# PACT: LEARNING DIVERSE DIAGNOSTIC STRATEGIES VIA PRIVILEGED SYNTHESIS AND BRANCH CONSENSUS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Artificial intelligence (AI)-based diagnosis has substantial potential to improve the accessibility, efficiency and quality of clinical care. Although existing AI diagnostic methods have demonstrated strong medical reasoning capabilities, most follow a single, fixed reasoning paradigm. In real-world clinical practice, however, physicians may adopt different strategies when questioning patients, gathering clinical evidence and formulating diagnoses. Here we introduce PACT (Periodic Anchor Consensus Training), a model-merging framework that separately learns paradigm-specific LoRA Branches and periodically merges them into a shared Anchor through sign consensus, enabling multiple diagnostic paradigms to be integrated while reducing cross-paradigm interference. Because high-quality clinical dialogues reflecting distinct diagnostic strategies are difficult to collect at scale, we further develop DPS (Doctor Patient Supervisor), which distils dialogues from large language models by simulating four diagnostic strategies; DPS uses complete electronic medical records (EMRs) for quality control while restricting the simulated doctor to patient-visible information, thereby limiting direct exposure to hidden clinical answers. We collect 10,000 de-identified internal-medicine EMRs to construct a dynamic, multi-turn diagnostic benchmark and report higher Strict Diagnosis Accuracy than the evaluated baselines.

## 1 INTRODUCTION

The rapid expansion of online medical consultation platforms has created an urgent need for large language model (LLM)-based diagnostic systems that can conduct safe, effective and clinically meaningful multi-turn consultations Zeng et al. (2020); Valizadeh & Parde (2022). Unlike single-turn medical question answering, diagnosis through dialogue is a sequential decision-making process under uncertainty: the model must actively elicit relevant information, adapt its questions to emerging evidence, resolve diagnostic ambiguities and determine when sufficient evidence has been gathered to reach a diagnosis (Croskerry, 2009; Norman, 2005). These capabilities are essential for translating the medical knowledge encoded in LLMs into practical diagnostic support for real-world consultations.

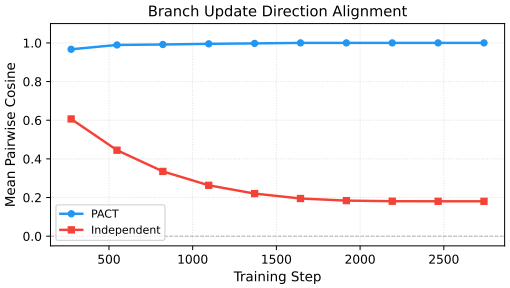


Figure 1: PACT aligns paradigm-specific Branches towards a shared optimization direction. Independently training LoRA Branches on different diagnostic reasoning paradigms causes their parameter updates to diverge, making post-hoc merging ineffective. In contrast, PACT progressively aligns paradigm-specific Branches towards a common optimization direction.

Although large language models (LLMs) have demonstrated promising capabilities in medical reasoning and diagnosis Zhang et al. (2023); Achiam et al. (2023), possessing medical knowledge does not necessarily translate into knowing how to conduct an effective consultation. Most existing LLM-based diagnostic systems are designed to follow a single reasoning paradigm, applying a largely



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

uniform questioning strategy across patients. Clinical diagnosis, however, rarely follows such a fixed path. Different physicians may approach the same case using different strategies; the same physician may adapt their approach to the characteristics of each patient; and even within a single consultation, the physician may shift between strategies as new evidence emerges. For example, a physician may initially trace symptoms step by step towards a likely diagnosis, subsequently compare competing hypotheses in parallel, prioritize questions that rule out urgent conditions, or revisit an earlier hypothesis when it conflicts with newly obtained evidence Croskerry (2009); Norman (2005). Existing LLM-based diagnostic systems, however, are typically trained on dialogues that exhibit a narrow and relatively uniform consultation strategy, encouraging the model to follow the same strategy during inference (Guo et al., 2026; Lai et al., 2025), which limits the models ability to exploit the complementary diagnostic knowledge encoded across different reasoning paradigms.

Bridging this gap requires learning distinct diagnostic reasoning paradigms while integrating them into a unified deployable model. We therefore propose **PACT** (Periodic Anchor Consensus Training), which combines multi-paradigm dialogue synthesis with consensus-based Branch training. This goal presents two challenges. First, mixing paradigms during supervised fine-tuning can blur their distinctive reasoning patterns, whereas training a separate LoRA Branch for each paradigm preserves specialization but produces incompatible updates. As shown in Figure 1, independently trained Branches have a mean pairwise cosine similarity of only 0.18, with 38% of parameters exhibiting conflicts. Second, suitable training data are scarce: clinician-authored dialogues are costly, and existing datasets rarely contain controlled, paradigm-specific diagnostic trajectories Zeng et al. (2020); Liu et al. (2022). Although LLM-based synthesis is scalable, unrestricted EMR access can leak hidden diagnostic information, while unsupervised role-playing often yields repetitive, hallucinated or clinically invalid dialogues. PACT addresses these challenges through periodic sign-consensus aggregation of paradigm-specific Branches and controlled dialogue synthesis with role-specific information access.

Specifically, to endow an LLM with multiple diagnostic reasoning paradigms, PACT constructs multiple LoRA Branches, with each Branch dedicated to learning a distinct diagnostic paradigm. During training, PACT periodically consolidates the compatible updates from these paradigm-specific Branches into a shared Anchor through sign-consensus merging, while regularizing each Branch towards the Anchor between successive aggregation rounds. This process encourages Branch specialization while consolidating compatible parameter updates into a unified LLM. The resulting Anchor can be deployed as a standard single-LoRA model, requiring no paradigm routing or additional adapters at inference. To provide the multi-paradigm supervision required to train these Branches, we further introduce DPS (Doctor Patient Supervisor), a role-separated data-distillation framework inspired by Learning Using Privileged Information (LUPI) Vapnik & Vashist (2009). DPS grants the Supervisor privileged access to complete electronic medical records for clinical validation while restricting the Doctor to information disclosed by the Patient during consultation. This design restricts direct access to privileged information during dialogue synthesis across four reasoning paradigms: *Serial*, *Tree*, *Heap*, and *Feedback*.

Since existing medical benchmarks primarily focus on single-turn question answering or static diagnosis, they provide limited support for assessing whether an LLM agent can actively elicit missing evidence through multi-turn consultation. We therefore construct a dynamic, interactive diagnostic benchmark from 10,000 de-identified internal-medicine electronic medical records, in which agents must sequentially question simulated patients before reaching a diagnosis. PACT improves Strict Diagnosis Accuracy over Mixed-SFT using the same training data. Supplementary checks examine alternative evaluators, patient simulation, synthesis models, and blinded physician ratings. Our contribution can be summarized as:

- We propose PACT, a unified framework that equips a single deployable LLM with multiple complementary diagnostic reasoning paradigms for adaptive multi-turn consultation.
- PACT incorporates two key designs: consensus-based training that periodically integrates paradigm-specific LoRA Branches into a shared Anchor, and DPS, a role-separated data-distillation module that synthesizes multi-paradigm dialogues with role-specific information access and Supervisor review.
- We construct a dynamic multi-turn diagnostic benchmark from 10,000 de-identified internal-medicine electronic medical records. Experiments show Strict Diagnosis Accuracy gains over Mixed-SFT across four backbones under the reported evaluation settings.

## 2 Related Work

**LLM-based medical diagnosis.** Recent medical LLMs have demonstrated strong capabilities in medical question answering, instruction following and clinical reasoning Zhang et al. (2023); Chen et al. (2023; 2024); Zhang et al. (2024); Jeong et al. (2024), with emerging work further extending them to multi-turn diagnostic inquiry (Guo et al., 2026). Nevertheless, most existing methods learn a relatively uniform consultation strategy. This contrasts with clinical diagnosis, in which physicians flexibly combine sequential inquiry, differential diagnosis, urgency-based prioritization and iterative hypothesis revision Elstein et al. (1978); Bowen (2006); Pauker & Kassirer (1980); Croskerry (2009); Norman (2005). Although chain-of-thought and tree-of-thought prompting elicit related reasoning structures at inference time Wei et al. (2022); Yao et al. (2023), PACT incorporates several different complementary diagnostic paradigms directly into LLM training.

**Medical dialogue synthesis.** Existing corpora, such as MedDialog and MedDG, are largely collected from naturally occurring consultations and provide limited control over the underlying reasoning process or doctor–patient information asymmetry Zeng et al. (2020); Liu et al. (2022). AMIE generates diagnostic dialogues through self-play with feedback from a critic aware of the ground-truth diagnosis Tu et al. (2025); DPS emphasizes explicit EMR information partitioning and paradigm-controlled supervision. Synthesis must balance clinical validity against leakage of privileged information from electronic medical records (EMRs) to the simulated Doctor. Inspired by Learning Using Privileged Information Vapnik & Vashist (2009), our DPS separates the information available to the Doctor from that available to a privileged Supervisor, which validates EMR-grounded dialogues with lightweight LLM-as-a-Judge intervention Zheng et al. (2023).

**Parameter-efficient adaptation and Branch aggregation.** LoRA supports lightweight model specialization Hu et al. (2021). Existing merging methods, including model soups, task arithmetic, DARE and TIES, primarily combine independently trained models post hoc and therefore rely on their parameter updates being sufficiently compatible Wortsman et al. (2022); Ilharco et al. (2022); Yu et al. (2024); Yadav et al. (2023). KnOTS aligns LoRA updates through SVD before merging Stoica et al. (2024), while ColD Fusion iteratively recycles task-specific finetuning and fusion to improve a shared model Don-Yehiya et al. (2023). Independently optimized adapters, however, can diverge substantially, reflecting the gradient-conflict and client-drift problems studied in multi-task and federated learning Yu et al. (2020); Liu et al. (2021); Li et al. (2020); Karimireddy et al. (2020). PACT promotes training-time compatibility by periodically consolidating paradigm-specific LoRA Branches into a shared Anchor and aligning them between aggregation rounds, ultimately producing a unified adapter without additional overhead.

## 3 Methodology

Our goal is to train a diagnostic LLM that can acquire missing clinical evidence through multi-turn interaction, apply multiple reasoning paradigms, and produce an accurate diagnosis within a limited number of turns. We formulate this task as a partially observable multi-turn decision process and use the PACT method proposed in Figure 2 to solve the problem. Next, we introduce it in detail.

