

FedCoT: 用于大型语言模型的通信高效联邦推理增强

Chuan Li^{1*}, Qianyi Zhao^{1*}, Fengran Mo², Cen Chen^{1†}

¹East China Normal University, China

²University of Montreal, Canada

51275903068@stu.ecnu.edu.cn, 51255903037@stu.ecnu.edu.cn,

fengran.mo@umontreal.ca, cench@ecnu.edu.cn

Abstract

在联邦学习环境中有效增强大型语言模型 (LLMs) 的推理能力仍然具有挑战性, 尤其是在严格的计算、通信和隐私约束下平衡性能收益时。这一挑战在医疗保健领域尤为严重, 因为涉及临床、运营和面向患者的上下文的决策不仅要求准确的输出, 还需要可解释、可追溯的推理, 以确保安全性、责任和遵从法规。传统的 LLM 联邦微调方法未能解决这一需求: 它们主要优化答案的正确性, 而忽略了推理质量, 使得链式思考 (CoT) 能力依赖于模型的固有预训练能力。此外, 现有的提高推理质量的方法通常依赖于从集中模型进行隐私侵犯的知识蒸馏。此外, 传统的 LLM 联邦微调中的通信开销仍然很大。我们通过提出 FedCoT 来解决这一差距, 这是一个专门设计用于在联邦环境中增强推理的新框架。FedCoT 利用了一种轻量级的链式思维增强机制: 本地模型生成多条推理路径, 并且通过一个精简的判别器动态选择最有前途的一条。这种方法提高了推理的准确性和鲁棒性, 同时提供了有价值的可解释性, 这在医疗应用中尤为关键。为了高效管理客户端的异质性, 我们采用了一种改进的聚合方法, 该方法构建于先进的 LoRA 模块堆叠之上, 结合了客户端分类器意识, 从而实现了跨多样化客户端的无噪声聚合。针对医疗推理任务的综合实验表明, FedCoT 在严格的资源限制下显著提升了客户端推理性能, 同时完全保留了数据隐私。我们的工作建立了一种原则性的途径, 用于可解释且资源高效的联邦推理增强。

引言

大型语言模型 (LLMs) 的发展在复杂推理任务中实现了先进的性能 (Touvron et al. 2023; Bai et al. 2023; Guo et al. 2025a; Team et al. 2025; Chen et al. 2024a), 这基于思维链 (Wei et al. 2022) 提高了效果和可解释性。令人期待的性能归功于强化学习 (RL) 算法 (Christiano et al. 2017; Schulman et al. 2017; Shao et al. 2024; Zhou et al. 2025)。然而, 推理模型的 RL 训练范式严重依赖于计算资源 (Tian, Shi, and Li 2023; Havrilla et al. 2024), 这使得它们在分布式边缘环境中不切实际, 特别是在隐私约束下, 例如, 在医疗场景中训练数据不能直接在不同节点/机构间共享 (Chen et al. 2023)。

无训练技术 (Wang et al. 2022; Xie et al. 2023; Wu et al. 2024) 通过提示工程或测试时缩放可以缓解训练阶段数据和模型分布的问题。尽管它们易于部署, 但性

能提升相当有限, 无法充分利用分布式设备网络的协作潜力。直观地看, 联邦学习 (FL) (McMahan et al. 2017) 可以作为在隐私保护保证与模型性能之间实现更好权衡的替代方案。然而, 现有基于 FL 的 LLM 训练范式 (Wu et al. 2025; Wei et al. 2025; Zhang et al. 2024a) 主要依赖于联邦有监督微调或简单地结合参数高效技术, 这仍然遇到高通信开销, 因而导致次优的性能提升。

在医疗领域中, 链式思考 (CoT) 是不可或缺的。例如, 客户的决策不仅需要准确, 还需要可靠且具有可追溯的理由。此外, 这些隐私敏感场景需要更严格的数据和模型使用以防止敏感信息泄露。然而, 现有研究 (Magister et al. 2022; Li et al. 2022; Wang et al. 2023; Chen et al. 2024c) 通常通过从专有模型中提取知识或在不同来源之间直接共享来获得数据和理由, 这可能会暴露敏感信息并违反数据隐私原则。

