, we construct answer-level preference pairs under the full structured rationale prefix c_{ans}, which contains the reasoning domains before the answer field. The positive continuation a^+ corresponds to the reference answer, using the dataset gold answer when available and the teacher answer otherwise. A negative continuation a^- corresponds to an incorrect option. The objective supports multiple incorrect options; in our main implementation, we use a single wrong option for efficiency and stability. We use the same length-normalized scoring function as in Eq. 5 and define:

$$\Delta_\theta^{\text{ans}} = \ell_\theta(a^+ \mid c_{\text{ans}}) - \ell_\theta(a^- \mid c_{\text{ans}}). \quad (8)$$

The answer-calibration loss is:

$$\mathcal{L}_{\text{ans}} = \frac{\sum_j \nu_j \left[-\log \sigma \left(\Delta_{\theta,j}^{\text{ans}} - m_{\text{ans}} \right) \right]}{\sum_j \nu_j}, \quad (9)$$

where ν_j is the answer-pair weight and m_{ans} is the answer preference margin.

The final ACPD objective is:

$$\mathcal{L}_{\text{ACPD}} = \mathcal{L}_{\text{sft}} + \lambda_{\text{dom}}(t)\mathcal{L}_{\text{dom}} + \lambda_{\text{ans}}(t)\mathcal{L}_{\text{ans}}. \quad (10)$$

Here, t denotes the training step. The coefficients $\lambda_{\text{dom}}(t)$ and $\lambda_{\text{ans}}(t)$ use fixed soft ramps: early training emphasizes domain-wise CoT scaffold learning, domain preference repair is introduced after the student has learned the structured format, and answer calibration is ramped in later to avoid destabilizing local rationale learning. Additional implementation details are provided in Appendix A.

4 Experiments

4.1 Experimental Setup

Datasets and teacher profiles. We use MedMCQA as the in-domain benchmark, a large-scale medical multiple-choice QA dataset covering diverse medical subjects (Pal et al., 2022). All distilled models are trained on the same MedMCQA distillation split and evaluated on a fixed internal 5k validation split held out from the official MedMCQA training set before teacher profiling.

Each example is converted into a teacher-correct structured profile with four domains: *Symptom Mapping*, *Pathological Reasoning*, *Differential Exclusion*, and *Final Verdict*, followed by the final answer. The teacher sees only the question and options, and profiles are retained only when the teacher's generated answer matches the dataset gold label. Before training, we construct offline profiles for 27,035 filtered training examples. For each four-option question, ACPD profiles the three target domains—*Pathological Reasoning*, *Differential Exclusion*, and *Final Verdict*—by having the student produce $K = 5$ candidate reasoning paths per domain; the teacher model then computes the mean token log-likelihood of each candidate under the same prefix. Out-of-domain benchmarks are described in Section 4.5.

Models and training protocol. Main experiments use Qwen3-8B as the student backbone and report the epoch-1 checkpoint. The backbone-transfer comparison in Figure 3 repeats the same protocol with Llama-3.1-8B. Across methods, we keep the training split, maximum length of 896, learning rate of 2×10^{-5}, and greedy decoding fixed; only per-device batch size and gradient accumulation are adjusted for memory. Full optimization details are provided in Appendix A.

Baselines. We compare ACPD with Raw Qwen3-8B, Standard KD (Hinton et al., 2015), Structured CoT-SFT, MI-CoT Distillation (Chen et al., 2024), Reasoning Scaffolding (Wen et al., 2026), and CWT (Chen et al., 2025a). These baselines cover answer-level distillation, full-rationale imitation, rationale-answer coupling, structured reasoning-flow transfer, and chunk-wise CoT distillation.

Evaluation metrics. We report final-answer accuracy as the primary metric. To evaluate localized repair, we additionally report collapse-sensitive accuracy, repair rate, Differential Exclusion repair, and fixed-path reasoning perplexity under teacher forcing.

4.2 Main Results

We first compare final-answer accuracy across two student backbones, and then analyze whether the gains come from repairing collapse-prone reasoning regions. Figure 3 reports overall accuracy on the MedMCQA 5k validation split. ACPD achieves the highest accuracy on both Qwen3-8B and Llama-3.1-8B, improving over the strongest non-ACPD baseline by 0.50 points and 0.82 points, respectively. This indicates that the benefit of collapse-aware preference repair is not specific to a single student backbone.

We report four collapse-sensitive metrics. A_{high} Acc. measures final-answer accuracy on high-severity collapse examples, defined as examples whose initial LCM is below -0.1, whose top student candidate is teacher-disapproved, and whose normalized prior entropy is below 0.6. Clean Acc. measures accuracy on examples whose profiled-domain LCMs are all above 0.1. Repair Rate measures the proportion of initially collapsed examples whose LCM becomes positive after distillation under the same teacher-defined candidate partition, and thus reflects teacher-relative repair. DE Repair further reports repair performance on Differential Exclusion, which is the most frequent and central option-level collapse domain in our structured reasoning profiles.

Table 1 shows that ACPD achieves the best performance across all four collapse-sensitive metrics, including high-severity collapse accuracy, clean-example accuracy, overall repair rate, and Differential Exclusion repair. Under the teacher-relative LCM evaluation used here, this indicates that ACPD improves local preference-repair metrics without reducing final-answer performance on non-collapse examples.

The comparison with CWT is particularly informative. CWT improves structured rationale learning by decomposing long CoT traces into smaller units, but it does not distinguish stable domains from collapse-prone domains. ACPD improves over CWT on A_{high} Acc. by 2.50 points and Repair Rate by 2.29 points, showing that LCM-guided preference repair provides additional benefit beyond domain-wise rationale learning.

ACPD also improves Clean Acc., suggesting that targeted repair does not simply overfit collapse-prone examples at the expense of stable examples. This supports the design of answer calibration: in the medical multiple-choice QA setting studied here, it constrains local preference repair to remain consistent with the final answer while preserving answer-level performance.

4.3 Collapse-Sensitive Evaluation

Figure 4 first examines where local preference collapse occurs before repair. The diagnosis is computed on the profiled MedMCQA training split using the same LCM construction as ACPD. Differ-

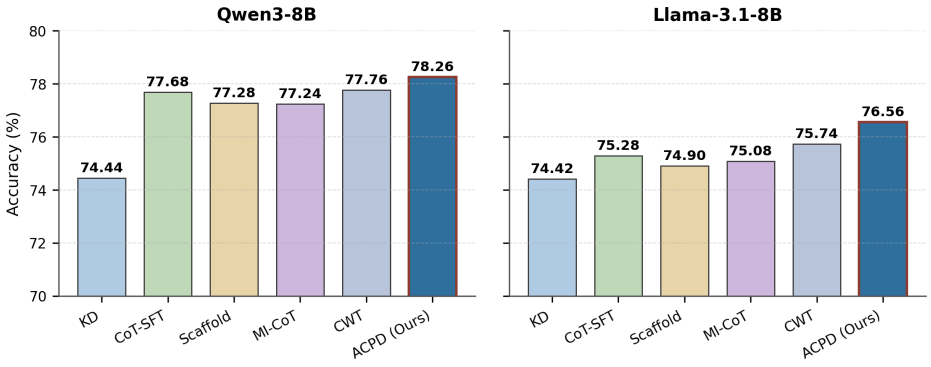


Figure 3: Overall accuracy across two student backbones on the MedMCQA 5k validation split. All methods use the same training split and evaluation protocol; rationale-based methods use the same structured medical CoT supervision setting. ACPD (Ours) achieves the highest accuracy on both Qwen3-8B and Llama-3.1-8B, suggesting that collapse-aware preference repair is not tied to a single backbone.

Method	A_{high} **Acc.** ↑	**Clean Acc.** ↑	**Repair Rate** ↑	**DE Repair** ↑	**Acc. Diff.**	**95% CI**
Standard KD	77.61	74.25	45.19	50.11	+3.82	[2.62, 5.02]
Structured CoT-SFT	79.29	75.71	51.50	58.98	+0.58	[0.12, 1.08]
Reasoning Scaffolding	80.29	77.39	54.15	55.63	+0.98	[0.52, 1.98]
MI-CoT Distillation	83.58	77.47	55.84	63.10	+1.02	[0.18, 2.10]
CWT	81.60	77.43	54.32	62.29	+0.50	[0.34, 1.05]
ACPD	**84.10**	**77.99**	**56.61**	**63.91**	–	–

Table 1: Collapse-sensitive in-domain results on the MedMCQA 5k validation split. Metrics are percentages. Acc. Diff. is overall accuracy difference between ACPD and each baseline; 95% CI is from paired bootstrap.

ential Exclusion has the highest collapsed-domain rate and the highest rate of teacher-disapproved top candidates, indicating that option-level exclusion is the most fragile structured reasoning domain. This motivates our collapse-sensitive evaluation and the dedicated DE Repair metric. This distributional result should be interpreted under the fixed profiling protocol. The fixed-candidate alignment and candidate-generation sensitivity analyses in Appendix C.4 support this interpretation: regenerated candidate sets can change exact instance-level labels, whereas the aggregate weakest-domain pattern remains Differential Exclusion.