### 3.1 Periodic Anchor Consensus Training

As mentioned before, different diagnostic paradigms provide complementary supervision, but naively mixing their dialogues may blur paradigm-specific reasoning patterns. Conversely, training an independent adapter for each paradigm prevents cross-paradigm knowledge transfer and requires adapter selection at inference time. We therefore propose Periodic Anchor Consensus Training , which separates specialization from knowledge consolidation. PACT trains a dedicated LoRA Branch for each paradigm and maintains an *Anchor*: a shared LoRA adapter that serves as the regularization reference during training and as the final adapter deployed at inference time. Unlike the Branches, the Anchor is not optimized on any individual dataset; instead, it is periodically reconstructed from their consensus.

Let $\{D_k\}_{k=1}^{K}$ denote the dialogue datasets for $K$ diagnostic paradigms. PACT associates each $D_k$ with a paradigm-specific LoRA update $\Delta^{(k)}$ and maintains a shared Anchor update $\Delta^{\mathrm{anc}}$. All updates

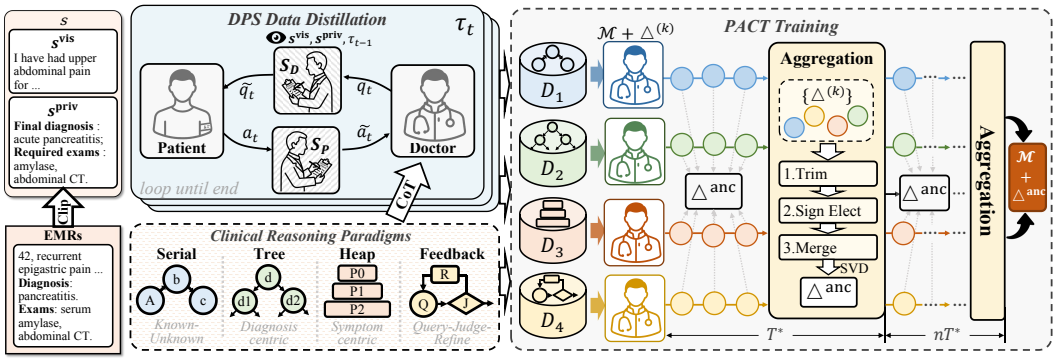


Figure 2: Overview of the PACT framework. **Left:** DPS splits each EMR into patient-visible memory and privileged clinical state, generates paradigm-conditioned Doctor–Patient dialogues, and uses Patient/Doctor Supervisors for minimal PASS-or-REWRITE quality control. **Right:** PACT trains one Branch LoRA per reasoning paradigm and periodically aggregates Branches into a global Anchor via sign-consensus merging, with regularization constraining Branch drift between aggregation rounds.

are initialized to zero. Each Branch is optimized on its corresponding dataset while being regularized toward the current Anchor:

$$\mathcal{L}_{\text{total}}^{(k)} = \mathcal{L}_{\text{task}}^{(k)} + \lambda \left\| \Delta^{(k)} - \Delta^{\text{anc}} \right\|_1, \tag{1}$$

where $\mathcal{L}_{\text{task}}^{(k)}$ is the next-token cross-entropy loss on $D_k$, and $\lambda$ controls Branch–Anchor coupling. The task loss promotes paradigm-specific learning, whereas the $\ell_1$ penalty limits excessive Branch drift and encourages sparse deviations from the shared adapter.

Regularization aligns the Branches with the Anchor but does not transfer their newly learned knowledge into it. After every $T^*$ epochs, PACT therefore updates the Anchor using trim–elect–merge aggregation. First, it retains the largest-magnitude entries in each Branch:

$$\hat{\Delta}^{(k)} = \text{Trim}\big(\Delta^{(k)}, \rho\big), \tag{2}$$

where $\rho$ is the proportion of entries retained. Trimming suppresses weak updates that are more likely to reflect optimization noise.

The retained Branch updates may still disagree in direction. For each parameter coordinate $i$, PACT elects the sign with the larger total retained magnitude:

$$\gamma_i = \text{sign}\left( \sum_{k=1}^{K} \hat{\Delta}_i^{(k)} \right). \tag{3}$$

We use $\text{sign}(0) = 0$, so a zero signed sum contributes zero. This is magnitude-based election, not a count of Branch votes. PACT then identifies the Branches that support the elected direction:

$$\mathcal{S}_i = \left\{ k \;\middle|\; \hat{\Delta}_i^{(k)} \neq 0,\ \text{sign}\big(\hat{\Delta}_i^{(k)}\big) = \gamma_i \right\}. \tag{4}$$

Only these sign-consistent updates contribute to the coordinate-wise merge:

$$\bar{\Delta}_i = \begin{cases} \dfrac{1}{|\mathcal{S}_i|} \displaystyle\sum_{k \in \mathcal{S}_i} \hat{\Delta}_i^{(k)}, & |\mathcal{S}_i| > 0, \\ 0, & |\mathcal{S}_i| = 0. \end{cases} \tag{5}$$

This procedure prevents opposing updates from cancelling one another and retains parameter changes supported by the dominant cross-paradigm direction.

Because coordinate-wise merging can increase matrix rank, the merged update cannot generally be represented by a rank-$r$ LoRA adapter. PACT restores the required parameterization by applying truncated singular value decomposition to each adapted weight matrix:

$$\Delta^{\text{anc}} \leftarrow \text{SVD}_r\big(\bar{\Delta}\big), \tag{6}$$

where $\mathrm{SVD}_r(\cdot)$ denotes the best rank-$r$ approximation under the Frobenius norm. The resulting Anchor is therefore a compact representation of the knowledge on which the paradigm-specific Branches agree.

The updated Anchor serves as the regularization reference in the next training period. The Branches are not reset after aggregation, allowing them to preserve paradigm-specific progress while adapting toward the latest shared consensus. PACT thus alternates between two complementary stages: Branch optimization develops specialized reasoning capabilities, and Anchor aggregation consolidates their compatible components into a single adapter.

After training, all Branches are discarded, and only the frozen base model $\mathcal{M}$ with the final Anchor $\Delta^{\mathrm{anc}}$ is deployed. The resulting model is equivalent in architecture and inference cost to a standard single-LoRA model and requires neither paradigm labels nor adapter routing. Appendix Algorithm 1 provides the complete procedure. Analyses of Branch drift and sign conflict are presented in Figures 3 and 5 and Appendix G.4; theoretical and implementation details are provided in Appendices C and D.

### 3.2 Doctor–Patient–Supervisor Dialogue Distillation

PACT requires paradigm-specific datasets containing multi-turn consultations in which diagnostic evidence is elicited incrementally rather than exposed at the outset. Because such datasets are scarce, we introduce Doctor–Patient–Supervisor (DPS), a controlled distillation framework for synthesizing them. Each DPS sample is a consultation trajectory conditioned on a diagnostic reasoning paradigm: a Doctor asks questions without access to the diagnosis, a Patient answers from the case record, and privileged Supervisors validate both sides of the interaction. The resulting datasets differ in how evidence is sought and organized while remaining grounded in the same clinical cases.

Given a complete clinical case $s$, DPS partitions its information as

$$s = \big(s^{\mathrm{vis}}, s^{\mathrm{priv}}\big), \tag{7}$$

where $s^{\mathrm{vis}}$ contains patient-reportable information that may be disclosed when appropriately elicited, such as symptoms and medical history, and $s^{\mathrm{priv}}$ contains information unavailable to the conversational Doctor, such as the reference diagnosis, unrevealed test results, and treatment plan. The Doctor observes only the dialogue history, whereas the Supervisors may inspect the complete case for validation. The reference diagnosis remains outside the Doctor context, while the Supervisor uses the complete case to review candidate questions.

For paradigm $k$, the Doctor generates a candidate question from the current dialogue trajectory $\tau_{t-1}$:

$$q_t \sim p_D\big(\cdot \mid \tau_{t-1}, k\big). \tag{8}$$

A Doctor-Supervisor $S_D$ then applies a Pass-or-Rewrite operation:

$$\tilde{q}_t = S_D\big(q_t; \tau_{t-1}, s, k\big). \tag{9}$$

The Supervisor passes a valid question unchanged and minimally rewrites it only when it is redundant, clinically inconsistent, prematurely diagnostic, or incompatible with paradigm $k$. Although $S_D$ accesses the complete case, it is restricted to validating the candidate question and must not disclose privileged information.

The Patient answers the validated question using only patient-reportable evidence:

$$a_t \sim p_P\big(\cdot \mid s^{\mathrm{vis}}, \tilde{q}_t\big). \tag{10}$$

A Patient-Supervisor $S_P$ checks the response against the complete case:

$$\tilde{a}_t = S_P\big(a_t; s, \tau_{t-1}, \tilde{q}_t\big). \tag{11}$$

It passes a faithful answer unchanged and minimally corrects hallucinated, contradictory, or improperly disclosed content. In particular, the corrected response may contain only information from $s^{\mathrm{vis}}$ that is relevant to $\tilde{q}_t$. The validated exchange is appended to the trajectory:

$$\tau_t = \big(\tau_{t-1}, \tilde{q}_t, \tilde{a}_t\big). \tag{12}$$

Iterating this procedure produces a clinically grounded consultation in which evidence is revealed through questioning rather than provided directly to the Doctor.

| Method | Size | Diagnosis | | | Examination | | | Process | |
|---|---|---|---|---|---|---|---|---|---|
| | | Strict | Relaxed | Soft | F1 | P | R | Rounds | Coverage |
| Gemini-3.1 Pro | – | 29.00 | 63.93 | 52.63 | 9.22 | 11.03 | 8.74 | 6.61 | 83.50 |
| DeepSeek-R1 | – | 29.39 | 72.20 | 56.85 | 19.26 | 20.21 | 22.45 | **3.32** | 87.00 |
| Doubao-Seed-1.6 | – | 30.08 | 71.69 | 57.75 | <u>25.71</u> | <u>29.53</u> | <u>27.45</u> | 3.96 | 89.60 |
| GPT-5 | – | 33.45 | 64.10 | 54.06 | 12.83 | 14.45 | 13.92 | 5.64 | 79.02 |
| GPT-4.1 | – | <u>36.70</u> | <u>78.82</u> | <u>64.86</u> | 17.26 | 20.40 | 19.76 | <u>3.85</u> | **94.04** |
| PACT (Qwen3) | 8B | **55.31**$^{\dagger}$ | **85.79**$^{\dagger}$ | **76.16**$^{\dagger}$ | **61.33**$^{\dagger}$ | **57.50** | **74.27** | 4.02 | <u>93.39</u> |

Table 1: Comparison with proprietary LLMs on dynamic multi-turn diagnosis. Best results are **bolded** and second best are <u>underlined</u>. $^{\dagger}$ denotes a significant improvement over GPT-4.1 under paired bootstrap ($p < 0.05$).