为了在隐私保护环境下进行推理训练而无需在分布式节点中共享整个模型, 我们提出了 FedCoT, 一个基于联邦学习的框架, 以在不泄露数据的情况下, 通过链式思维提示技术增强模型的推理能力。我们的 FedCoT 的核心是一个动态链式思维区分机制, 以实现跨客户端的推理能力提升。具体而言, 我们部署了一种轻量级鉴别器, 以实时评估推理路径候选, 这使得优化的推理轨迹可以被区分。基于 FLoRA 算法的模块化 LoRA 堆栈联邦微调的主要机制 (Wang et al. 2024b), 我们将这种方法应用于轻量级 BERT (Devlin et al. 2019) 模型。关键是我们结合了一种特定任务的预测器 (基于 BERT 的分类器), 专注于推理路径鉴别。为了聚合这些特定任务的分类器, 我们采用了加权聚合方案, 确保整个联邦具有稳健的辨别能力。我们的 FedCoT 目标同时防止隐私风险, 并在隐私保护场景中实现稳健的性能。在五个医疗领域数据集上的实验结果证明了我们的方法的有效性, 通过超过现有强基线。

我们的贡献总结如下:

- 据我们所知, 我们是第一个在联邦学习环境中利用 CoT 技术来增强大型语言模型推理能力的研究, 同时实现了隐私保护和低资源消耗。
- 我们提出了一种端到端的联邦推理增强框架, 集成了动态推理路径鉴别。FedCoT 将模块化堆叠扩展到基于 BERT 的鉴别器, 并采用加权聚合方案, 有效应对客户端异质性, 同时保持稳健的性能。
- 我们在多个医学问答基准上进行了全面的实验。结果表明, FedCoT 与强大的基线相比显著提升了推理

*These authors contributed equally.

†Corresponding author

性能和效率，稳健地验证了我们联邦推理增强框架的有效性。

相关工作

由 Wei et al. (2022) 提出的

大语言模型的推理增强 CoT 提示，被视为通过提供的指示增强大语言模型推理能力的一种有效机制，已经催生了众多不需要训练的变体 (Wang et al. 2022; Chen et al. 2024b; Li et al. 2025; Nair and Wang 2024; Wan et al. 2024; Guo et al. 2025b)。后续研究通过使用 CoT 生成的逻辑用于模型监督微调 (Kim et al. 2023; Magister et al. 2022; Li et al. 2022; Wang et al. 2023; Hsieh et al. 2023)，或使用强化学习进行训练 (Team et al. 2025; Shao et al. 2024)，将 CoT 与参数更新相结合。然而，这些方法假设数据集中访问，忽视了为联邦客户端生成逻辑所带来的隐私限制和计算负担 (McMahan et al. 2017)，因此在分布式环境下，没有为 CoT 能力增强提供专门的机制 (Zhang et al. 2024b; Dritsas and Trigka 2025; Wang et al. 2024a; Li et al. 2019)。为此，我们的方法利用分布式客户端的信息来增强 CoT 推理能力，同时严格确保隐私和低资源消耗。

大语言模型的联邦学习 FL 是在隐私保护环境下训练 LLM 的关键解决方案之一 (Wei et al. 2025; Wu et al. 2025; Chen et al. 2022; Wu, Chen, and Wang 2020; Tariq et al. 2023, 2024; Ye et al. 2023; Qian et al. 2024)。在训练阶段，梯度和数据聚合导致了通信效率的需求，通常通过参数高效的调优来实现。例如，低秩适应 (LoRA) 技术 (Zhang et al. 2024a) 和矩阵分解方法 (Wang et al. 2024b) 能够在异构环境下实现高效聚合。此外，知识蒸馏 (Fan et al. 2024) 和联邦 RL (Tian, Shi, and Li 2023) 提供了其他优化路径。然而，现有的 FL-LLM 研究既没有明确增强 CoT 推理能力，也没有缓解中央教师模型中提取推理的内在隐私风险 (Havrilla et al. 2024; Li et al. 2022)。此外，基于联邦 RL 的方法对于资源受限的客户端而言，会带来过高的计算和通信开销 (Qi et al. 2021; Krouka et al. 2021; Imteaj et al. 2022)。因此，以往的研究在开发重量轻的、隐私保护的框架以适应联邦 CoT 增强方面仍存在重要空白，而我们的研究提供了首次探索。