A cross-teacher audit further supports this aggregate pattern: when the Qwen3-32B teacher model is replaced with Gemma4-31B, Differential Exclusion remains the domain with the lowest mean LCM (Appendix C.5).

Fixed-path reasoning perplexity. We further evaluate whether LCM-diagnosed collapse regions correspond to more difficult reasoning spans. To ensure a controlled comparison, we do not compute perplexity on each model's generated rationales. Instead, all models are evaluated under teacher forcing on the same structured teacher profiles, using

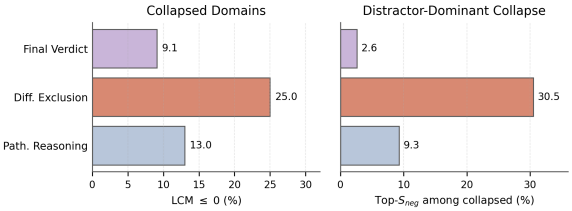


Figure 4: Pre-repair collapse diagnosis on the profiled MedMCQA training split. Differential Exclusion has the highest collapsed-domain rate and the highest rate of teacher-disapproved top candidates, motivating domain-level collapse repair and the DE Repair metric.

the same tokenized reference paths and the same initial collapse-domain annotations. Thus, differences in Table 2 reflect how well each model assigns likelihood to an identical structured reasoning path, rather than differences in generation length, formatting, or content.

Table 2 shows that ACPD obtains the lowest perplexity on both collapse and non-collapse reasoning spans, indicating stronger compatibility with the structured teacher rationale. Across methods, collapse-domain perplexity is consistently higher than non-collapse perplexity, supporting the interpretation that LCM identifies locally challenging

Method	Coll. PPL ↓	Non-coll. PPL ↓
Raw Qwen3-8B	1.8932	1.8171
Standard KD	1.8519	1.7331
Structured CoT-SFT	1.5229	1.4615
Reasoning Scaffolding	1.5188	1.4630
MI-CoT	1.5928	1.5654
CWT	1.5558	1.4995
ACPD	**1.4826**	**1.4051**

Table 2: Fixed-path reasoning perplexity on the same structured teacher profiles. Coll. and Non-coll. denote collapse-domain and non-collapse reasoning spans defined by the same initial LCM diagnosis. Lower PPL indicates better compatibility with the teacher-provided structured rationale.

reasoning regions rather than arbitrary segments.

4.4 Ablation Study

We conduct ablations to isolate the contribution of each ACPD component. Starting from the full model, we remove domain preference repair, answer calibration, or LCM-guided selection while keeping the same CWT scaffold and training protocol. As shown in Table 3, all three removals reduce MedMCQA validation accuracy, confirming that ACPD benefits from both local repair and answer-level calibration. Removing domain repair hurts more than removing answer calibration, indicating that answer-level calibration alone cannot replace explicit preference repair inside collapse-prone domains. The largest drop comes from the broad/uniform ablation that removes LCM-guided selection and applies domain repair to all eligible profiled domains with valid preference pairs; it falls below the CWT baseline, suggesting that ACPD's gain comes from selectively targeting diagnosed fragile domains rather than uniformly adding preference supervision. Under the equal-budget random-domain control, ACPD remains higher than random allocation in overall accuracy and collapse-sensitive metrics, further supporting that the benefit of LCM-guided allocation is not equivalent to simply adding an equal amount of domain-level preference supervision.

4.5 Out-of-Domain Generalization

We evaluate out-of-domain generalization on MedQA, a USMLE-style exam benchmark with 1,273 test examples (Jin et al., 2021); the English clinical subset of HeadQA, covering medicine, pharmacology, and nursing with 1,375 examples (Vilares and Gómez-Rodríguez, 2019); and MedExpQA, a multilingual medical exam benchmark with doctor-written explanations (Alonso et al., 2024). For MedExpQA, we use the pooled English split of 622 examples because the official English test split is small, and no model is trained or tuned on MedExpQA. As shown in Table 4, ACPD ranks first on MedQA and MedExpQA and second on HeadQA; its unweighted macro-average over the three displayed benchmark accuracies is the highest among the compared methods (70.33, versus 69.67 for MI-CoT). ACPD also outperforms raw Qwen3-8B on all three datasets, suggesting that LCM-guided repair does not simply overfit to MedMCQA.

Methods (Qwen3-8B)	Acc. ↑	Δ
CWT	77.76%	-0.50
ACPD	**78.26%**	–
- w/o domain repair	77.72%	-0.54
- w/o answer calibration	77.91%	-0.35
- w/o LCM selection	77.46%	-0.80
Random-domain repair (equal-budget)	77.72%	-0.54

Table 3: Ablation studies on the MedMCQA validation split. Δ is computed relative to ACPD. *w/o domain repair* removes LCM-guided domain preference repair; *w/o answer calibration* removes answer-level preference calibration; *w/o LCM selection* applies domain repair without LCM-guided selective targeting.

Method	MedQA	HeadQA	MedExpQA
Raw Qwen3-8B	64.80	71.20	66.88
Standard KD	59.50	68.10	67.36
Reasoning Scaffolding	66.77	72.15	66.08
MI-CoT	<u>67.56</u>	**74.40**	67.04
CWT	67.24	66.33	<u>68.97</u>
ACPD	**67.79**	<u>73.10</u>	**70.10**

Table 4: Out-of-domain accuracy on MedQA, HeadQA, and MedExpQA. All reported methods use Qwen3-8B as the student backbone. The best result is in bold and the second-best result is underlined.

5 Conclusion

This paper addresses a localized failure mode in medical CoT distillation: a student can preserve the structured rationale format while locally preferring a teacher-disapproved reasoning continuation. We propose ACPD, which diagnoses such teacher-relative preference collapse with Local Collapse Margin and converts the diagnosis into targeted domain preference repair with answer calibration. On MedMCQA, ACPD yields modest final-answer accuracy gains alongside improved teacher-relative collapse-sensitive repair metrics and fixed-path reasoning compatibility. Across the

three out-of-domain benchmarks, ACPD ranks first on two and second on one, yielding the highest unweighted macro-average accuracy among the compared methods. These results suggest that reasoning distillation should not only transfer rationales, but also identify and repair where student preferences deviate inside the reasoning process. A key future direction is extending collapse-aware repair beyond fixed medical QA domains to more open-ended and clinically realistic reasoning settings.

Limitations

ACPD relies on structured teacher rationales and teacher-labeled candidate continuations. This design provides an explicit signal for diagnosing local reasoning collapse, but the signal is bounded by the teacher model and the chosen domain schema. If the teacher produces incomplete rationales or mislabels candidates, LCM may reflect teacher-specific preferences rather than clinically optimal reasoning. Therefore, LCM should be interpreted as a teacher-relative diagnostic rather than a direct measure of clinical correctness.

Our implementation focuses on medical multiple-choice QA with a fixed set of reasoning domains, which enables controlled diagnosis and repair but does not fully capture open-ended clinical decision making, multi-turn patient interaction, or multimodal medical reasoning. In open clinical settings, reasoning and management decisions may be revised as new information arrives, as documented in a longitudinal acute-care dysphagia study (Gunasekaran et al., 2024). Accordingly, ACPD's four domains are a task-specific, auditable scaffold for medical multiple-choice QA rather than a universal clinical-reasoning ontology. Candidate construction and teacher scoring constitute an offline profiling cost before training; at deployment, ACPD uses only the trained student model, without teacher calls or LCM recomputation. Although our experiments show gains on medical QA benchmarks, these metrics do not replace expert clinical validation before deployment in safety-critical settings. Future work may explore more flexible domain induction, cheaper candidate profiling, and human expert evaluation in clinically realistic scenarios.

References

Iñigo Alonso, Maite Oronoz, and Rodrigo Agerri. 2024. MedExpQA: Multilingual benchmarking of large language models for medical question answering. *Artificial Intelligence in Medicine*, 155:102938.

Xiaoshu Chen, Sihang Zhou, Ke Liang, Xiaoyu Sun, and Xinwang Liu. 2025a. Skip-thinking: Chunk-wise chain-of-thought distillation enable smaller language models to reason better and faster. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 12142–12157, Suzhou, China. Association for Computational Linguistics.

Xin Chen, Hanxian Huang, Yanjun Gao, Yi Wang, Jishen Zhao, and Ke Ding. 2024. Learning to maximize mutual information for chain-of-thought distillation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6857–6868, Bangkok, Thailand. Association for Computational Linguistics.