DPS instantiates four complementary reasoning paradigms inspired by established models of clinical reasoning Elstein et al. (1978); Bowen (2006); Pauker & Kassirer (1980); Schön (2017). *Serial* reasoning evaluates diagnostic hypotheses sequentially. *Tree* reasoning maintains competing hypotheses and asks discriminative questions. *Heap* reasoning prioritizes questions by urgency and expected diagnostic value. *Feedback* reasoning periodically reassesses the available evidence and revises the inquiry strategy. These paradigms define generation strategies rather than mutually exclusive models of clinician cognition.

For each paradigm $k$, DPS produces a dataset

$$D_k = \left\{ \tau^{(n,k)} \right\}_{n=1}^{N_k}, \tag{13}$$

where each $\tau^{(n,k)}$ is a validated, information-asymmetric consultation generated under paradigm $k$. Thus, DPS supplies PACT with clinically consistent dialogues that share the same task objective but exhibit distinct evidence-acquisition patterns. Each PACT Branch learns from one $D_k$, and periodic Anchor aggregation consolidates the compatible knowledge learned across paradigms. Prompt templates and validation criteria are provided in Appendix A, dialogue-quality results in Table 6, and potential biases from privileged supervision in Appendix H.

# 4 Experiments

## 4.1 Experimental Setup

Existing medical benchmarks mainly test single-turn QA or static diagnosis, and therefore do not measure whether a diagnostic agent can collect missing evidence through multi-turn consultation. We therefore construct a dynamic multi-turn diagnosis benchmark and evaluate all models under the same simulator-based protocol.

**Data.** We collect 10,000 de-identified Internal Medicine electronic medical records (EMRs). From this pool, we sample 6,357 EMRs for DPS synthesis, producing 25,428 validated dialogue candidates (4 paradigms per EMR). After quality filtering—removing EMRs where $\geq 2$ paradigm-specific dialogues fail diagnosis matching—we retain 20,684 validated dialogues from 5,171 unique EMRs, evenly distributed across the four reasoning paradigms. These 5,171 EMRs constitute the synthesis-derived training pool and are split at the EMR level into train/validation/test partitions with an 85%/5%/10% ratio. The dynamic evaluation set is constructed separately from the remaining de-identified Internal Medicine pool and contains 438 unique held-out EMR cases; no EMR in this set appears in the DPS synthesis or training partitions. The retained synthesis pool covers 145 unique diagnostic labels.

**Model and Training.** We use Qwen3-8B Yang et al. (2025) as the primary base model, with Qwen3.5-9B, Llama3.1-8B, and Ministral-3-8B for cross-architecture analysis. LoRA adapters are configured with rank $r = 64$, $\alpha = 64$, and applied to all attention projection matrices. Training uses AdamW with learning rate $5 \times 10^{-4}$, cosine schedule, warmup ratio 0.03, and bfloat16 precision. The effective batch size is 64 (per-device 2 $\times$ gradient accumulation 8 $\times$ 4 GPUs). For PACT, we set the synchronization strength $\lambda = 200$, aggregation interval $T^* = 0.25$ epochs, and TIES density $\rho = 0.5$. Additional dataset, preprocessing, evaluation, and hyperparameter details are provided in Appendix B, with sensitivity to $\lambda$ and $T^*$ reported in Appendix Figure 4.

| | | Diagnosis | | | Examination | | | Process | |
|---|---|---|---|---|---|---|---|---|---|
| Method | Size | Strict | Relaxed | Soft | F1 | P | R | Rounds | Coverage |
| Reflect-PubMed | 8B | 14.69 | 51.20 | 39.40 | — | 2.57 | 10.73 | 2.71 | 67.70 |
| DoctorAgent-RL | 8B | 20.36 | 61.43 | 48.74 | — | 2.01 | 7.53 | 4.36 | 65.65 |
| UltraMedical | 8B | 21.77 | <u>77.71</u> | <u>58.67</u> | <u>25.70</u> | <u>27.56</u> | <u>28.07</u> | 3.73 | 71.27 |
| HuatuoGPT2-7B | 7B | 27.85 | 63.64 | 51.24 | 9.99 | 11.71 | 10.47 | <u>1.80</u> | 54.39 |
| HuatuoGPT-o1-8B | 8B | 30.54 | 70.61 | 57.26 | — | 1.54 | 3.36 | **1.72** | 63.88 |
| Doctor-R1 | 8B | <u>33.50</u> | 72.09 | 57.80 | 21.00 | 22.57 | 23.92 | 4.88 | <u>77.00</u> |
| PACT (Qwen3) | 8B | **55.31** | **85.79** | **76.16** | **61.33** | **57.50** | **74.27** | 4.02 | **93.39** |

Table 2: Comparison with medical-specialized LLMs under the same protocol. Best results are **bolded** and second best are <u>underlined</u>; fewer rounds are better. — denotes an unreported value.

**Evaluation Protocol.** We adopt a *dynamic multi-turn evaluation* protocol where the trained model interacts with a GPT-4o-powered patient simulator Hurst et al. (2024) for up to 8 rounds per held-out EMR case. A separate GPT-4o judge scores the completed dialogue on three dimensions: *Diagnosis* (Strict DA for exact top-1 match, Relaxed DA for hierarchical/top-3 credit, and Soft Score as a graded composite), *Examinations* (Exam F1/Precision/Recall against the gold-standard list), and *Process* (Symptom Coverage of key clinical information and average dialogue rounds). Detailed scoring rubrics are in Appendix B. Each EMR case is evaluated over four independent rollouts to reduce simulator variance; statistical tests (Appendix E) are paired at the EMR-case level.

**Baselines.** We compare against *proprietary LLMs* evaluated zero-shot under the same protocol: GPT-5 Singh et al. (2025), GPT-4.1 OpenAI (2025), Gemini-3.1 Pro Google DeepMind (2026), DeepSeek-R1 Guo et al. (2025), and Doubao-Seed-1.6 Seed (2025); and *medical-specialized LLMs*: HuatuoGPT2-7B Chen et al. (2023), HuatuoGPT-o1-8B Chen et al. (2024), UltraMedical Zhang et al. (2024), Doctor-R1 Lai et al. (2025), DoctorAgent-RL Feng et al. (2026), and Reflect-PubMed Jeong et al. (2024). We also test Qwen3-8B, Qwen3.5-9B Yang et al. (2025), Llama3.1-8B Grattafiori et al. (2024), and Ministral-3-8B Liu et al. (2026) for cross-architecture generalization.

### 4.2 Main Results

Tables 1 and 2 compare PACT with proprietary and medically specialized LLMs, respectively. PACT achieves 55.31% Strict DA and a Soft Score of 76.16, substantially outperforming the strongest zero-shot proprietary baseline, GPT-4.1, in Strict DA (+18.61; 95% CI: [15.24, 22.03], $p < 0.0001$), Soft Score (+11.30; 95% CI: [8.68, 13.94], $p < 0.0001$), and Exam F1 (+44.06; 95% CI: [41.26, 46.77], $p < 0.0001$). PACT also improves over its Qwen3-8B base model by 22.78 Strict DA points. Although GPT-4.1 achieves slightly higher symptom coverage than PACT (94.04% vs. 93.39%), the difference is not statistically significant ($p = 0.061$). This result suggests that broad symptom elicitation alone is insufficient for accurate diagnosis; PACT additionally benefits from the structured diagnostic reasoning acquired through multi-paradigm consensus training.

Most medically specialized LLMs are optimized for single-turn question answering or static reasoning rather than interactive diagnosis Lai et al. (2025); Feng et al. (2026). Doctor-R1, which incorporates multi-turn interaction training, is therefore the strongest medically specialized baseline. On Qwen3-8B, Mixed-SFT using the same DPS data already substantially outperforms Doctor-R1 in Strict DA (51.48% vs. 33.50%; Tables 2 and 3). PACT further improves over Mixed-SFT by 3.82 Strict DA points (95% CI: [1.31, 6.28], $p = 0.003$) and increases Exam F1 from 59.83% to 61.33%.

| | Strict DA | | | Exam F1 | | |
|---|---|---|---|---|---|---|
| Base model | Base | Mixed | PACT | Base | Mixed | PACT |
| Qwen3.5-9B | 35.90 | 48.29 | **49.77** | 15.00 | 62.45 | **63.11** |
| Qwen3-8B | 32.53 | 51.48 | **55.31** | 21.53 | 59.83 | **61.33** |
| Llama3.1-8B | 16.89 | 51.37 | **53.03** | 16.68 | 45.87 | **60.92** |
| Ministral-3-8B | 20.94 | 36.42 | **50.29** | 5.08 | 28.61 | **45.71** |

Table 3: Cross-architecture results obtained using different base models. Full metrics are reported in Appendix Table 15.

Table 3 further examines whether these gains depend on the choice of base model. Across Qwen3.5-9B, Qwen3-8B, Llama3.1-8B, and Ministral-3-8B, PACT consistently outperforms both the corresponding base model and Mixed-SFT in Strict DA and Exam F1. The gains over Mixed-SFT are particularly pronounced for Llama3.1-8B in Exam F1 (+15.05) and Ministral-3-8B in both Strict

DA (+13.87) and Exam F1 (+17.10). These results demonstrate that PACT is not tied to a specific backbone and generalizes across distinct model families.

Because our primary experiments consider dialogues of up to eight rounds, we further assess whether PACT remains effective over longer interactions. As shown in Table 4(a), increasing $T_{\max}$ from 8 to 16 changes Strict DA by only $+0.28$ points and increases the average number of dialogue rounds by merely $0.11$. These results demonstrate that PACT remains robust under an extended interaction horizon and that its performance gains are not artifacts of the eight-round limit. Moreover, because the evaluations in Tables 1 and 2 are conducted in Chinese, we further evaluate PACT in an English setting to assess its cross-lingual generalizability. In the English-translated setting, PACT outperforms the Base model by $14.44$ points in Strict DA and $18.65$ points in Exam F1, while reducing the average number of dialogue rounds from $3.73$ to $1.27$ (Table 4 (b)).

| Method | $T_{\max}$ | DA | F1 | Avg Rnd |
|---|---|---|---|---|
| Mixed-SFT | 8 | 51.48 | 59.83 | 4.17 |
| Mixed-SFT | 16 | 51.8 | 60.09 | 4.35 |
| PACT | 8 | 55.31 | 61.33 | 4.02 |
| PACT | 16 | **55.59** | **61.55** | 4.13 |

(a) Maximum dialogue horizon.

| Setting | DA | F1 | Rnd | Cov |
|---|---|---|---|---|
| English Base | 33.39 | 20.13 | 3.73 | **77.05** |
| English PACT | **47.83** | **38.78** | **1.27** | 62.79 |

(b) English-translated evaluation.