预备知识

思维链提示

Chain-of-Thought (CoT) 提示 (Wei et al. 2022) 通过在输入 x 和输出 y 之间生成中间推理路径 τ 来增强复杂推理，作为指导和解释。形式上，给定一个由 θ 参数化的模型和提示 I ，CoT 生成过程定义为：

$$[\tau, y] \sim p_{\theta}(x|I)$$

尽管 CoT 在一系列场景中有效，但如果不利用协同设备进行特定训练，仅在分布式环境中使用它是不能很好地运行的。

LoRA 联邦聚合

低秩调整 (LoRA) (Hu et al. 2022; Mao et al. 2024) 通过参数矩阵分解 $\Delta W = BA$ ($A \in \mathbb{R}^{r \times n}, B \in \mathbb{R}^{m \times r}$

) 实现了高效的微调。标准的联邦学习通过平均聚合更新客户端模型，如

$$A = \sum_i^N u_i A_i, \quad B = \sum_i^N u_i B_i \quad (1)$$

，其中 N 表示客户端的数量， u_i 表示由数据量比率导出的客户端权重。此聚合方案引入了跨客户端噪声项，如

$$\begin{aligned} \Delta W &= (u_0 B_0 + u_1 B_1)(u_0 A_0 + u_1 A_1) \\ &= u_0^2 B_0 A_0 + u_1^2 B_1 A_1 \\ &\quad + \underbrace{u_0 u_1 (B_0 A_1 + B_1 A_0)}_{\text{noise term}} \end{aligned} \quad (2)$$

。然而，由于噪声项随着本地客户端呈二次增长，全球更新将会偏离。此外，它将导致异构秩之间的维度不匹配，即 $r_1 \neq r_2$ ，导致参数更新失效。

方法论

我们提出了一种基于联邦学习的框架 FedCoT，以在隐私约束下增强大型语言模型 (LLMs) 的推理能力。FedCoT 的概述如图 ?? 所示，具有动态路径选择和参数高效聚合。由客户端 LLM 本地生成的候选推理路径，所产生的信号监督轻量级鉴别器的训练，然后其 LoRA 模块和分类器在服务器端聚合，构建出一个全局鉴别器。该联邦模型随后使客户端能够在推理过程中动态选择最佳推理路径，提供增强验证链 (CoT) 的隐私保护答案。

在联邦学习框架下，每个客户端应首先基于其关联的本地数据集 $\{x_j, y_j\}$ 的问题生成候选推理路径，这些路径随后作为后续判别器训练的基础。形式上， K 个候选推理路径通过对 LLM p_{θ} 的多样性采样生成，这可以类似地被视为强化学习场景中的演员模型，对应输入 x_j 为：

$$[\tau_{j,k}, \hat{y}_{j,k}] \sim p_{\theta}(x_j|I)$$

其中本地真实值 y_j 被分配为二元标签，如方程 3 所示，然后用于判别器训练的判别数据集形成如方程 4。

$$z_{j,k} = \mathbf{1}(\hat{y}_{j,k} = y_j), \quad k = 1, 2, \dots, K \quad (3)$$

$$\mathcal{D} = \{(\mathbf{h}_{j,k}, z_{j,k}) \mid \mathbf{h}_{j,k} = [x_j \parallel \tau_{j,k} \parallel \hat{y}_{j,k}]\} \quad (4)$$

整个本地候选生成过程使多样化推理路径的隐私保护探索成为可能。

我们将推理路径区分表述为一个由 Shi et al. (2024) 激发的二元分类任务，其中一个在 BERT 规模上的轻量级判别器可以有效评估候选的正确性。

在我们的 FedCoT 框架中，客户端从以下选项之一初始化其本地模型：(1) 服务器提供的全局模块（用于非初始轮次）或 (2) 本地的基础预训练模型（用于第一轮）。在每个全局通信轮中，客户端接收并使用最新聚合参数初始化模型，这些参数包括 LoRA 矩阵和分类器，它们涵盖了来自整个联邦的信息，同时防止模型完全偏离本地领域。