Yao Chen, Jiawei Sheng, Wenyuan Zhang, and Tingwen Liu. 2025b. Improving reasoning capabilities in small models through mixture-of-layers distillation with stepwise attention on key information. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 4952–4971, Suzhou, China. Association for Computational Linguistics.

Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. 2024. A general theoretical paradigm to understand learning from human preferences. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 4447–4455. PMLR.

Sulekha Gunasekaran, Joanne Murray, and Sebastian Doeltgen. 2024. Clinical reasoning during dysphagia assessment and management in acute care: A longitudinal qualitative study. *International Journal of Language & Communication Disorders*, 59(4):1463–1477.

Zexue He, Yu Wang, An Yan, Yao Liu, Eric Chang, Amilcare Gentili, Julian McAuley, and Chun-Nan Hsu. 2023. MedEval: A multi-level, multi-task, and multi-domain medical benchmark for language model evaluation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8725–8744, Singapore. Association for Computational Linguistics.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*.

Namgyu Ho, Laura Schmid, and Se-Young Yun. 2023. Large language models are reasoning teachers. In

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 14852–14882, Toronto, Canada. Association for Computational Linguistics.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8003–8017, Toronto, Canada. Association for Computational Linguistics.

Dongwei Jiang, Guoxuan Wang, Yining Lu, Andrew Wang, Jingyu Zhang, Chuyu Liu, Benjamin Van Durme, and Daniel Khashabi. 2025. RATIONALYST: Pre-training process-supervision for improving reasoning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26547–26566, Vienna, Austria. Association for Computational Linguistics.

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.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. In *Advances in Neural Information Processing Systems*, volume 35, pages 22199–22213.

Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, Danny Hernandez, Dustin Li, Esin Durmus, Evan Hubinger, Jackson Kernion, Kamile Lukosiute, Karina Nguyen, Newton Cheng, Nicholas Joseph, Nicholas Schiefer, Oliver Rausch, Robin Larson, Sam McCandlish, Sandipan Kundu, and 11 others. 2023. Measuring faithfulness in chain-of-thought reasoning. *Preprint*, arXiv:2307.13702.

Liunian Harold Li, Jack Hessel, Youngjae Yu, Xiang Ren, Kai-Wei Chang, and Yejin Choi. 2023. Symbolic chain-of-thought distillation: Small models can also "think" step-by-step. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2665–2679, Toronto, Canada. Association for Computational Linguistics.

Yuetai Li, Xiang Yue, Zhangchen Xu, Fengqing Jiang, Luyao Niu, Bill Yuchen Lin, Bhaskar Ramasubramanian, and Radha Poovendran. 2025. Small models struggle to learn from strong reasoners. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 25366–25394, Vienna, Austria. Association for Computational Linguistics.

Yusheng Liao, Shuyang Jiang, Zhe Chen, Yu Wang, and Yanfeng Wang. 2024. MedCare: Advancing medical LLMs through decoupling clinical alignment and knowledge aggregation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 10562–10581, Miami, Florida, USA. Association for Computational Linguistics.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. 2024. Let's verify step by step. In *International Conference on Learning Representations*.

Fenglin Liu, Zheng Li, Hongjian Zhou, Qingyu Yin, Jingfeng Yang, Xianfeng Tang, Chen Luo, Ming Zeng, Haoming Jiang, Yifan Gao, Priyanka Nigam, Sreyashi Nag, Bing Yin, Yining Hua, Xuan Zhou, Omid Rohanian, Anshul Thakur, Lei Clifton, and David A. Clifton. 2024. Large language models are poor clinical decision-makers: A comprehensive benchmark. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13696–13710, Miami, Florida, USA. Association for Computational Linguistics.

Jianqiao Lu, Zhiyang Dou, Hongru Wang, Zeyu Cao, Jianbo Dai, Yingjia Wan, Yunlong Feng, and Zhijiang Guo. 2024. AutoPSV: Automated process-supervised verifier. In *Advances in Neural Information Processing Systems*, volume 37, pages 79935–79962.

Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. 2023. Teaching small language models to reason. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1773–1781, Toronto, Canada. Association for Computational Linguistics.

Saeel Sandeep Nachane, Ojas Gramopadhye, Prateek Chanda, Ganesh Ramakrishnan, Kshitij Sharad Jadhav, Yatin Nandwani, Dinesh Raghu, and Sachindra Joshi. 2024. Few shot chain-of-thought driven reasoning to prompt LLMs for open-ended medical question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 542–573, Miami, Florida, USA. Association for Computational Linguistics.

Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. 2022. MedMCQA: A large-scale multi-subject multi-choice dataset for medical domain question answering. In *Proceedings of the Conference on Health, Inference, and Learning*, volume 174 of *Proceedings of Machine Learning Research*, pages 248–260. PMLR.

Tej Deep Pala, Panshul Sharma, Amir Zadeh, Chuan Li, and Soujanya Poria. 2025. Error typing for smarter rewards: Improving process reward models with error-aware hierarchical supervision. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 14940–14954, Suzhou, China. Association for Computational Linguistics.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741.

Amrith Setlur, Chirag Nagpal, Adam Fisch, Xinyang Geng, Jacob Eisenstein, Rishabh Agarwal, Alekh Agarwal, Jonathan Berant, and Aviral Kumar. 2025. Rewarding progress: Scaling automated process verifiers for llm reasoning. In *International Conference on Learning Representations*.

Zhenyi Shen, Hanqi Yan, Linhai Zhang, Zhanghao Hu, Yali Du, and Yulan He. 2025. CODI: Compressing chain-of-thought into continuous space via self-distillation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 677–693, Suzhou, China. Association for Computational Linguistics.

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

Kumar Shridhar, Alessandro Stolfo, and Mrinmaya Sachan. 2023. Distilling reasoning capabilities into smaller language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7059–7073, Toronto, Canada. Association for Computational Linguistics.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Abubakr Babiker, Nathanael Schärli, Aakanksha Chowdhery, Philip Mansfield, Dina Demner-Fushman, and 13 others. 2023. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180.

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. Toward expert-level medical question answering with large language models. *Nature Medicine*, 31:943–950.

Kaya Stechly, Karthik Valmeekam, and Subbarao Kambhampati. 2025. On the self-verification limitations of large language models on reasoning and planning tasks. In *International Conference on Learning Representations*.

Yu Sun, Xingyu Qian, Weiwen Xu, Hao Zhang, Chenghao Xiao, Long Li, Deli Zhao, Wenbing Huang, Tingyang Xu, Qifeng Bai, and Yu Rong. 2025. ReasonMed: A 370k multi-agent generated dataset for advancing medical reasoning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 26446–26467, Suzhou, China. Association for Computational Linguistics.

Miles Turpin, Julian Michael, Ethan Perez, and Samuel R. Bowman. 2023. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. In *Advances in Neural Information Processing Systems*, volume 36, pages 74952–74965.

David Vilares and Carlos Gómez-Rodríguez. 2019. HEAD-QA: A healthcare dataset for complex reasoning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 960–966, Florence, Italy. Association for Computational Linguistics.

Somin Wadhwa, Silvio Amir, and Byron C. Wallace. 2024. Investigating mysteries of CoT-augmented distillation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6071–6086, Miami, Florida, USA. Association for Computational Linguistics.

Peifeng Wang, Zhengyang Wang, Zheng Li, Yifan Gao, Bing Yin, and Xiang Ren. 2023a. SCOTT: Self-consistent chain-of-thought distillation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5546–5558, Toronto, Canada. Association for Computational Linguistics.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023b. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*.

Zihan Wang, Yunxuan Li, Yuexin Wu, Liangchen Luo, Le Hou, Hongkun Yu, and Jingbo Shang. 2024. Multi-step problem solving through a verifier: An empirical analysis on model-induced process supervision. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7309–7319, Miami, Florida, USA. Association for Computational Linguistics.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pages 24824–24837.

Xiangyu Wen, Junhua Huang, Zeju Li, Min Li, Jianyuan Zhong, Zhijian Xu, Mingxuan Yuan, Yongxiang Huang, and Qiang Xu. 2026. Reasoning scaffolding: Distilling the flow of thought from LLMs. In *International Conference on Learning Representations*.

Shahram Yazdani and Maryam Hoseini Abardeh. 2019. Five decades of research and theorization on clinical reasoning: A critical review. *Advances in Medical Education and Practice*, 10:703–716.

Jaehoon Yun, Jiwoong Sohn, Jungwoo Park, Hyunjae Kim, Xiangru Tang, Daniel Shao, Yong Hoe Koo, Ko Minhyeok, Qingyu Chen, Mark Gerstein, Michael Moor, and Jaewoo Kang. 2025. Med-PRM: Medical reasoning models with stepwise, guideline-verified process rewards. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 16554–16571, Suzhou, China. Association for Computational Linguistics.