Table 4: Dialogue-horizon sensitivity and English transfer on Qwen3-8B.

### 4.3 Ablation Study

**Impact of the proposed training paradigm.** The training paradigm is crucial for jointly learning diagnostic and examination capabilities. We compare PACT with Mixed-SFT, sequential learning, FedAvg McMahan et al. (2017), and three post-hoc merging methods. As shown in Table 5, PACT achieves the highest Strict DA and Exam F1, significantly outperforming Mixed-SFT in Strict DA ($+3.82$, $p = 0.003$). The Strict DA gain over Mixed-SFT is significant; the Exam F1 and Soft Score differences are not significant at the 0.05 level (Appendix E).

| Strategy | DA | F1 | Cov |
|---|---|---|---|
| Mixed-SFT | 51.48 | 59.83 | 93.72 |
| Sequential | 50.17 | 57.86 | **94.72** |
| FedAvg | 54.39 | 18.74 | 94.11 |
| Indep. + Linear | 52.91 | 56.27 | 93.92 |
| Indep. + TIES | 40.41 | 59.03 | 89.81 |
| Indep. + DARE-TIES | 52.45 | 57.16 | 93.66 |
| PACT | **55.31**$^*$ | **61.33** | 93.39 |

Table 5: Training-strategy ablation on Qwen3-8B. $^*$: paired-bootstrap gain over Mixed-SFT ($p < 0.05$).

**Impact of the DPS synthesis module.** High-quality data distillation is critical to PACT. We therefore compare full DPS with single-model synthesis, Doctor–Patient role separation, and variants without either Supervisor. As shown in Table 6, full DPS achieves the lowest label leakage and invalid-behavior rates and the highest diagnosis and examination alignment, supporting these controls under the reported audit. The measured explicit label leakage rate is 2.1%.

| Setting | Leak.↓ | Inv.↓ | Diag.↑ | Exam↑ | Rew. |
|---|---|---|---|---|---|
| Single-model | 18.7 | 24.5 | 71.2 | 62.8 | – |
| Doctor-Patient | 6.8 | 17.9 | 78.4 | 68.5 | – |
| w/o Patient-S | 4.9 | 11.8 | 84.6 | 74.3 | 13.2 |
| w/o Doctor-S | 7.6 | 9.4 | 81.5 | 70.9 | 9.7 |
| DPS full | **2.1** | **5.8** | **91.3** | **82.6** | 16.4 |

Table 6: DPS data quality and Supervisor ablation (%).

**Cross-model checks and clinician assessment.** Table 7 summarizes three cross-model checks. Replacing GPT-4o with DeepSeek-V3 preserves the four-system ranking on 100 cases (Spearman $\rho$=1.00). The alternative-patient test evaluates PACT with a different simulator, and the alternative-synthesis test assesses DPS dialogue quality. In a double-blind assessment by three physicians, PACT receives the highest Overall Clinical Competence score ($3.87\pm0.55$), exceeding Mixed-SFT ($3.48\pm0.63$; $p$=0.001). Protocols appear in Appendix G.7.

| Method | GPT-4o | DS-V3 | Rank |
|---|---|---|---|
| GPT-4.1 | 37.00 | 35.50 | Yes |
| Doctor-R1 | 34.00 | 32.50 | Yes |
| Mixed-SFT | 52.00 | 50.50 | Yes |
| PACT | **56.00** | **54.00** | Yes |

(a) Cross-judge Strict DA ($n$=100).

| Patient simulator | Strict DA | Exam F1 |
|---|---|---|
| GPT-4o | 56.00 | 62.15 |
| DeepSeek-R1 | 53.75 | 59.83 |

(b) Cross-simulator PACT ($n$=100).

| Synthesis model | Diag. | Leak. | Exam |
|---|---|---|---|
| GPT-4o | 91.3% | 2.1% | 82.6% |
| DeepSeek-R1 | 87.5% | 3.4% | 78.2% |

(c) Cross-synthesis DPS ($n\approx$200).

| Method | Reasoning | Efficiency | Overall |
|---|---|---|---|
| GPT-4.1 | 2.84±0.72 | 2.96±0.81 | 2.78±0.69 |
| Doctor-R1 | 2.71±0.68 | 2.62±0.75 | 2.58±0.71 |
| Mixed-SFT | 3.52±0.65 | 3.41±0.70 | 3.48±0.63 |
| **PACT** | **3.91±0.58** | **3.82±0.62** | **3.87±0.55** |

(d) Double-blind clinician ratings ($n$=50).

Table 7: Robustness across judges, simulators, synthesis models, and three physicians. Clinician scores are mean ± standard deviation on a 1–5 scale.

**Further analyses.** We evaluate the contribution of each reasoning paradigm by removing its training dialogues from Mixed-SFT, and separately ablate the regularization and aggregation methods in PACT. Tables 8a–8c show that removing any paradigm reduces Exam F1 by 5.3–6.5 points, L1 Anchor gives the strongest Strict DA, and TIES preserves examination quality better than Linear Avg or SLERP Li et al. (2020); Goddard et al. (2024); Yadav et al. (2023). Table 3 shows that PACT consistently outperforms Mixed-SFT on Strict DA across four architectures. Independent Branch deployment further reveals complementary specializations despite high pairwise delta alignment (Figure 6) (Appendix G.5).

| Setting | DA | F1 | Cov |
|---|---|---|---|
| All four | 51.48 | **59.83** | **93.72** |
| w/o Serial | 51.43 | 54.51 | 93.21 |
| w/o Tree | 50.34 | 53.87 | 93.22 |
| w/o Heap | 50.23 | 53.64 | 92.69 |
| w/o Feedback | **51.9** | 53.29 | 92.94 |

(a) Reasoning-paradigm data ablation under Mixed-SFT.

| Regularizer | DA | F1 | Cov |
|---|---|---|---|
| None | 33.45 | 56.96 | 89.83 |
| L2 Prox | 33.16 | 58.00 | 89.23 |
| Cosine | 54.05 | **62.32** | 91.02 |
| L1 Anchor | **55.31** | 61.33 | **93.39** |

(b) Regularization.

| Aggregator | DA | F1 | Cov |
|---|---|---|---|
| Linear Avg | 55.19 | 33.74 | 93.28 |
| SLERP | 54.45 | 36.43 | **93.68** |
| TIES | **55.31** | **61.33** | 93.39 |

(c) Model merge methods.

Table 8: Component ablations on Qwen3-8B. DA, F1, and Cov denote Strict DA, Exam F1, and Symptom Coverage.

## 5 Conclusion

We introduce PACT, a model-merging framework for integrating multiple diagnostic reasoning paradigms within a unified AI diagnostic system. PACT learns paradigm-specific LoRA branches and periodically merges them into a shared anchor through sign consensus, thereby preserving complementary diagnostic strategies while mitigating cross-paradigm interference. To address the scarcity of high-quality clinical dialogues representing diverse diagnostic practices, we further develop DPS, a DoctorPatientSupervisor framework that distils multi-turn dialogues from large language models by simulating four diagnostic strategies. DPS uses complete EMRs for quality control while restricting the simulated doctor to patient-visible information, reducing direct exposure to hidden clinical answers. Using 10,000 de-identified internal-medicine EMRs, we construct a dynamic, multi-turn diagnostic benchmark on which PACT achieves higher Strict Diagnosis Accuracy than the evaluated baselines. Together, these results demonstrate the potential of combining diverse reasoning paradigms to improve interactive AI diagnosis.

**Limitation** Several directions remain for future work. Our evaluation primarily uses retrospective data from a single clinical department, complemented by synthetic dialogues and model-based assessment. Broader validation across institutions, specialties and native-language clinical settings would help further assess the generalizability of PACT. Expanding clinician evaluations and examining additional diagnostic scenarios could also provide a more comprehensive understanding of model behaviour and support its continued development towards prospective clinical applications.

### AI USE STATEMENT

Generative AI systems are intrinsic to the experimental methodology: they are used for DPS dialogue synthesis, as evaluation-time simulators and judges, and as research baselines, with their roles and controls documented throughout the paper. Generative AI tools were also used to assist language editing and LaTeX formatting. The authors retain responsibility for the final content.

### ETHICS STATEMENT

The study uses de-identified EMRs and evaluates research prototypes for diagnostic dialogue; it does not study clinical deployment or provide medical advice. The clinician study evaluates generated transcripts rather than intervening in patient care. Because diagnostic systems can produce unsafe or overconfident outputs, all results are framed as research evidence rather than evidence of clinical readiness. The internal EMRs and derived artifacts remain restricted to research use and are not released for unrestricted redistribution.

### REPRODUCIBILITY STATEMENT

The main paper describes the reported data split, evaluation metrics, baselines, and optimization procedure. Appendix B provides dataset and scoring details; Appendix C states the theoretical assumptions; Appendix D gives implementation details; and Appendices E and G report statistical tests, full results, sensitivity analyses, and robustness checks.

## REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report, 2023.

Judith L Bowen. Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine*, 355(21):2217–2225, 2006.

Junying Chen, Xidong Wang, Ke Ji, Anningzhe Gao, Feng Jiang, Shunian Chen, Hongbo Zhang, Dingjie Song, Wenya Xie, Chuyi Kong, et al. Huatuogpt-ii, one-stage training for medical adaption of llms. *arXiv preprint arXiv:2311.09774*, 2023.

Junying Chen, Zhenyang Cai, Ke Ji, Xidong Wang, Wanlong Liu, Rongsheng Wang, Jianye Hou, and Benyou Wang. Huatuogpt-o1, towards medical complex reasoning with llms, 2024.

Pat Croskerry. A universal model of diagnostic reasoning. *Academic medicine*, 84(8):1022–1028, 2009.

Shachar Don-Yehiya, Elad Venezian, Colin Raffel, Noam Slonim, and Leshem Choshen. ColD Fusion: Collaborative descent for distributed multitask finetuning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 788–806, 2023. doi: 10.18653/v1/2023.acl-long.46. URL https://aclanthology.org/2023.acl-long.46/.

Arthur S. Elstein, Lee S. Shulman, and Sarah A. Sprafka. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Harvard University Press, 1978.

Yichun Feng, Jiawei Wang, Lu Zhou, Zhen Lei, and Yixue Li. Doctoragent-rl: A multi-agent collaborative reinforcement learning system for multi-turn clinical dialogue. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 16952–16956. IEEE, 2026.

Charles Goddard, Shamane Siriwardhana, Malikeh Ehghaghi, Luke Meyers, Vladimir Karpukhin, Brian Benedict, Mark McQuade, and Jacob Solawetz. Arcee's mergekit: A toolkit for merging large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pp. 477–485, 2024.