形式上，给定一个问题-推理对 $(x_j, \tau_{j,k})$ ，判别器 $d_{\theta} : \mathcal{X} \times \mathcal{T} \rightarrow [0, 1]$ 通过 sigmoid 激活函数输出一个标准评分，该评分被优化以最小化如下的二元交叉熵损失：

$$\mathcal{L} = -[z_{j,k} \log d_{\theta}(\mathbf{h}_{j,k}) + (1 - z_{j,k}) \log(1 - d_{\theta}(\mathbf{h}_{j,k}))] \quad (5)$$

，其中 $z_{j,k} \in \{0,1\}$ 表示验证的正确性， h 编码候选推理路径。客户端 LoRA 参数和分类器的联邦聚合然后将这些局部分布合成为在全球优化的决策边界，具有增强的泛化能力。

模块化全局聚合

我们采用并集成了 FLoRA (Wang et al. 2024b)，以在保护数据隐私的情况下实现 LoRA 矩阵的无噪声聚合。当聚合本地 LoRA 模块时，全局模型更新 ΔW 可以表达为

$$\begin{aligned}\Delta W &= \sum_{i=1}^N B_i A_i \\ &= (B_1 \oplus B_2 \oplus \dots \oplus B_N) \\ &\quad \cdot (A_1 \oplus A_2 \oplus \dots \oplus A_N)\end{aligned}\quad (6)$$

，其中“ $\oplus$ ”表示矩阵堆叠操作，即沿着行方向垂直堆叠 A_i ，沿着列方向水平堆叠 B_i 。根据块矩阵乘法原理，这两个全局生成矩阵 $B \cdot A$ 的乘积在数学上等价于各个本地更新的总和 $\sum_{i=1}^N B_i A_i$ 。

这种方法使全局聚合的鉴别器更可靠且更适应异质性，异质性来源于不同客户的能力（例如，较弱客户使用较小的 LoRA 等级，较强客户使用较大的等级）。我们也可以通过给简单任务分配较小等级而给复杂任务分配较大等级来有意创造异质性。不论其来源，叠加方法通过统一合并整合这些不同的 LoRA 矩阵，确保顺利的联邦学习。

此外，每个客户的分类器权重在每个全局轮中使用加权平均方法进行聚合，以整合下游任务中的信息作为

$$W^{cls} = \sum_i u_i W_i^{cls} \quad (7)$$

在推理阶段，每个客户端利用最终的全局判别器模型对多个候选推理路径进行评分，然后选择得分最高的路径作为最终输出，以实现动态推理，如下所示：

$$r(h_{j,k}) = \sigma(d_\theta(h_{j,k})) \quad (8)$$

$$\hat{y}_j = \arg \max_{k \in \{1, \dots, K\}} r(h_{j,k}) \quad (9)$$

该综合过程描述为附录中提供的一个算法，是我们 FedCoT 中联邦推理的总体过程。

实验

实验装置

我们在五个生物医学问答 (QA) 数据集上评估我们的方法，这些数据集遵循之前的研究作为隐私保护基准，包括 BioASQ、MedMCQA、MedQA、MMLU-MED 和 PubMedQA。统计数据在表中提供。这些数据集涵盖不同的任务类别，从医学考试问题到基于文献的问答，使我们能够全面评估模型在复杂推理任务中的表现。详细信息在附录中提供。

跨数据孤岛环境 这五个数据集分别被视为独立的客户端，并且在训练过程中严格保护每个客户端数据的隐私。这个跨孤岛设置反映了模型在不同任务上的推理能力，并验证了其在数据分布异质性下的鲁棒性，这与现实世界中的数据孤岛 (Huang et al. 2021, 2023; Tang and Wong 2021; Liu et al. 2022) 相符。

Datasets	Train	Test	Source
PubMedQA (2019)	450	500	Experts
BioASQ (2015)	494	124	Articles
MMLU (2020)	1299	163	Examination
MedMCQA (2022)	3000	4183	Examination
MedQA (2021)	10178	1273	Examination