Chen Zhan, Xiaoyu Tan, Gengchen Ma, Yu-Jie Xiong, Xiaoyan Jiang, and Xihe Qiu. 2026. From answers to arguments: Toward trustworthy clinical diagnostic reasoning with toulmin-guided curriculum goal-conditioned learning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13985–14004, San Diego, California, United States. Association for Computational Linguistics.

Dan Zhang, Sining Zhoubian, Ziniu Hu, Yisong Yue, Yuxiao Dong, and Jie Tang. 2024. ReST-MCTS*: Llm self-training via process reward guided tree search. In *Advances in Neural Information Processing Systems*, volume 37, pages 64735–64772.

Jiayi Zhang, Simon Yu, Derek Chong, Anthony Sicilia, Michael R. Tomz, Christopher D. Manning, and Weiyan Shi. 2025a. Verbalized sampling: How to mitigate mode collapse and unlock LLM diversity. *Preprint*, arXiv:2510.01171.

Lunjun Zhang, Arian Hosseini, Hritik Bansal, Mehran Kazemi, Aviral Kumar, and Rishabh Agarwal. 2025b. Generative verifiers: Reward modeling as next-token prediction. In *International Conference on Learning Representations*.

Ming Zhang, Yujiong Shen, Zelin Li, Huayu Sha, Binze Hu, Yuhui Wang, Chenhao Huang, Shichun Liu, Jingqi Tong, Changhao Jiang, Mingxu Chai, Zhiheng Xi, Shihan Dou, Tao Gui, Qi Zhang, and Xuanjing Huang. 2025c. LLMEval-Med: A real-world clinical benchmark for medical llms with physician validation. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 4888–4914, Suzhou, China. Association for Computational Linguistics.

A Implementation Details

Reproducibility and Release Plan. Upon publication, we will release, through a public repository, the materials needed to reproduce this work, including training profiles, the teacher-scoring and partition procedures, definitions and annotation conventions for LCM and collapse, ACPD training code, and the necessary instructions for running and reproducing the experiments.

Table 5 summarizes the main implementation settings. Unless otherwise specified, the appendix reports the Qwen3-8B setting used for the main ACPD and baseline comparisons. Models are trained on the same MedMCQA training split and evaluated on the same 5k validation split with greedy decoding. For the main comparison, we report results from the epoch-1 checkpoint to keep the distillation budget aligned across methods.

Setting	Value
Student backbone	Qwen3-8B
Teacher model	Qwen3-32B
Teacher output format	Four-domain CoT + final answer
Optimizer	AdamW
Learning rate	2×10^{-5}
Max length	896
Checkpoint	epoch-1
Precision	bfloat16
Decoding	Greedy, temp. 0.0, top-$p = 1.0$
Candidate count	$K = 5$
CWT scaffold	Domain-wise masked CE
Domain repair	Representative pairwise loss over $S_{\text{pos}}/S_{\text{neg}}$
Answer calibration	Pairwise loss over correct/wrong answers
Teacher-score partition	25th percentile by mean token log-likelihood
Collapse typing	A_{high}: LCM < -0.1, top candidate in S_{neg}, normalized prior entropy < 0.6; clean: all profiled-domain LCMs > 0.1
Preference weights	$\lambda_{\text{dom}} = 0.1$, $\lambda_{\text{ans}} = 0.05$; margins $m_{\text{dom}} = m_{\text{ans}} = 0$
Loss ramps	Domain: $0.10 \rightarrow 0.60$; answer: $0.50 \rightarrow 0.90$ training progress
Repair weights	Collapsed domains: 1.0; fragile domains: 0.5; answer negatives use one wrong option
Bootstrap CI	Paired bootstrap, 5,000 resamples

Table 5: Implementation details for ACPD and evaluation.

We use Qwen3-32B as the teacher/profiling model for structured rationale generation, candidate scoring, and profile construction. ACPD uses fixed soft ramps for preference losses. Early training emphasizes domain-wise structured CoT imitation. Domain preference repair is introduced after the student has learned the structured reasoning format, and answer calibration is applied later to align repaired reasoning paths with the final medical choice. Different baselines may use different per-device batch sizes due to memory requirements, but the student backbone, training split, maximum sequence length, learning rate, and evaluation protocol are kept fixed.

For the domain-wise CWT scaffold, each training view supervises only one target domain or the final answer while masking non-target rationale spans. This keeps the supervision unit local while preserving the full structured reasoning context. For ACPD, domain preference pairs are added only when the corresponding profiled domain satisfies the LCM repair threshold, and answer calibration is applied only to collapse-prone examples. This design avoids applying preference losses uniformly to all samples and keeps the additional training pressure focused on diagnosed local failures.

A.1 Checkpoint Sensitivity

We evaluate ACPD checkpoints after one, two, and three epochs to justify the epoch-1 checkpoint used in the main comparison. As shown in Table 6, additional epochs reduce validation accuracy. This suggests that one-pass collapse-aware repair is sufficient in our setting, while further training may overfit fixed preference pairs or weaken the initially targeted repair signal.

Checkpoint	Accuracy
Epoch 1	**78.26**
Epoch 2	75.34
Epoch 3	74.70

Table 6: Checkpoint sensitivity of ACPD on the MedMCQA 5k validation split.

B Practical LCM Construction

This appendix provides implementation details for the LCM diagnosis defined in Section 3.3. The main text gives the formal definition; here we specify how candidate continuations are generated, scored, normalized, and filtered in practice.

For each profiled domain, we condition the student on the teacher-completed prefix before that domain and collect verbalized candidate continuations with self-assigned probabilities. We normalize these probabilities within the same domain and use them only as relative local preference masses. The teacher then scores each continuation using

mean token log-likelihood under the same prefix, and the score threshold partitions candidates into S_{pos} and S_{neg}.

A profiled domain can be used for LCM diagnosis whenever it has valid scored candidates. Preference-pair construction additionally requires both S_{pos} and S_{neg} to be non-empty; domains without such pair coverage are skipped for pair construction and counted in the coverage statistics in Appendix C.4. This separation ensures that LCM-based repair is applied only when the local teacher-approved versus teacher-disapproved contrast is well defined.

C Dataset, Prompts, and Candidate Construction

C.1 Dataset Processing

We use MedMCQA as the source medical multiple-choice QA benchmark. Since the official test labels are not publicly available in the same format required by our collapse profiling pipeline, we create a fixed internal split from the official MedMCQA training set. Specifically, we reserve 5k examples as the validation split for final-answer evaluation and collapse-sensitive analysis, and use the remaining training examples for distillation. The validation split is fixed before teacher profiling and is never used for student training. Each example is converted into a structured teacher format with four reasoning domains and one final answer field.

Teacher-correct filtering. For each MedMCQA training example, the teacher model generates a structured four-domain rationale and a final answer using only the question and options. The gold answer is not included in the teacher prompt. We then parse the JSON output, verify that all required reasoning domains are present and non-empty, and compare the teacher's top-level answer with the dataset gold label. Examples are retained only when the structured output is valid and the teacher answer is correct. This filtering prevents the student from imitating incorrect teacher rationales while avoiding gold-answer leakage during teacher generation. All distilled methods use the same filtered teacher-profile split.

C.2 Teacher Prompt

The teacher prompt is designed to produce profiles that are usable for both rationale distillation and local collapse diagnosis. First, the gold answer is withheld from the teacher prompt, so the teacher must infer the answer from the question, options, and medical knowledge rather than restating a provided label. This avoids gold-answer leakage during teacher generation. Second, the prompt enforces a fixed four-domain schema, which makes the resulting rationales comparable across examples and enables domain-level profiling in ACPD. In particular, requiring a dedicated *Differential Exclusion* field encourages the teacher to compare candidate options explicitly, which is necessary for diagnosing option-level preference collapse.

The JSON-only constraint is included for reliable parsing and downstream training. Since ACPD constructs candidate continuations, teacher-approved and teacher-disapproved sets, and answer-level preference pairs from the structured profile, malformed outputs would introduce noise into both distillation and LCM computation. We therefore require all four domains to be present and non-empty, and require the final answer to appear as a top-level field. These constraints make the teacher profile a stable supervision object rather than a free-form explanation.

C.3 Verbalized Candidate Construction

This subsection records the candidate artifacts used by the profiling pipeline. For each profiled reasoning domain, we store the candidate continuation text, the normalized student probability, the teacher score, and the resulting teacher-approved or teacher-disapproved partition label. These fields are reused consistently for LCM computation, domain preference-pair construction, and coverage analysis.