Google DeepMind. Gemini 3.1 Pro model card. https://deepmind.google/models/model-cards/gemini-3-1-pro/, 2026. Google DeepMind, February 2026.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Yue Guo, Fanfu Wang, Jianwei Lv, Xincheng Shi, Yuchen Li, Youya Wang, Yunsheng Zeng, Yujing Liu, Yunhao Qiao, Gen Li, et al. Dr. assistant: Enhancing clinical diagnostic inquiry via structured diagnostic reasoning data and reinforcement learning. *arXiv preprint arXiv:2601.13690*, 2026.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv e-prints*, pp. arXiv–2106, 2021.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022.

Minbyul Jeong, Jiwoong Sohn, Mujeen Sung, and Jaewoo Kang. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. *Bioinformatics*, 40(Supplement_1):i119–i129, 2024.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pp. 5132–5143. PMLR, 2020.

Yunghwei Lai, Kaiming Liu, Ziyue Wang, Weizhi Ma, and Yang Liu. Doctor-r1: Mastering clinical inquiry with experiential agentic reinforcement learning. *arXiv preprint arXiv:2510.04284*, 2025.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

Alexander H Liu, Kartik Khandelwal, Sandeep Subramanian, Victor Jouault, Abhinav Rastogi, Adrien Sadé, Alan Jeffares, Albert Jiang, Alexandre Cahill, Alexandre Gavaudan, et al. Ministral 3. *arXiv preprint arXiv:2601.08584*, 2026.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in neural information processing systems*, 34:18878–18890, 2021.

Wenge Liu, Jianheng Tang, Yi Cheng, Wenjie Li, Yefeng Zheng, and Xiaodan Liang. Meddg: an entity-centric medical consultation dataset for entity-aware medical dialogue generation. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 447–459. Springer, 2022.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. Pmlr, 2017.

Geoffrey Norman. Research in clinical reasoning: past history and current trends. *Medical education*, 39(4):418–427, 2005.

OpenAI. Introducing GPT-4.1 in the API. https://openai.com/index/gpt-4-1/, 2025. OpenAI, April 2025.

Stephen G Pauker and Jerome P Kassirer. The threshold approach to clinical decision making. *New England Journal of Medicine*, 302(20):1109–1117, 1980.

Donald A Schön. *The reflective practitioner: How professionals think in action*. Routledge, 2017.

ByteDance Seed. Introduction to techniques used in seed1. 6. *Online. Accessed*, pp. 12–21, 2025.

Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, et al. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*, 2025.

George Stoica, Pratik Ramesh, Boglarka Ecsedi, Leshem Choshen, and Judy Hoffman. Model merging with SVD to tie the Knots. *arXiv preprint arXiv:2410.19735*, 2024. URL https://arxiv.org/abs/2410.19735.

Tao Tu, Mike Schaekermann, Anil Palepu, et al. Towards conversational diagnostic artificial intelligence. *Nature*, 642:442–450, 2025. doi: 10.1038/s41586-025-08866-7. URL https://www.nature.com/articles/s41586-025-08866-7.

Mina Valizadeh and Natalie Parde. The ai doctor is in: A survey of task-oriented dialogue systems for healthcare applications. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6638–6660, 2022.

Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural networks*, 22(5-6):544–557, 2009.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International conference on machine learning*, pp. 23965–23998. PMLR, 2022.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. *Advances in neural information processing systems*, 36:7093–7115, 2023.

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

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

Le Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In *Forty-first International Conference on Machine Learning*, 2024.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020.

Guangtao Zeng, Wenmian Yang, Zeqian Ju, Yue Yang, Sicheng Wang, Ruisi Zhang, Meng Zhou, Jiaqi Zeng, Xiangyu Dong, Ruoyu Zhang, et al. Meddialog: Large-scale medical dialogue datasets. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pp. 9241–9250, 2020.

Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu, Zhihong Chen, Guiming Chen, Jianquan Li, Xiangbo Wu, Zhang Zhiyi, Qingying Xiao, et al. Huatuogpt, towards taming language model to be a doctor. In *Findings of the association for computational linguistics: EMNLP 2023*, pp. 10859–10885, 2023.

Kaiyan Zhang, Sihang Zeng, Ermo Hua, Ning Ding, Zhang-Ren Chen, Zhiyuan Ma, Haoxin Li, Ganqu Cui, Biqing Qi, Xuekai Zhu, et al. Ultramedical: Building specialized generalists in biomedicine. *Advances in Neural Information Processing Systems*, 37:26045–26081, 2024.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36:46595–46623, 2023.

## A Prompt Specifications and Generation Controls

DPS synthesis uses Chinese prompts to match the source EMRs and target dialogue setting. The English descriptions below summarize role instructions, control rules, and input placeholders.

**Shared Doctor instruction.** You are an experienced and logically rigorous doctor conducting an online consultation. Your goal is to collect information through multi-turn interaction and, before the final turn, provide both examination recommendations and a preliminary diagnosis. Examination recommendations include laboratory tests, imaging tests, and other auxiliary examinations. The preliminary diagnosis should be based on the chief complaint, history, physical signs, and available examinations; because information is incomplete, it should be framed as a preliminary rather than final diagnosis. Information should be organized in a medical-record format covering basic information, chief complaint, history of present illness, past medical history, allergy history, and physical examination. The Doctor receives the dialogue history `{{history}}` and the latest patient message `{{message}}`, reasons internally, and outputs only the direct patient-facing reply.

**Serial Doctor template.** The Doctor follows a step-by-step consultation policy. The internal reasoning template contains: (1) known information organized by medical-record fields, (2) diagnostic hypotheses matching the current evidence, (3) missing information to be collected, (4) a decision to either collect more information or give the final opinion, and (5) a response strategy. This template encourages linear hypothesis testing and asks one or two high-priority questions per turn.

**Tree Doctor template.** The Doctor maintains a set of candidate diseases and uses high-discrimination questions to narrow the differential diagnosis. The internal reasoning template contains: (1) known information, (2) current suspected diseases, (3) discriminative features among these diseases, (4) the best question for maximizing information gain, (5) a collect-or-finalize decision, and (6) a response strategy.

**Heap Doctor template.** The Doctor prioritizes information by diagnostic importance, urgency, and specificity. The internal reasoning template contains: (1) known information, (2) analysis of known symptoms with high/medium/low importance weights, (3) the most important symptoms to lock onto, (4) deeper inquiry over missing high-value evidence, (5) a collect-or-finalize decision, and (6) a response strategy.

**Feedback Doctor template.** The Doctor performs a generate–critique–revise rehearsal before responding. The internal reasoning template contains: (1) known information, (2) an initial response plan, (3) error detection for logical gaps, insufficient evidence, or previously denied symptoms, (4) control revision, and (5) the final decision. This template encourages reflection before producing the patient-facing reply.

**Patient template.** The Patient is instructed to role-play an online consultation patient using only patient-visible facts `{{patient_memory}}` and the dialogue history `{{history}}`. If the history is empty, the Patient starts with a short natural opening based only on the chief complaint. The Patient must convert medical-record language into lay expressions, answer only what the Doctor asks, avoid disclosing the full record at once, and answer "not sure", "did not notice", or "have not done that test" when the requested information is absent from the visible memory.

**Doctor-Supervisor template.** The Doctor-Supervisor acts as a privileged clinical supervisor with access to the full ground-truth EMR `{{ground_truth}}`, target diagnosis `{{diagnosis}}`,

| Item | Value |
|---|---|
| De-identified EMR pool | 10,000 |
| Clinical department | Internal Medicine |
| EMRs sampled for DPS synthesis | 6,357 |
| Raw synthesized dialogues | 25,428 |
| Retained unique EMRs after filtering | 5,171 |
| Retained synthesized dialogues | 20,684 |
| Reasoning paradigms per EMR | 4 |
| Train / validation / test split | 85% / 5% / 10% |
| Unique diagnostic labels in retained pool | 145 |
| Held-out dynamic evaluation EMRs | 438 |
| Rollouts per held-out EMR | 4 |
| Total dynamic evaluation episodes | 1,752 |
| Maximum dialogue rounds | 8 |

Table 9: Dataset statistics for DPS synthesis and dynamic evaluation.

dialogue history `{{history}}`, and Doctor draft `{{draft_response}}`. It checks whether the Doctor should receive PASS or REWRITE. Required rewrites include severe diagnostic drift, missed high-risk clues, logical dead ends, missing key examinations at closure, failure to use already collected evidence, over-suspicion, and over-examination. The Supervisor must preserve knowledge isolation: it may use privileged information to judge the process, but its rewritten patient-facing response must not reveal facts that the Doctor could not know from the dialogue. Rewrites are limited to adding necessary questions, removing low-value questions, or appending missing examination suggestions.

**Patient-Supervisor template.** The Patient-Supervisor receives the complete case $s$, including its patient-visible subset $s^{\text{vis}}$ (`{{patient_profile}}`), dialogue history `{{history}}`, the current validated Doctor question $\tilde{q}_t$, and the Patient draft `{{draft_response}}`. Consistent with the main method, the complete case is available for validation, but the rewritten Patient reply must contain only patient-visible facts relevant to the question. Hidden diagnoses or undisclosed examinations must not be introduced; information absent from the visible profile should be treated as unknown. Required rewrites address fabricated symptoms, medical jargon, over-answering, and role inconsistencies. The output specifies a correction analysis, a PASS-or-REWRITE decision, and a rewritten reply when necessary. Only the accepted Patient reply enters the dialogue; the Supervisor analysis and privileged case remain outside the Doctor context.

# B Dataset, Evaluation, and Hyperparameter Details

## B.1 Dataset Statistics

Table 9 provides the dataset statistics used in our experiments. The reported split unit is the EMR.

## B.2 Evaluation Rubric

Table 10 summarizes the metrics used by the GPT-4o judge in dynamic multi-turn evaluation. The judge evaluates both outcome quality and consultation process quality after each completed dialogue.

## B.3 Training and Evaluation Hyperparameters

Table 11 lists the main training and evaluation hyperparameters. Unless otherwise specified, the same settings are used for the primary Qwen3-8B experiments and cross-architecture runs.