Table 1: 实验中使用的医学问答数据集的规模和来源。

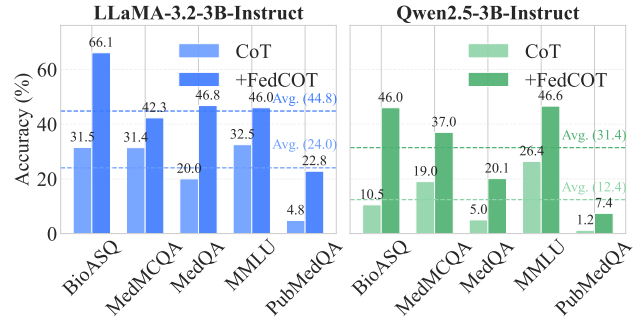


Figure 1: 通过在我们的 FedCoT 之上进行联邦推理微调，以提高 3B LLM 的性能。

提示模板 我们设计的 CoT 模板遵循一个标准化结构，其特征为简洁的指示以减少冗长，以及一个结构化的、条目化的要求格式。我们还在模板中加入了一次性 CoT 演示，完整的信息在附录中提供。

我们的实验使用不同的模型进行评估，包括 Qwen2.5-7B-Instruct (Bai et al. 2023)、LLaMA-3-8B-Instruct (Touvron et al. 2023) 作为主要评估的核心 LLM，以及作为判别模型的 Longformer-base-4096 (Beltagy, Peters, and Cohan 2020)，遵循 Shi et al. (2024)。我们在联邦和非联邦场景下与无训练和基于训练的基线进行比较，以如下方式评估我们的 FedCoT：(1) Self-Consistency (Wang et al. 2022)，一种利用多样化采样和多数投票的无训练方法；(2) Local-SFT，每个客户端在使用其本地训练数据在 actor 模型上执行 SFT；(3) Fed-SFT，客户通过直接平均使用其本地数据集在 actor 模型上协作进行联邦监督微调；(4) FedIT (Zhang et al. 2024a)，设置与 Fed-SFT 相同，但使用加权平均。采用准确性作为主要评估指标，以保持与先前研究的一致性 (Chen et al. 2024a)。所有评估都在 CoT 提示下测量准确性，这不仅量化性能，还通过固有的分步推理符合对可解释性和响应安全性的现实医疗需求。

超参数设置 我们为每个查询样本生成 8 个候选响应，最长为 512 个标记。BioASQ、MedMCQA、MedQA、MMLU 和 PubMedQA 数据集集中的每个客户端模型的 LoRA 等级分别设定为 4、32、32、16 和 4。在联邦 LLMs 的监督微调期间，模型训练使用了等级为 32 的统一 LoRA。全局轮数设定为 2，在 LLMs 的 SFT 中作为基线，本地训练周期设定为 1，批量大小为 2。在我们的判别器训练中，全局轮数设定为 3，本地训练周期设定为 1，批量大小为 16。