The profiled domains include Pathological Reasoning, Differential Exclusion, and Final Verdict. Symptom Mapping is used as part of the structured reasoning scaffold but is not included in the candidate-level collapse profiling.

C.4 Candidate Coverage and Probing Validity

We use teacher-prefix-conditioned verbalized sampling as a local preference probe rather than as a calibrated uncertainty estimator. Given the teacher-completed prefix before a target reasoning domain, the student verbalizes multiple candidate continuations and assigns a probability to each candidate. We normalize these probabilities and treat them only as relative local preference masses.

Table 7 shows that verbalized probing produces sufficient positive/negative candidate coverage and exposes measurable teacher-relative preference re-

Teacher Prompt

System role. You are a rigorous medical teacher model for structured reasoning distillation.

Task. Read one medical multiple-choice question and produce a structured four-domain reasoning chain based on accurate medical knowledge.

Important constraints.

1. The gold answer is not provided.
2. The best answer must be inferred only from the question, options, and medical knowledge.
3. The output must preserve exactly four domain labels: Symptom Mapping, Pathological Reasoning, Differential Exclusion, and Final Verdict.
4. Differential Exclusion must explicitly compare candidate options and rule out at least two incorrect options.
5. The output must be valid JSON only.
6. The answer field must contain exactly one of A, B, C, or D.

Required schema.

```
{ "id": "string",
"domains": {
"Symptom Mapping": "string",
"Pathological Reasoning": "string",
"Differential Exclusion": "string",
"Final Verdict": "string"
},
"answer": "A/B/C/D"
}
```

versals. Examples without both teacher-approved and teacher-disapproved candidates are excluded from preference-pair construction but remain part of standard scaffold training and final-answer evaluation.

Domain	Pair	Coll.	Top-neg	LCM
Path. Reason.	65.7	13.0	9.3	0.503
Diff. Excl.	69.0	25.0	30.5	0.350
Final Verdict	47.8	9.1	2.6	0.530
All	60.8	16.5	20.5	0.453

Table 7: Candidate coverage and validity check for teacher-prefix-conditioned verbalized preference probing. Values are percentages except LCM. Pair denotes domains with both S_{pos} and S_{neg}; Coll. denotes $\text{LCM} \leq 0$; Top-neg denotes the fraction of collapsed domains whose highest-probability verbalized candidate is in S_{neg}.

Table 7 shows that verbalized probing produces sufficient positive/negative candidate coverage and exposes measurable teacher-relative preference reversals. Differential Exclusion has the highest collapsed-domain rate, the highest Top-S_{neg} rate among collapsed domains, and the lowest mean LCM, supporting our observation that option-level exclusion is the most fragile structured reasoning domain. We do not interpret the verbalized probabilities as calibrated confidence; they are used only as relative local preference masses for LCM-based diagnosis.

To directly validate this probe, we fixed candidate continuations for 400 held-out validation examples (1,200 reasoning nodes). Under the same question, answer options, teacher-completed prefix, and candidate text, we compared the student's normalized verbalized probabilities with next-token probabilities over candidate IDs elicited from the same Qwen3-8B model by a forced-choice prompt. This setup separates probability assignment from variation in candidate generation. For LCM-related metrics, we retained the same teacher-defined candidate partition and recomputed the margin from each probability distribution. All intervals use sample-cluster bootstrap 95% CIs.

Metric	Result	95% CI
Rank ρ	0.514	[0.487, 0.541]
Top-1	0.674	[0.647, 0.702]
Pairwise	0.727	[0.715, 0.739]
Bad-mass ρ	0.914	[0.895, 0.930]
LCM sign	0.883	[0.865, 0.901]

Table 8: Fixed-candidate alignment of verbalized and forced-choice preferences. All values are reported on a 0–1 scale; Rank ρ and Bad-mass ρ are Spearman correlations. Teacher-disapproved mass and LCM are recomputed under the same teacher-defined candidate partition.

Rank ρ compares within-node candidate rankings, whereas Top-1 and Pairwise measure agreement at the best-candidate and candidate-pair levels. Bad-mass ρ compares the total probability assigned to teacher-disapproved candidates, and LCM sign compares the direction of the margin recomputed from the two probability distributions. Candidate-ranking and node-level mass/LCM-sign permutation tests all yield $p < 0.001$. Thus, within fixed candidate sets, verbalized probabilities align with the model's forced-choice relative preferences and with LCM-relevant mass allocation, supporting their use as relative local preference masses.

The fixed-candidate audit isolates the validity of probability assignment, but candidate regeneration can change exact instance-level LCM types. On the same 200 held-out examples, we varied candidate count and decoding temperature. Across all

candidate-count variants, Differential Exclusion accounted for 76.81–87.39% of collapsed samples; across temperature variants, it accounted for 79.00–81.72%. In a separate prompt-paraphrase audit on 1,000 held-out examples, changing only the system prompt yielded 82.42% domain-level LCM-sign agreement, and Differential Exclusion remained the domain with the lowest mean LCM under both prompts. Together, these analyses define LCM as a teacher-relative local preference probe under a fixed profiling protocol: candidate regeneration can affect exact instance-level labels, whereas the aggregate weakest-domain pattern remains Differential Exclusion.

C.5 Teacher-Model Sensitivity

We conduct a matched 200-example audit to assess whether the aggregate diagnosis depends on the main Qwen3-32B teacher model. We replace Qwen3-32B with Gemma4-31B for structured profile construction and candidate scoring, while holding the student model, domain schema, candidate-generation settings, and audited examples fixed. The sample-level collapse rates are 84.5% under Qwen3-32B and 88.5% under Gemma4-31B; these values are the proportions of the matched examples diagnosed as collapsed under each teacher model, not model probabilities. Differential Exclusion remains the domain with the lowest mean LCM under both teacher models.

D Blinded Medical Plausibility Audit

LCM and repair metrics are constructed from a teacher-defined candidate partition and are therefore teacher-relative. To assess whether this partition has external medical plausibility, we conducted a two-rater blinded pairwise audit.

Candidate-pair construction. For each audited item, we held fixed the same medical multiple-choice question, answer options, and teacher-completed prefix preceding the target reasoning domain. The student generated multiple candidate reasoning fragments at that domain and assigned relative preference probabilities. The teacher scored the candidates under the same prefix and partitioned them into a teacher-approved set S_{pos} and a teacher-disapproved set S_{neg}. For each hard item, we selected the highest teacher-scored fragment in S_{pos} and the highest student-probability fragment in S_{neg}. The latter is the locally student-attractive but teacher-disapproved reasoning path that can give rise to preference collapse.

Sampling and blinding. The audit contains 200 hard candidate pairs: 100 from Differential Exclusion, 50 from Pathological Reasoning, and 50 from Final Verdict. Of these pairs, 120 were diagnosed as collapsed and 80 as fragile. Two independent medical-background annotators evaluated all 200 pairs, yielding 400 individual judgments. Each annotator saw only the question, answer options, fixed upstream reasoning context, target domain, and anonymized Candidate A and Candidate B. They did not see model identities, teacher labels, teacher scores, student preference probabilities, LCM values, gold answers, or teacher answers. Candidate order was independently randomized for the two annotators.

Annotation and results. For each pair, annotators selected whether Candidate A or Candidate B was medically better, whether both were acceptable, whether both were problematic, or whether they could not judge. After annotation, we used a private key to decode Candidate A/B into $S_{\text{pos}}/S_{\text{neg}}$. Across 400 individual judgments, annotators explicitly preferred S_{pos} in 66.0% of judgments (conservative item-level 95% CI: 59.2–72.2%), whereas they explicitly preferred S_{neg} in only 7.5% (4.6–12.0%). In Differential Exclusion, the strict support rate for S_{pos} was 65.0% (55.3–73.6%). The remaining judgments did not explicitly favor either candidate, for example because both were acceptable, both were problematic, or the annotator could not judge. We compute confidence intervals with items, rather than the 400 individual judgments, as the statistical unit to avoid treating two judgments on the same item as fully independent.

Taken together, under a fixed medical question and reasoning context, teacher-approved local candidates are usually aligned with independent medical judgment, whereas student-attractive teacher-disapproved candidates are rarely explicitly preferred. The same directional evidence holds in Differential Exclusion, the domain of primary interest to ACPD. LCM and repair metrics remain teacher-relative, but their comparison direction is externally medically plausible rather than a wholly arbitrary teacher-specific preference.

E Additional Robustness Checks

We conduct four robustness checks to validate the stability of LCM-based diagnosis under threshold variation, probability perturbation, domain-level statistical testing, and candidate-length effects.