**Preprocessing.** Dialogue data are converted into chat-style training examples by mapping Patient utterances to user turns and Doctor utterances to assistant turns. Loss is applied only to assistant

| Dimension | Metric | Definition |
|---|---|---|
| Diagnosis | Strict DA | The top-1 diagnosis exactly matches the ground-truth diagnosis or an accepted synonym. |
| Diagnosis | Relaxed DA | Credits strict matches, hierarchical matches, or top-3 diagnostic hits. |
| Diagnosis | Soft Score | Assigns 1.0 / 0.7 / 0.3 / 0 for strict, hierarchical, top-3, and missed diagnoses. |
| Examination | Exam Precision | Fraction of recommended examinations that match the gold-standard examination list. |
| Examination | Exam Recall | Fraction of gold-standard examinations covered by the model's recommendations. |
| Examination | Exam F1 | Harmonic mean of Exam Precision and Exam Recall. |
| Process | Symptom Coverage | Coverage of key clinical information collected during the dialogue. |
| Efficiency | Avg Rnd | Average number of dialogue rounds before termination. |

Table 10: Evaluation rubric for dynamic multi-turn diagnosis.

| Hyperparameter | Value |
|---|---|
| LoRA rank $r$ | 64 |
| LoRA $\alpha$ | 64 |
| LoRA target modules | Attention projection matrices |
| Optimizer | AdamW |
| Learning rate | $5 \times 10^{-4}$ |
| Schedule | Cosine |
| Warmup ratio | 0.03 |
| Precision | bfloat16 |
| Effective batch size | 64 |
| Number of GPUs | 4 |
| PACT synchronization strength $\lambda$ | 200 |
| Aggregation interval $T^*$ | 0.25 epochs |
| TIES density $\rho$ | 0.5 |
| Maximum training sequence length | 4096 |
| Dynamic evaluation episodes per EMR | 4 |
| Dynamic evaluation maximum rounds | 8 |

Table 11: Main hyperparameters used for training and evaluation.

| Stage | EMRs | Dialogues |
|---|---|---|
| Raw DPS synthesis | 6,357 | 25,428 |
| Round/validity filtered | 6,128 | 24,512 |
| Record-level quality filtered | 5,171 | 20,684 |

Table 12: DPS filtering stages for the synthesis-derived training pool.

responses. Residual thinking tags are removed before training, and models without a built-in chat template are assigned a ChatML-style template for consistent formatting.

### B.4 Dataset Filtering Analysis

Table 12 summarizes the filtering stages used to construct the final DPS training pool. We first remove very short or invalid dialogue trajectories, and then apply record-level quality filtering: if at least two paradigm-specific dialogues from the same EMR fail diagnosis matching, all four dialogues from that EMR are removed. This rule preserves paradigm balance in the retained set, but may bias the training pool toward cases with clearer diagnostic trajectories.

The record-level quality filter removes 957 EMRs from the round-filtered pool, corresponding to 3,828 paradigm-specific dialogues. Among removed EMRs, 685 have two failed paradigms, 217 have three failed paradigms, and 55 fail across all four paradigms. This record-level filtering preserves four paradigm-specific dialogues for each retained EMR.

# C THEORETICAL DISCUSSION

We analyze the materialized LoRA updates used by PACT through exact algebraic properties and conditional optimization bounds in the low-rank factor space.

## C.1 LOW-RANK PARAMETERIZATION

For one adapted weight matrix, let $B \in \mathbb{R}^{m\times r}$, $A \in \mathbb{R}^{r\times n}$, and $\Delta = BA$. The fixed LoRA scaling is absorbed into the factors; in our configuration $\alpha/r = 1$. Let $C$ denote the Anchor during one training period. For Branch $k$, define

$$F_k(A, B; C) = \mathcal{L}_k(BA) + \lambda\|BA - C\|_1, \tag{14}$$

where $\|\cdot\|_1$ is the entrywise sum of absolute values and $\mathcal{L}_k$ is a fixed empirical task loss. For multiple adapted matrices, the penalty is summed over modules and all norms below apply to their concatenated entries. The results concern this objective; any implementation-specific normalization must be incorporated into its effective coefficient.

## C.2 MAGNITUDE-BASED SIGN CONSENSUS

Simple averaging can cancel updates of opposite signs. PACT uses the same trim–elect–merge operator as Eqs. 2–5, following TIES Yadav et al. (2023). For the retained values $\hat{\Delta}_i^{(k)}$, write

$$\begin{aligned} p_i &= \sum_{k:\hat{\Delta}_i^{(k)}>0} \hat{\Delta}_i^{(k)}, \qquad n_i = \sum_{k:\hat{\Delta}_i^{(k)}<0} |\hat{\Delta}_i^{(k)}|, \\ \gamma_i &= \operatorname{sign}(p_i - n_i) = \operatorname{sign}\left(\sum_k \hat{\Delta}_i^{(k)}\right), \qquad \operatorname{sign}(0) = 0. \end{aligned} \tag{15}$$

Define $\mathcal{S}_i = \{k : \hat{\Delta}_i^{(k)} \neq 0,\ \operatorname{sign}(\hat{\Delta}_i^{(k)}) = \gamma_i\}$. The merged coordinate is

$$\bar{\Delta}_i = \begin{cases} |\mathcal{S}_i|^{-1} \sum_{k\in\mathcal{S}_i} \hat{\Delta}_i^{(k)}, & |\mathcal{S}_i| > 0, \\ 0, & |\mathcal{S}_i| = 0. \end{cases} \tag{16}$$

Thus election is based on total magnitude, not the number of supporting Branches. A zero signed sum, including the all-zero case, produces zero. For example, retained values $(1, 1, 1, -10)$ elect the negative sign. This operator avoids cancellation within the selected group, but neither sign agreement nor a large retained magnitude guarantees low task loss. Trimming and subsequent rank projection can still discard useful information.

## C.3 CONDITIONAL CONTROL OF BRANCH DRIFT

**Proposition 1 (fixed-Anchor objective comparison).** Fix $C$ and $\lambda > 0$ during a training period. Let $\Delta_0 = B_0A_0$ and $\Delta_1 = B_1A_1$ be its starting and ending updates. Suppose the same empirical task loss satisfies $\mathcal{L}_k(\Delta) \geq \ell_k$ and the endpoints obey

$$F_k(A_1, B_1; C) \leq F_k(A_0, B_0; C) + \varepsilon, \qquad \varepsilon \geq 0. \tag{17}$$

Then

$$\|\Delta_1 - C\|_F \leq \|\Delta_1 - C\|_1 \leq \|\Delta_0 - C\|_1 + \frac{\mathcal{L}_k(\Delta_0) - \ell_k + \varepsilon}{\lambda}. \tag{18}$$

*Proof.* Expanding Eq. 17 and rearranging gives $\lambda\|\Delta_1 - C\|_1 \leq \mathcal{L}_k(\Delta_0) - \mathcal{L}_k(\Delta_1) + \lambda\|\Delta_0 - C\|_1 + \varepsilon$. Use the loss lower bound and $\|X\|_F \leq \|X\|_1$. □

Cross-entropy permits $\ell_k = 0$. The proposition applies directly to low-rank factor iterates and requires no PL assumption, parameter-count surrogate, or unconstrained stationarity. It bounds possible drift relative to the period's initial residual; it does not imply contraction. AdamW and stochastic mini-batch updates do not automatically guarantee Eq. 17. Its applicability must be checked using the same objective at both endpoints, rather than losses from different mini-batches.

Because $C$ has rank at most $r$ per module, it is itself representable by LoRA factors. If the stronger, separately checkable condition $F_k(A_1, B_1; C) \le \mathcal{L}_k(C) + \varepsilon$ holds, the same argument gives

$$\|\Delta_1 - C\|_1 \le \frac{\mathcal{L}_k(C) - \ell_k + \varepsilon}{\lambda}, \qquad \mathcal{L}_k(\Delta_1) \le \mathcal{L}_k(C) + \varepsilon. \tag{19}$$

PACT does not reset Branches to $C$, so this stronger condition is not guaranteed by its initialization at each period. Neither bound implies that increasing $\lambda$ monotonically improves diagnostic performance, since the optimization endpoints and achieved error also change.

**Factor-space stationarity.** For completeness, let $G = \nabla_\Delta \mathcal{L}_k(BA)$ and choose an entrywise L1 subgradient $S \in \partial\|BA - C\|_1$. At a first-order stationary point of the factorized objective, a compatible choice of $S$ satisfies

$$(G + \lambda S)A^\mathsf{T} = 0, \qquad B^\mathsf{T}(G + \lambda S) = 0. \tag{20}$$

These are the factor-space conditions; they do not in general imply $G + \lambda S = 0$. Moreover, $S$ has $mn$ materialized coordinates per module, not $r(m + n)$ factor coordinates. For every nonzero residual entry, $|S_{ij}| = 1$ even if that residual is small. We therefore make no claim that closeness to the Anchor or trimming during aggregation makes this subgradient sparse.

### C.4 Rank Projection and Periodic Updates

If $\bar{\Delta}$ is the coordinate-wise merged matrix and $C^+ = \mathrm{SVD}_r(\bar{\Delta})$, truncated SVD gives the exact projection error

$$\|\bar{\Delta} - C^+\|_F^2 = \sum_{j>r} \sigma_j(\bar{\Delta})^2. \tag{21}$$

This characterizes matrix approximation, not diagnostic quality. If, additionally, $\mathcal{L}_k$ is $L_k$-Lipschitz in the materialized update over the segment between these matrices, then

$$|\mathcal{L}_k(C^+) - \mathcal{L}_k(\bar{\Delta})| \le L_k \left( \sum_{j>r} \sigma_j(\bar{\Delta})^2 \right)^{1/2}. \tag{22}$$

For multiple modules the squared projection errors are summed. The inequality relates rank-projection error to task-loss variation under the stated Lipschitz assumption.

Proposition 1 applies while an Anchor is fixed. When it changes from $C$ to $C^+$, the next period's initial residual satisfies $\|\Delta_1 - C^+\|_1 \le \|\Delta_1 - C\|_1 + \|C^+ - C\|_1$. Thus Anchor movement and projection error also matter; within-period regularization alone does not prove global stability. The interval $T^*$ controls how often the reference is updated and is an empirical design choice. The analysis provides no optimal interval and no guarantee that a consensus update reduces every Branch's task loss.

## D PACT Implementation Details

**Algorithm 1** PACT: Periodic Anchor Consensus Training

**Require:** Base model $\mathcal{M}$, datasets $\{D_k\}_{k=1}^K$, $\lambda$, $T^*$, TIES density $\rho$

```
Initialize Branches {Δ^(k)}_{k=1}^K and Anchor Δ^anc
for each training phase do
    Train each Branch on D_k with L_task^(k) + λ‖Δ^(k) − Δ^anc‖_1
    if phase reaches aggregation interval T* then
        Δ̂^(k) ← Trim(Δ^(k), ρ) for each k
        γ_i ← sign(Σ_k Δ̂_i^(k)) for each i
        Apply sign-consensus merge and rank-r projection to update Δ^anc
        Update the Anchor reference; Branches continue from their local parameters
    end if
end for
return M + Δ^anc
```

**Materialized LoRA deltas.** For each LoRA module, we compute the effective update as the materialized product $\Delta = BA$. The Anchor regularization in Eq. 1 is applied to these materialized deltas, and gradients are back-propagated to the low-rank factors $B$ and $A$.