E.1 Threshold Sensitivity

We vary the teacher-score percentile threshold used for candidate partitioning from 15% to 35%. As shown in Table 9, increasing the threshold predictably changes the size of the detected collapse set, while the agreement of detected collapse domains remains stable under moderate perturbations. Here, HS denotes high-severity collapse samples, Coll. denotes the collapse rate, Type denotes agreement of the diagnosed collapse type with the default 25% threshold, and D_c denotes agreement of the first detected collapse domain with the default 25% threshold. This check evaluates whether LCM diagnosis is sensitive to the teacher-score cutoff used for candidate partitioning.

τ pct.	HS	Coll.	Type	D_c
15%	0.0375	0.1429	0.8306	0.8366
20%	0.0566	0.2126	0.9075	0.9083
25%	0.0778	0.2913	1.0000	1.0000
30%	0.1004	0.3712	0.9017	0.9036
35%	0.1221	0.4512	0.8054	0.8067

Table 9: Threshold sensitivity of candidate partitioning.

E.2 Robustness to Probability Noise

To test whether LCM depends on finely calibrated absolute probabilities, we inject Gaussian noise into saved candidate-level student probabilities and recompute LCM-based rankings. Table 10 shows that Kendall's τ remains high under moderate perturbations, indicating that LCM mainly relies on relative preference structure.

Noise σ	Mean τ	Median τ	Min τ	Max τ
0.05	0.8739	0.8740	0.8731	0.8749
0.10	0.7960	0.7959	0.7949	0.7975

Table 10: LCM ranking robustness under probability noise.

E.3 Domain-Level Statistical Reliability

We conduct non-parametric statistical tests to verify that domain-level LCM differences are not random fluctuations. A Friedman test over domain-wise LCM values is significant ($\chi^2 = 5579.51$, $p < 0.001$), and post-hoc comparisons show significant differences among all domain pairs. As shown in Table 11, Differential Exclusion has the lowest mean LCM and the lowest average rank, indicating that it is the most fragile reasoning domain under our LCM diagnosis.

Domain	Median	Mean	Avg. rank
Path. Reasoning	0.9000	0.6674	2.0024
Diff. Exclusion	0.6000	0.4579	1.7027
Final Verdict	1.0000	0.7746	2.2949

Table 11: Domain-level LCM reliability statistics. Lower mean LCM and lower average rank indicate greater local fragility.

E.4 Length Bias Check

Since teacher scores are computed as mean token log-likelihood, the main scoring definition already normalizes for candidate length. As an additional sanity check, we compute the Spearman correlation between candidate length and teacher score. Table 12 shows that the correlations are statistically significant but small in magnitude, suggesting limited residual length dependence rather than a dominant length-driven explanation for LCM-based collapse diagnosis.

Length metric	Spearman ρ	p-value
Character length	0.2283	< 0.001
Word length	0.2304	< 0.001

Table 12: Correlation between candidate length and teacher score.

Taken together, these robustness checks support the stability of the LCM diagnosis from four complementary angles. Threshold sensitivity shows that the diagnosed collapse structure is not an artifact of a single teacher-score cutoff. Probability-noise analysis indicates that LCM relies on relative preference ordering rather than finely calibrated probability values. Domain-level statistical testing confirms that Differential Exclusion is consistently more fragile than the other profiled domains, while the length-bias check suggests that candidate length is not the dominant explanation for the teacher-score partition. These results strengthen the interpretation of LCM as a teacher-relative local preference diagnostic rather than a brittle scoring heuristic.

F ACPD Pseudocode

Algorithm 1 Answer-Calibrated Collapse-Aware Preference Distillation

Input: Training examples $\mathcal{D}$, teacher profiles, student model P_θ, profiled domains $\mathcal{D}_{\text{prof}} = \{\text{PR}, \text{DE}, \text{FV}\}$, repair threshold ϵ.
Output: Updated student model.

1. For each example $x \in \mathcal{D}$, construct a structured teacher rationale with four domains: SM, PR, DE, and FV, followed by the final answer.
2. For each profiled domain $d \in \mathcal{D}_{\text{prof}}$, build the teacher-completed prefix c_d and generate verbalized student candidates $\mathcal{C}_d = \{(s_{d,k}, p_{d,k})\}_{k=1}^{K}$, with $\sum_k p_{d,k} = 1$.
3. Score each $s_{d,k}$ with the teacher under the same prefix using mean token log-likelihood, and partition candidates into S^d_{pos} and S^d_{neg}.
4. Compute the local collapse margin:
$$\text{LCM}_d = \sum_{s_{d,k} \in S^d_{\text{pos}}} p_{d,k} - \sum_{s_{d,k} \in S^d_{\text{neg}}} p_{d,k}.$$
5. Select repair domains $D_{\text{rep}} = \{d \in \mathcal{D}_{\text{prof}} : \text{LCM}_d \leq \epsilon\}$.
6. For each pair-valid $d \in D_{\text{rep}}$ with non-empty S^d_{pos} and S^d_{neg}, construct one representative preference pair:
$$s_d^+ = \arg \max_{s_{d,k} \in S^d_{\text{pos}}} p_{d,k}, \qquad s_d^- = \arg \max_{s_{d,k} \in S^d_{\text{neg}}} p_{d,k}.$$
7. Construct answer-level preference pairs by contrasting the reference answer with an incorrect option under the full structured rationale prefix.
8. Train the student with
$$\mathcal{L}_{\text{ACPD}} = \mathcal{L}_{\text{sft}} + \lambda_{\text{dom}}(t)\mathcal{L}_{\text{dom}} + \lambda_{\text{ans}}(t)\mathcal{L}_{\text{ans}}.$$

G Qualitative Repair Case Study

This appendix provides qualitative evidence for how ACPD changes local reasoning behavior in collapse-prone medical domains. We focus on in-domain examples from the MedMCQA 5k evaluation split whose risk slice is `A_high` and whose diagnosed collapsed domain is either *Differential Exclusion* or *Pathological Reasoning*. These examples are used as qualitative case studies rather than as additional quantitative evidence.

Selection protocol. We first identify evaluation samples with $\text{LCM} \leq 0$ in the profiled domains. We then select representative cases from *Differential Exclusion* and *Pathological Reasoning* where ACPD predicts the correct answer while CWT predicts an incorrect answer. The selected cases are chosen for interpretability and expose clear local distractor preferences. This selection is intentionally diagnostic rather than statistically representative: the goal is to inspect the type of local failure that ACPD is designed to repair, while the aggregate effect is evaluated separately by repair rate and collapse-sensitive accuracy in the main experiments.

How to read the cases. These cases analyze local reasoning behavior rather than provide additional quantitative evidence. We focus on the diagnosed collapsed domain and examine whether each model gives stronger support to the teacher-aligned reasoning path than to the competing distractor. Cases 1–3 show *Differential Exclusion* failures, while Case 4 shows a *Pathological Reasoning* failure. A model may mention relevant biomedical concepts yet still fail if its local reasoning favors the wrong distractor; ACPD targets this teacher-relative preference collapse rather than rationale length or fluency. Table 13 summarizes the local contrast in each case.

Compared baselines. The compared baselines highlight complementary failure modes. Standard KD transfers answer-level behavior without local-domain pressure, while Structured CoT-SFT and Reasoning Scaffolding improve rationale format or procedure but can still make an incorrect distractor appear coherent. CWT is the strongest structural baseline because it trains local rationale chunks, yet it treats chunks as supervision units rather than diagnosed failure locations. MI-CoT strengthens rationale-answer coupling and solves some collapse-prone cases, but it does not explicitly locate or repair domains where student preferences shift toward teacher-disapproved continuations.

Error-pattern analysis. In Case 1, most incorrect models recognize that both Mycoplasma and spirochetes are relevant to motility without external flagella, but resolve the ambiguity toward the surface cue that Mycoplasma can exhibit gliding motility. Standard KD, Structured CoT-SFT, Reasoning Scaffolding, and CWT therefore produce fluent rationales while failing to prioritize the teacher-aligned contrast between external flagella and internal axial filaments. ACPD instead treats Mycoplasma as a plausible distractor and assigns stronger local support to spirochetes, matching the diagnosed repair target.

In Case 2, the baselines fail through different distractors: Standard KD and CWT overemphasize

Case	Local Failure and ACPD Repair
Case 1	Mycoplasma motility distracts the model; ACPD emphasizes internal axial filaments in spirochetes.
Case 2	Phagocytosis and healing distract the model; ACPD separates macrophage function and broad healing from antigen presentation.
Case 3	Carbolic-acid irritation distracts the model; ACPD links erethism to mercury poisoning rather than surface irritation cues.
Case 4	Necrotic diathesis and button-collar lesion distract the model; ACPD identifies Azzopardi effect as the specific pathological cue.

Table 13: Summary of the local reasoning failures illustrated in the qualitative cases. Cases 1–3 involve *Differential Exclusion*, while Case 4 involves *Pathological Reasoning*.

healing or granuloma formation, while Structured CoT-SFT and Reasoning Scaffolding are pulled toward phagocytosis. This shows that preserving a structured exclusion paragraph does not guarantee correct option ranking. ACPD repairs the contrast by separating macrophage-associated phagocytosis and broad healing from the teacher-preferred antigen-presentation function. MI-CoT also solves these cases, indicating that the examples are not impossible for strong rationale-answer coupling methods; ACPD differs by explicitly targeting the diagnosed collapsed domain.

In Case 3, local collapse appears in a compact toxicological association rather than a long mechanistic rationale. Standard KD and CWT link blushing and vasodilation to carbolic-acid irritation, whereas ACPD preserves the teacher-aligned association between erethism and mercury poisoning. This suggests that LCM can expose failures where a short medical cue is attached to a plausible but incorrect distractor.

Case 4 extends beyond option-exclusion failures. CWT recognizes the necrotic background of small-cell lung carcinoma but maps the finding to the broader term *necrotic diathesis*. The collapsed domain is *Pathological Reasoning*: the failure is identifying the specific mechanism in which DNA from necrotic tumor cells deposits along vessel walls. ACPD repairs this mechanism-level preference by selecting the Azzopardi effect, showing that collapse-aware repair can target local concept formation rather than only final option elimination.

Case 1: Motility without external flagella

Question: The following organism is motile with no external flagella.

Options: (A) Brucella (B) Mycoplasma (C) Chlamydia (D) Spirochetes

Gold / Collapse: Gold: **D**; risk slice: `A_high`; diagnosed collapsed domain: *Differential Exclusion*; $\mathrm{LCM} = -0.7$

Teacher rationale: Brucella is non-motile, Chlamydia is non-motile and intracellular, and the stem classically points to spirochetes because they are motile through internal axial filaments rather than external flagella. Mycoplasma is a plausible distractor because some species exhibit gliding motility without flagella, but under the teacher's structured interpretation, the stronger discriminative cue is internal axial-filament motility in spirochetes.

Standard KD **Pred.: B; Gold: D; Correct: No**

Collapsed-domain reasoning: Option A (Brucella) is a non-flagellated, intracellular bacterium that is not motile in the traditional sense. Option B (Mycoplasma) is a non-flagellated, cell-wall-deficient bacterium that is motile through a gliding mechanism. Option C (Chlamydia) is a non-flagellated, obligate intracellular bacterium that is not motile in the traditional sense. Option D (Spirochetes) is a non-flagellated, spiral-shaped bacterium that is motile through a unique corkscrew-like movement. Therefore, Option B (Mycoplasma) is the most likely answer.

Final Verdict: The organism described in the question is motile but lacks external flagella. Among the options provided, Mycoplasma is the only organism that is motile through a gliding mechanism and lacks external flagella. Therefore, the correct answer is Option B (Mycoplasma).

Structured CoT-SFT **Pred.: B; Gold: D; Correct: No**

Collapsed-domain reasoning: Option A (Brucella) is not motile. Option C (Chlamydia) is an obligate intracellular parasite and is not motile. Option D (Spirochetes) are motile due to axial filaments, not external flagella. Option B (Mycoplasma) is motile through gliding, which does not involve external flagella, making it the correct choice.

Final Verdict: The correct answer is Option B, as Mycoplasma is motile without external flagella.

MI-CoT **Pred.: D; Gold: D; Correct: Yes**

Collapsed-domain reasoning: Option A (Brucella) is non-motile and lacks flagella. Option B (Mycoplasma) is not selected as the best answer in this setting. Option C (Chlamydia) is non-motile and obligate intracellular. Option D (Spirochetes) is motile via axial filaments, making it the correct choice.

Final Verdict: Option D is the most likely answer as spirochetes are motile without external flagella.

Reasoning Scaffolding **Pred.: B; Gold: D; Correct: No**

Collapsed-domain reasoning: Option A (Brucella) is non-motile and lacks flagella. Option C (Chlamydia) is obligate intracellular and non-motile. Option D (Spirochetes) are motile due to axial filaments, not external flagella. Option B (Mycoplasma) is motile through gliding, not flagella, making it the correct choice.

Final Verdict: The correct answer is Option B, as Mycoplasma is motile without external flagella.

CWT **Pred.: B; Gold: D; Correct: No**

Collapsed-domain reasoning: Option A (Brucella) is not motile and lacks flagella. Option C (Chlamydia) is an obligate intracellular parasite and is not motile. Option D (Spirochetes) are motile due to axial filaments, not external flagella. Option B (Mycoplasma) is motile through a gliding mechanism and lacks external flagella, making it the correct choice.

Final Verdict: Based on the absence of external flagella and the presence of motility, the correct answer is Option B.

ACPD **Pred.: D; Gold: D; Correct: Yes**

Collapsed-domain reasoning: Option A (Brucella) is not motile. Option B (Mycoplasma) is a plausible distractor because it lacks external flagella, but the question is better aligned with organisms that use internal axial filaments. Option C (Chlamydia) is intracellular and non-motile. Option D (Spirochetes) are motile due to axial filaments, which are internal rather than external flagella, making it the teacher-aligned choice.

Final Verdict: The correct answer is Option D, as spirochetes are motile through axial filaments rather than external flagella.

Table 14: Case panel for motility without external flagella. Several baselines contain the relevant biomedical terms, but their local *Differential Exclusion* preference is pulled toward the Mycoplasma distractor. ACPD recovers the teacher-aligned contrast between external flagella and internal axial filaments.

Case 2: Epithelioid cell function in tuberculosis
Question: The most important function of epithelioid cells in tuberculosis is:
Options: (A) Phagocytosis (B) Secretory (C) Antigenic (D) Healing
Gold / Collapse: Gold: **C**; risk slice: `A_high`; diagnosed collapsed domain: *Differential Exclusion*; $\mathrm{LCM} = -0.6$
Teacher rationale: Phagocytosis is primarily associated with macrophage activity, secretory function is not the primary role, and healing is too broad to identify the key function of epithelioid cells in tuberculosis. The teacher-preferred option is antigenic function, corresponding to antigen presentation in the immune response to tuberculosis.

Standard KD **Pred.: D; Gold: C; Correct: No**
Collapsed-domain reasoning: Option A (Phagocytosis) is incorrect because epithelioid cells are not primarily involved in phagocytosis. Option B (Secretory) is incorrect because epithelioid cells are not primarily secretory in the context of tuberculosis. Option C (Antigenic) is incorrectly down-ranked. Option D (Healing) is selected because epithelioid cells are interpreted as important in the healing process of granulomas in tuberculosis.
Final Verdict: The most important function of epithelioid cells in tuberculosis is healing. Therefore, the correct answer is Option D.

Structured CoT-SFT **Pred.: A; Gold: C; Correct: No**
Collapsed-domain reasoning: Option B (Secretory) is incorrect because epithelioid cells are not primarily secretory in function. Option C (Antigenic) is treated as misleading, even though antigen presentation is the teacher-preferred interpretation. Option D (Healing) is incorrect because healing is a broader process involving multiple cell types, not the primary role of epithelioid cells.
Final Verdict: The most important function of epithelioid cells in tuberculosis is phagocytosis, making Option A the correct answer.

MI-CoT **Pred.: C; Gold: C; Correct: Yes**
Collapsed-domain reasoning: Option A (Phagocytosis) is incorrect because epithelioid cells are not primarily phagocytic; this is the role of macrophages. Option B (Secretory) is incorrect as epithelioid cells do not have secretory function as their main role. Option D (Healing) is incorrect because healing is a broader process involving multiple cell types, not specifically epithelioid cells.
Final Verdict: Option C is the most accurate as epithelioid cells are antigen-presenting cells that play a key role in the immune response to tuberculosis.

Reasoning Scaffolding **Pred.: A; Gold: C; Correct: No**
Collapsed-domain reasoning: Option A (Phagocytosis) is selected because epithelioid cells are interpreted as primarily involved in phagocytosis and antigen presentation. Option B (Secretory) is incorrect as epithelioid cells are not primarily secretory in this context. Option C (Antigenic) is down-ranked as secondary to phagocytic activity. Option D (Healing) is incorrect as epithelioid cells are not directly involved in tissue healing in tuberculosis.
Final Verdict: The most important function of epithelioid cells in tuberculosis is phagocytosis, making Option A the correct answer.

CWT **Pred.: D; Gold: C; Correct: No**
Collapsed-domain reasoning: Option A (Phagocytosis) is incorrect because phagocytosis is primarily the function of macrophages, not epithelioid cells. Option B (Secretory) is not a primary function of epithelioid cells. Option C (Antigenic) is treated as vague and not the most accurate description of their role. Option D (Healing) is selected because epithelioid cells contribute to granuloma formation, which is interpreted as part of healing in tuberculosis.
Final Verdict: The most important function of epithelioid cells in tuberculosis is their role in the healing process by forming granulomas. Therefore, the correct answer is Option D.