**Rank projection.** Sign-consensus aggregation is performed on materialized Branch deltas. Because the coordinate-wise aggregated matrix is generally full-rank, the result is projected back to the original LoRA rank $r$ through truncated SVD before it is stored as the updated Anchor.

**Memory-efficient regularization.** Computing the L1 Anchor penalty over all LoRA modules in one graph can increase peak memory. We therefore compute and back-propagate the regularization term module by module, freeing each intermediate graph immediately after its backward pass. This preserves the same objective while reducing peak memory in implementation.

**Computational overhead.** PACT adds training work from four Branch adapters, Anchor regularization, and aggregation every $T^*$ epochs. Aggregation uses materialized LoRA delta matrices and rank projection. At inference, PACT deploys only one Anchor adapter and requires no additional Branch routing.

## E Statistical Significance Tests

We conduct paired bootstrap significance tests over EMR cases rather than over individual dialogue episodes. For each model and metric, we first average the four independent episodes belonging to the same held-out EMR, yielding 438 paired case-level scores. We then draw 20,000 bootstrap samples over EMR cases and compute two-sided $p$-values from the bootstrap distribution of paired differences. Table 13 reports the main comparisons; Delta denotes PACT minus the comparison method.

| Comparison | Metric | Delta | 95% CI | $p$-value |
|---|---|---|---|---|
| PACT–Mixed-SFT | Strict DA | +3.82 | [1.31, 6.28] | 0.003 |
| PACT–Mixed-SFT | Soft Score | +1.28 | [-0.65, 3.21] | 0.198 |
| PACT–Mixed-SFT | Exam F1 | +1.50 | [-0.25, 3.26] | 0.093 |
| PACT–Sequential | Strict DA | +5.14 | [2.71, 7.56] | $< 0.001$ |
| PACT–Sequential | Exam F1 | +3.47 | [1.68, 5.27] | 0.001 |
| PACT–GPT-4.1 | Strict DA | +18.61 | [15.24, 22.03] | $< 0.0001$ |
| PACT–GPT-4.1 | Soft Score | +11.30 | [8.68, 13.94] | $< 0.0001$ |
| PACT–GPT-4.1 | Exam F1 | +44.06 | [41.26, 46.77] | $< 0.0001$ |
| PACT–GPT-4.1 | Symp Cov | -0.64 | [-1.31, 0.03] | 0.061 |

Table 13: Paired bootstrap significance tests over EMR cases. Delta denotes PACT minus the comparison method.

## F Post-hoc Merge Details

Table 14 reports the detailed post-hoc merging results used to summarize the independent-branch baseline in Table 5. Different merge operators favor different metrics, indicating that post-hoc merging is sensitive to the aggregation rule.

| Method | Strict DA | Exam F1 | Symp Cov |
|---|---|---|---|
| Indep. + Linear | 52.91 | 56.27 | 93.92 |
| Indep. + TIES | 40.41 | 59.03 | 89.81 |
| Indep. + DARE-TIES | 52.45 | 57.16 | 93.66 |

Table 14: Post-hoc merge details for independently trained Qwen3-8B Branches.

# G Analysis and Additional Experimental Results

| Base Model | Config | Size | Strict DA | Relaxed DA | Soft Score | Avg Rnd | Symp Cov |
|---|---|---|---|---|---|---|---|
| Qwen3.5-9B | Base | 9B | 35.90 | 71.06 | 59.24 | 5.78 | 78.15 |
| | Mixed-SFT | 9B | 48.29 | 83.33 | — | 4.24 | 93.52 |
| | PACT | 9B | 49.77 | 82.23 | 71.73 | 2.06 | 79.52 |
| Qwen3-8B | Base | 8B | 32.53 | 73.97 | 59.03 | 4.70 | 77.11 |
| | Mixed-SFT | 8B | 51.48 | 86.02 | 74.88 | 4.17 | 93.72 |
| | PACT | 8B | 55.31 | 85.79 | 76.16 | 4.02 | 93.39 |
| Llama3.1-8B | Base | 8B | 16.89 | 70.03 | 50.94 | 3.77 | 75.04 |
| | Mixed-SFT | 8B | 51.37 | 86.74 | — | 2.56 | 74.79 |
| | PACT | 8B | 53.03 | 84.19 | 74.54 | 4.27 | 93.14 |
| Ministral-3-8B | Base | 8B | 20.94 | 52.34 | 41.83 | 2.71 | 43.28 |
| | Mixed-SFT | 8B | 36.42 | 78.53 | 64.35 | 3.69 | 88.54 |
| | PACT | 8B | 50.29 | 78.39 | 69.51 | 2.00 | 75.38 |

(a) Diagnosis and process metrics.

| Base Model | Config | Exam F1 | Exam P | Exam R |
|---|---|---|---|---|
| Qwen3.5-9B | Base | 15.00 | 16.79 | 16.70 |
| | Mixed-SFT | 62.45 | 69.62 | 61.97 |
| | PACT | 63.11 | 64.58 | 68.53 |
| Qwen3-8B | Base | 21.53 | 23.86 | 23.95 |
| | Mixed-SFT | 59.83 | 60.42 | 63.17 |
| | PACT | 61.33 | 57.50 | 74.27 |
| Llama3.1-8B | Base | 16.68 | 18.40 | 17.46 |
| | Mixed-SFT | 45.87 | 49.64 | 47.48 |
| | PACT | 60.92 | 66.82 | 61.15 |
| Ministral-3-8B | Base | 5.08 | 6.43 | 5.16 |
| | Mixed-SFT | 28.61 | 32.90 | 30.93 |
| | PACT | 45.71 | 49.03 | 48.33 |

(b) Examination metrics.

Table 15: Full cross-architecture generalization results. Each model shows zero-shot Base, Mixed-SFT on DPS data, and PACT. — denotes an unreported value.

## G.1 Pilot Drift–Perplexity Analysis

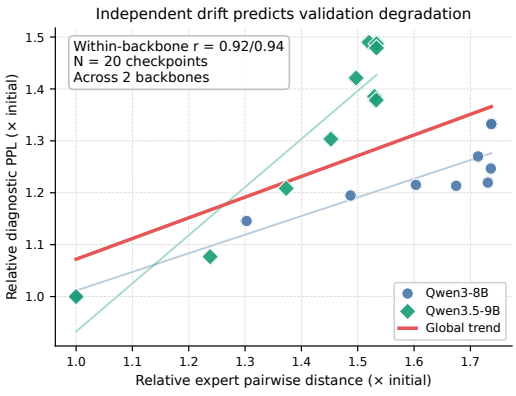


Figure 3: Pilot study on independent fine-tuning. Distances and PPL are normalized by each backbone's first checkpoint.

**Interpretation.** Across Qwen3-8B and Qwen3.5-9B, increasing Branch pairwise distance closely tracks degradation in diagnostic validation perplexity. The within-backbone correlations ($r$=0.92/0.94 over 20 checkpoints) support the use of parameter drift as a direct diagnostic of merge incompatibility rather than a purely geometric proxy.

### G.2 Hyperparameter Sensitivity

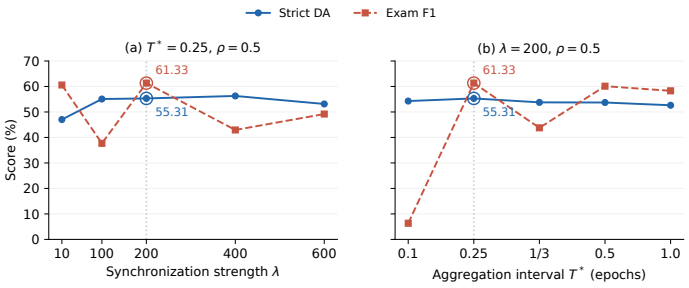


Figure 4: Sensitivity to $\lambda$ and $T^*$. Circled points indicate the default setting (Strict DA 55.31; Exam F1 61.33).

| $\rho$ | DA | F1 | Cov |
|---|---|---|---|
| 0.25 | 53.94 | 59.72 | 92.88 |
| 0.50 | **55.31** | **61.33** | **93.39** |
| 0.75 | 54.62 | 60.41 | 93.18 |
| 1.00 | 53.78 | 58.96 | 92.74 |

Table 16: TIES density with $\lambda$=200 and $T^*$=0.25. DA/F1/Cov denote Strict DA, Exam F1, and Symptom Coverage.

Figure 4 reports $\lambda \in \{10, 100, 200, 400, 600\}$ and $T^* \in \{0.1, 0.25, 1/3, 0.5, 1.0\}$ epochs. The default configuration $(\lambda, T^*, \rho) = (200, 0.25, 0.5)$ has Strict DA 55.31 and Exam F1 61.33, consistent with the main-result tables. Among the non-default settings, $\lambda = 400$ gives the highest Strict DA in the $\lambda$ sweep, while the default $\lambda = 200$ gives the highest Exam F1. In the interval sweep, the default $T^* = 0.25$ gives the highest reported Strict DA and Exam F1. These curves show metric-dependent trade-offs across synchronization strengths and aggregation intervals.

Table 16 shows that moderate trimming works best. A small density removes useful Branch-specific signal, while no trimming retains low-magnitude noisy coordinates and weakens sign-consensus aggregation. We therefore use $\rho = 0.5$ as the default.

### G.3 Parameter Drift Analysis

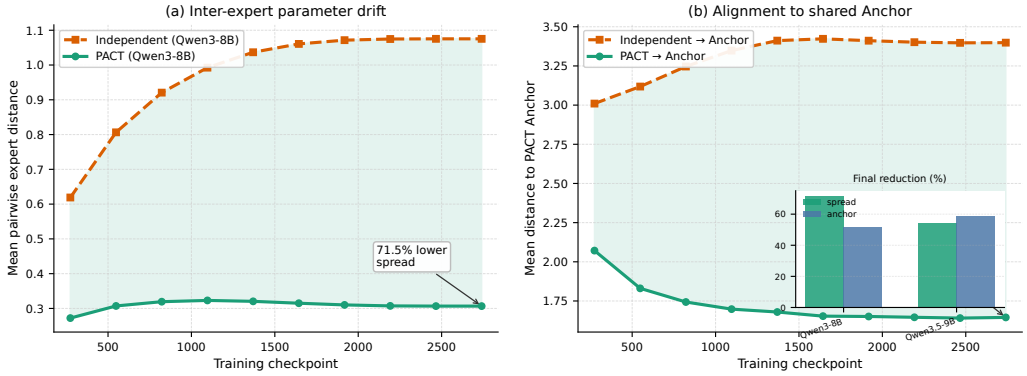


Figure 5: Parameter drift analysis across training checkpoints. (a) PACT substantially reduces the mean pairwise distance among paradigm-specific Branches compared with independent training. (b) PACT Branches remain closer to the shared Anchor, indicating that periodic aggregation and L1 Anchor Regularization align local paradigm updates to a common consensus. The inset reports final-checkpoint reductions on Qwen3-8B and Qwen3.5-9B.