ACPD **Pred.: C; Gold: C; Correct: Yes**
Collapsed-domain reasoning: Option A (Phagocytosis) is incorrect because epithelioid cells are not primarily phagocytic; this is more directly associated with macrophages. Option B (Secretory) is also incorrect as epithelioid cells are not known for secretory function as their main role. Option D (Healing) is vague and not the primary function being tested. Option C (Antigenic) is the most accurate, as epithelioid cells are involved in antigen presentation to T cells.
Final Verdict: The most important function of epithelioid cells in tuberculosis is antigenic, as they play a central role in antigen presentation and immune response. Therefore, the correct answer is Option C.

Table 15: Case panel for epithelioid-cell function in tuberculosis. The baselines fail through different distractors, either phagocytosis or healing, while ACPD preserves the teacher-aligned local ranking and selects antigenic function.

Case 3: Toxicological sign association
Question: Frequent blushing ("erethism") is associated with poisoning of
Options: (A) Mercury (B) Lead (C) Phenolic acid (D) Carbolic acid
Gold / Collapse: Gold: **A**; risk slice: `A_high`; diagnosed collapsed domain: *Differential Exclusion*; $\mathrm{LCM} = -0.1$
Teacher rationale: Lead poisoning is more commonly associated with abdominal pain, anemia, and neurological symptoms such as wrist drop, not blushing. Phenolic acid and carbolic acid are more associated with irritation or systemic toxicity, but not the specific symptom of erethism. The teacher-preferred answer is mercury poisoning.

Standard KD **Pred.: D; Gold: A; Correct: No**
Collapsed-domain reasoning: Mercury poisoning typically presents with neurological symptoms, such as tremors, ataxia, and cognitive impairment. Lead poisoning is characterized by abdominal pain, constipation, and anemia. Phenolic acid poisoning is less commonly reported and typically presents with gastrointestinal symptoms rather than the characteristic erethism seen with carbolic acid poisoning.
Final Verdict: Based on the characteristic clinical presentation of erethism and the known pathophysiology of carbolic acid poisoning, the most likely answer is Option D.

Structured CoT-SFT **Pred.: A; Gold: A; Correct: Yes**
Collapsed-domain reasoning: Lead poisoning is more commonly associated with neurological and hematological symptoms, such as encephalopathy and anemia, rather than blushing. Phenolic acid and carbolic acid are not typically associated with erethism; they are more relevant in the context of chemical burns or local irritation.
Final Verdict: Based on the clinical association of erethism with mercury poisoning and the lack of such a link with the other options, the most likely answer is Option A.

MI-CoT **Pred.: A; Gold: A; Correct: Yes**
Collapsed-domain reasoning: Option B (Lead) is associated with neurological and renal toxicity but not typically with blushing. Option C (Phenolic acid) and Option D (Carbolic acid) are not commonly linked to this symptom. Option A (Mercury) is known to cause vasomotor instability and is the most likely to be associated with frequent blushing.
Final Verdict: Option A is the most likely answer based on the known toxic effects of mercury on vascular reactivity.

Reasoning Scaffolding **Pred.: A; Gold: A; Correct: Yes**
Collapsed-domain reasoning: Lead poisoning typically presents with neurological and hematological symptoms, not vasomotor instability. Phenolic acid and carbolic acid are not commonly associated with erethism. Mercury, however, is well documented for causing this symptom.
Final Verdict: Based on the clinical association between mercury poisoning and erethism, the most likely answer is Option A.

CWT **Pred.: D; Gold: A; Correct: No**
Collapsed-domain reasoning: Mercury poisoning typically presents with neurological and renal symptoms, not blushing. Lead poisoning is associated with abdominal pain, anemia, and encephalopathy. Phenolic acid is not a recognized toxic agent in standard medical literature and is not associated with blushing. Carbolic acid, however, is a known cause of vasodilation and skin irritation, making it the most likely option.
Final Verdict: Based on the clinical association of blushing with vasodilation and the known effects of carbolic acid, the most likely answer is Option D.

ACPD **Pred.: A; Gold: A; Correct: Yes**
Collapsed-domain reasoning: Lead poisoning typically presents with neurological and renal symptoms, not vasodilation or blushing. Phenolic acid and carbolic acid are not commonly associated with erethism; they are more likely to cause irritation or systemic toxicity without the specific vascular effects seen in mercury poisoning.
Final Verdict: Based on the clinical association between mercury poisoning and frequent blushing, the most likely answer is Option A.

Table 16: Case panel for toxicological sign association. Standard KD and CWT are pulled toward a surface association between blushing and chemical irritation, whereas ACPD preserves the teacher-aligned association between erethism and mercury poisoning.

Case 4: Azzopardi effect in small-cell lung carcinoma
Question: Small cell Ca lung, DNA of the necrotic cells get deposited on vessels?
Options: (A) Psammoma body (B) Azzopardi effect (C) Button collar lesion (D) Necrotic diathesis
Gold / Collapse: Gold: **B**; risk slice: `A_high`; diagnosed collapsed domain: *Pathological Reasoning*; $\mathrm{LCM} = -0.5$
Teacher rationale: In small-cell lung carcinoma, tumor necrosis is common because of rapid growth and poor vascularity. Necrotic material, including DNA, may be deposited along vascular structures. The specific named phenomenon for DNA from necrotic tumor cells deposited on vessel walls is the Azzopardi effect, rather than psammoma bodies, button-collar lesions, or the broader concept of necrotic diathesis.

Standard KD **Pred.: B; Gold: B; Correct: Yes**
Collapsed-domain reasoning: In small cell lung cancer, necrotic cells can release DNA into the vascular space. This phenomenon is known as the Azzopardi effect.
Final Verdict: The phenomenon described in the question is the Azzopardi effect, which is the deposition of DNA from necrotic cells into the vascular space. Therefore, the correct answer is Option B.

Structured CoT-SFT **Pred.: D; Gold: B; Correct: No**
Collapsed-domain reasoning: In small cell lung cancer, the deposition of DNA from necrotic cells on blood vessels is a known phenomenon. This is due to the high mitotic activity and necrosis within the tumor, which leads to the release of nuclear material into the vascular system.
Final Verdict: The most accurate and specific term for the described phenomenon is not explicitly listed, but the best available option is Option D (Necrotic diathesis), as it most closely aligns with the general concept of necrotic changes in the tumor microenvironment.

MI-CoT **Pred.: C; Gold: B; Correct: No**
Collapsed-domain reasoning: In small cell lung cancer, the necrotic cells release DNA, which can be deposited on blood vessels, forming a characteristic pattern known as the button-collar lesion or Azzopardi effect. This is a well-documented histopathological feature in this tumor type.
Final Verdict: Option C is the most accurate and specific term for the described histopathological feature in small cell lung cancer.

Reasoning Scaffolding **Pred.: C; Gold: B; Correct: No**
Collapsed-domain reasoning: In small-cell lung cancer, the DNA of necrotic tumor cells can be released into the bloodstream and deposited on the endothelial lining of blood vessels, forming a characteristic pattern known as the button collar lesion. This is a hallmark of SCLC and is not typically seen in other types of lung cancer.
Final Verdict: The most accurate and specific term for the described histopathological feature in small cell lung cancer is button collar lesion. Therefore, the correct answer is Option C.

CWT **Pred.: D; Gold: B; Correct: No**
Collapsed-domain reasoning: In small cell lung cancer, the DNA of necrotic tumor cells can be released and deposited on blood vessels, a phenomenon that is a known histological feature of this aggressive malignancy.
Final Verdict: The most accurate and specific term for the described histological finding in small cell lung cancer is necrotic diathesis. Therefore, the correct answer is Option D.

ACPD **Pred.: B; Gold: B; Correct: Yes**
Collapsed-domain reasoning: In small cell lung cancer, the necrotic cells release DNA, which can be seen as a fine, thread-like material deposited on blood vessels. This is a characteristic feature of the Azzopardi effect, a known histopathological finding in this malignancy.
Final Verdict: The most accurate and specific histopathological finding described in the question is the Azzopardi effect, which is associated with the deposition of DNA from necrotic cells in small cell lung cancer. Therefore, the correct answer is Option B.

Table 17: Case panel for Azzopardi effect in small-cell lung carcinoma. Unlike the previous *Differential Exclusion* cases, this example illustrates a *Pathological Reasoning* collapse: several baselines recognize DNA deposition from necrotic tumor cells but map it to an overly broad or incorrect histological term. ACPD recovers the teacher-aligned pathological concept and selects Azzopardi effect.