We measure Branch drift directly in LoRA parameter space rather than relying only on qualitative PCA projections. Under independent training, the mean pairwise Branch distance grows from 0.619 to 1.075 on Qwen3-8B, a 73.7% increase. In contrast, PACT keeps the corresponding spread nearly flat, ending at 0.307. At the final checkpoint, PACT reduces inter-Branch spread by 71.5% and Branch–Anchor distance by 51.6% compared with independent training. The same trend appears on Qwen3.5-9B, where PACT reduces final Branch spread and Anchor distance by 54.0% and 59.1%, respectively.

### G.4 Sign Conflict Analysis

To measure sign disagreements during TIES merging, we simulate the Trim→Elect Sign pipeline offline at each checkpoint: we expand each Branch's LoRA delta $\Delta^{(k)} = B^{(k)}A^{(k)}$, apply magnitude trimming at density $\rho$=0.5, and measure the fraction of surviving positions where at least one Branch's sign disagrees with the magnitude-elected sign (**sign conflict ratio**). We also report the **mean flip rate**: the average fraction of each Branch's active positions opposed to the magnitude-elected sign.

Under independent training, sign conflict rises from 19.8% to 38.0% and saturates, confirming that Branches develop structurally incompatible parameter updates. The corresponding mean pairwise cosine between Branch task vectors drops from 0.61 to 0.18 (Figure 1). PACT reduces both metrics by over 99%: final sign conflict falls to 0.38% and mean flip rate to 0.11%, with Branch cosine maintained at ≈1.0. These measurements quantify sign agreement among retained coordinates; functional differences among Branches are examined in Appendix G.5.

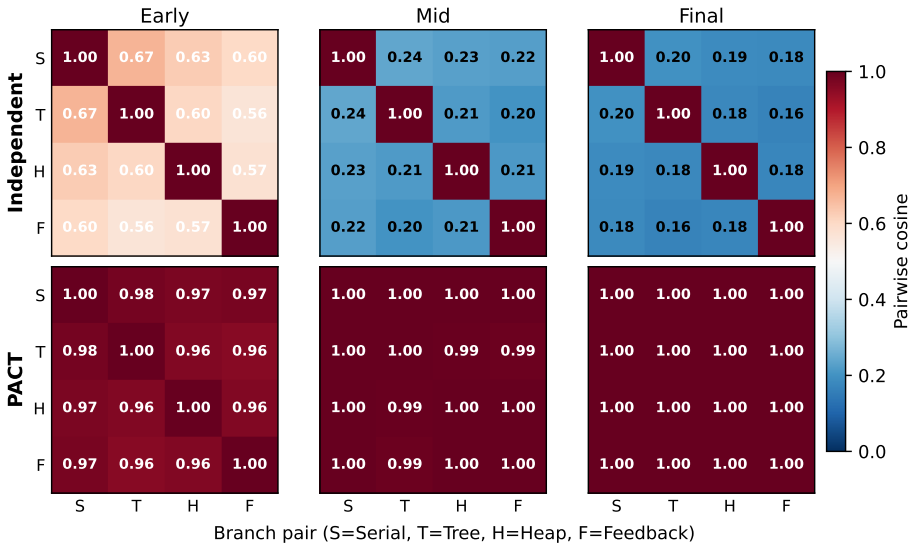


Figure 6: Full checkpoint-level Branch-alignment matrices for independent training and PACT. Independent Branches progressively diverge from early to final checkpoints, whereas PACT remains highly aligned throughout training.

### G.5 Branch Specialization and Complementarity

We deploy every trained PACT Branch independently on the full evaluation set and compare it with single-paradigm controls. Table 17 shows distinct performance profiles. Branch-Heap reaches the highest Branch-level Strict DA with the shortest dialogues, Branch-Tree reaches the highest Exam F1, and Branch-Serial reaches the highest symptom coverage. Differences among Branches are significant under paired bootstrap for Exam F1 and average rounds ($p$<0.05). The Anchor exceeds every Branch in Strict DA, showing that consensus captures complementary supervision rather than selecting one dominant strategy.

Figure 6 reports the checkpoint-level Branch cosine matrices. The final PACT values displayed as 1.00 are rounded to two decimal places. Table 17 complements this directional-alignment analysis with independently evaluated functional profiles.

Table 18 gives a qualitative example from a randomly selected 45-year-old female patient with recurrent epigastric pain. The Branches ask observably different first questions even though their parameter updates remain merge-compatible.

| Model | Training Data | Strict DA | Exam F1 | Symp Cov | Avg Rnd |
|---|---|---|---|---|---|
| Single-Serial | Serial only | 44.18 | 47.62 | 88.41 | 4.53 |
| Single-Tree | Tree only | 45.89 | 49.35 | 85.73 | 3.82 |
| Single-Heap | Heap only | 42.71 | 52.86 | 84.95 | 3.47 |
| Single-Feedback | Feedback only | 44.52 | 48.71 | 89.12 | 4.86 |
| Mixed-SFT | All four mixed | 51.48 | 59.83 | 93.72 | 4.17 |
| PACT Branch-S | Serial + Anchor reg. | 50.23 | 57.44 | **93.85** | 4.52 |
| PACT Branch-T | Tree + Anchor reg. | 49.89 | 61.22 | 91.45 | 4.08 |
| PACT Branch-H | Heap + Anchor reg. | 51.94 | 55.87 | 92.31 | **3.48** |
| PACT Branch-F | Feedback + Anchor reg. | 50.68 | 58.93 | 92.67 | 4.21 |
| **PACT Anchor** | Consensus of four | **55.31** | **61.33** | 93.39 | 4.02 |

Table 17: Functional specialization of single-paradigm controls, PACT Branches, and the final Anchor. Bold within the PACT rows marks the strongest value for each metric.

| Branch | First question | Pattern |
|---|---|---|
| Serial | "When did the pain start? Is it related to meals? Any nausea or vomiting?" | Stepwise history |
| Tree | "I need to consider gastritis, peptic ulcer, and pancreatitis. Does the pain radiate to the back? Any history of gallstones?" | Differential inquiry |
| Heap | "Is the pain severe enough to affect daily activities? Any weight loss or black stools?" | Urgency first |
| Feedback | "Let me organize: epigastric pain for two weeks. Before asking further, I should clarify its duration pattern and associated GI symptoms." | Self-reflection |
| Anchor | "How would you describe the pain—burning, cramping, or dull? Is it worse before or after eating?" | Balanced integration |

Table 18: Qualitative comparison of first-turn inquiry strategies for the same patient.

### G.6 Mixed-Training Conflict Diagnostics

To test the interference claim directly, we measure LoRA-gradient cosine on paradigm-specific mini-batches during Mixed-SFT. Mean pairwise cosine declines from 0.11 early in training to $-0.03$ late in training, with several pairs becoming negatively aligned. Validation dynamics provide a second check: the perplexity gap between Mixed-SFT and the PACT Anchor grows monotonically from 0.06 at step 500 to 0.29 at step 2500. Together, the diagnostics in Table 19 provide gradient- and validation-level evidence of interference, consistent with the drift–perplexity relation in Figure 3.

| Paradigm pair | Early | Late |
|---|---|---|
| Serial–Tree | 0.12 | $-0.04$ |
| Serial–Heap | 0.08 | $-0.11$ |
| Tree–Feedback | 0.14 | 0.02 |
| Mean pairwise | 0.11 | $-0.03$ |

(a) Gradient cosine.

| Step | Mixed PPL | PACT PPL | Gap |
|---|---|---|---|
| 500 | 4.21 | 4.15 | 0.06 |
| 1000 | 3.87 | 3.72 | 0.15 |
| 1500 | 3.74 | 3.51 | 0.23 |
| 2000 | 3.69 | 3.42 | 0.27 |
| 2500 | 3.67 | 3.38 | 0.29 |

(b) Validation perplexity.

Table 19: Mixed-training conflict diagnostics: paradigm-gradient alignment and validation dynamics.

### G.7 Judge, Simulator, Synthesis, and Human Robustness

**Automated robustness.** We re-evaluate four representative systems on the same 100-case subset with DeepSeek-V3 as an alternative outcome judge; all ranks are preserved and system-level Spearman correlation is 1.00 (Table 7a). With DeepSeek-R1 as the alternative patient simulator, PACT achieves 53.75% Strict DA and 59.83% Exam F1 on the 100-case subset (Table 7b). We also repeat DPS for approximately 200 EMRs with DeepSeek-R1 in the Doctor, Patient, and Supervisor roles. The

alternative synthesis produces 87.5% diagnosis alignment, 78.2% examination alignment, and 3.4% explicit label leakage (Table 7c).

**Clinician evaluation protocol.** Three board-certified physicians independently evaluate outputs from PACT, Mixed-SFT, GPT-4.1, and Doctor-R1 on 50 stratified cases balanced across difficulty levels. System identities are hidden, output order is randomized independently for each annotator, and every output is rated on three 1–5 Likert axes. The aggregate scores appear in Table 7d. Inter-annotator agreement is Krippendorff's $\alpha$=0.74. Human ratings and automated evaluation agree at Cohen's $\kappa$=0.71, with system-level Spearman $\rho$=1.00. A paired case-level test on Overall Clinical Competence gives $t$=3.42 and $p$=0.001 for PACT versus Mixed-SFT.

## H Limitations and Qualitative Failure Modes

**Evaluation scope.** Our evaluation combines synthetic supervision, LLM-based simulation and assessment, and de-identified EMRs from an Internal Medicine department. Cross-model and clinician evaluations are conducted on selected subsets, including 50 clinician-reviewed cases. Future studies could extend these evaluations to larger cohorts, additional institutions and specialties, and native English consultations. Prospective studies would further clarify the clinical utility of the proposed framework.

**Supervision design.** DPS uses diagnosis-aware Supervisor feedback to improve question selection, while privileged case information remains excluded from the Doctor context. Our audit identifies explicit diagnostic-label disclosure in only 2.1

**Qualitative analysis.** Dynamic evaluation identifies several opportunities for further improvement, including collecting more discriminative evidence before diagnosis, considering relevant examinations more systematically, and producing more focused differential diagnoses. Larger expert studies and multi-institutional prospective evaluations could provide a more comprehensive characterization of these behaviours and inform future safety assessment.