主要结果

总体结果如表 2 所示。我们可以观察到，FedCoT 明显优于基于五个数据集的两个骨干大型语言模型的其

Method	BioASQ		MedMCQA		MedQA		MMLU		PubMedQA		Avg.	
	Acc. (%)	Δ (%)	Acc. (%)	Δ (%)	Acc. (%)	Δ (%)	Acc. (%)	Δ (%)	Acc. (%)	Δ (%)	Acc. (%)	Δ (%)
LLaMA-3-8B-Instruct	37.90	—	29.80	—	27.20	—	38.70	—	9.20	—	28.56	—
+Self-Consistency	40.30	+2.40	31.50	+1.70	24.70	-2.50	41.10	+2.40	2.80	-6.40	28.08	-0.48
+Local-SFT	52.42	+14.52	39.30	+9.50	54.60	+27.40	55.21	+16.51	10.20	+1.00	42.35	+13.79
+Fed-SFT	51.61	+13.71	44.23	+14.43	45.48	+18.28	65.03	+26.33	10.60	+1.40	43.39	+14.83
+FedIT	42.74	+4.84	47.29	+17.49	53.73	+26.53	71.17	+32.47	13.20	+4.00	<u>45.63</u>	+17.07
+FedCoT (Ours)	65.30	+27.40	45.20	+15.40	56.10	+28.90	54.00	+15.30	41.00	+31.80	52.32	+23.76
Qwen2.5-7B-Instruct	73.40	—	43.70	—	29.50	—	50.30	—	38.80	—	47.14	—
+Self-Consistency	86.30	+12.90	47.10	+3.40	28.00	-1.50	57.10	+6.80	39.80	+1.00	51.66	+4.52
+Local-SFT	75.81	+2.41	35.02	-8.68	46.11	+16.61	49.08	-1.22	43.60	+4.80	49.92	+2.78
+Fed-SFT	81.45	+8.05	44.56	+0.86	37.86	+8.36	55.83	+5.53	41.20	+2.40	52.18	+5.04
+FedIT	82.26	+8.86	48.48	+4.78	44.30	+14.80	68.71	+18.41	47.20	+8.40	<u>58.19</u>	+11.05
+FedCoT (Ours)	96.80	+23.40	50.00	+6.30	52.50	+23.00	66.30	+16.00	64.80	+26.00	66.08	+18.94

Table 2: 在不同设置下，两种主干大语言模型上的五个隐私保护医疗数据集的不同方法表现。最佳结果以粗体显示，第二高的结果用 underline 表示。

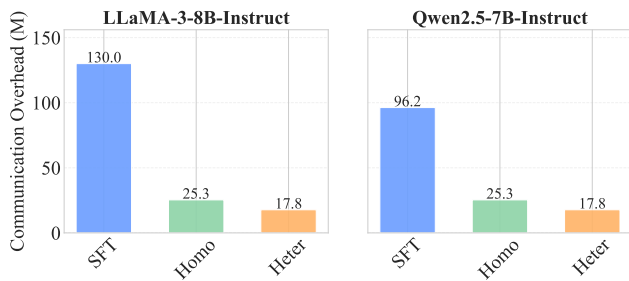


Figure 2: 对联邦 SFT 中的通信效率进行分析以及我们 FedCoT 的分析。“SFT”代表 Fed-SFT/FedIT，“Homo”代表 lora rank 为 32 的 FedCoT，“Heter”代表 lora rank 分别为 4, 32, 32, 16, 4 的 FedCoT。

他方法，这表明了我们 FedCoT 的卓越性能。具体而言，与直接使用 CoT 提示查询 LLaMA-3-8B-Instruct 和 Qwen2.5-7B-Instruct 相比，FedCoT 平均分别带来了 23.76 % 和 18.94 % 的绝对提升，这也超越了第二好的传统联邦微调方法，即 FedIT，在两个骨干大型语言模型上的提升分别超过 6 % 和 7 %。这些结果展示了我们 FedCoT 在联邦学习设置下的潜力和广泛适用性。我们还发现，免训练的方法 Self-Consistency 获得了轻微的改进，而 Local-SFT 方法无法达到与联邦方法相媲美的性能。这是因为在联邦学习场景下训练模型可以在约束条件下进一步受益于充分的数据使用。此外，我们的 FedCoT 在各种数据集上的改进更为稳定，表明其比其他方法更具鲁棒性。

效率比较 效率比较如图 2 所示，比较了联邦 SFT 方法（Fed-SFT/FedIT）与我们的 FedCoT 之间的通信效率。在联邦学习过程中，联邦 SFT 方法对 LLM actor LLaMA-3-8B-Instruct 和 Qwen2.5-7B-Instruct 进行 LoRA 微调。这里，130M 和 96M 的值表示在一次全局训练轮次中需要在联邦系统的所有客户端之间传输的参数总数。然而，即便采用 LoRA 微调，这样的参数数量仍然会给低资源客户端和整个联邦学习系统带来巨大的计算和通信开销。相比之下，我们的 FedCoT 通过微调一个轻量级模型极大地降低了训练和通信开销。因此，在联邦学习过程中 FedCoT 的参数数量仅为 25.3M

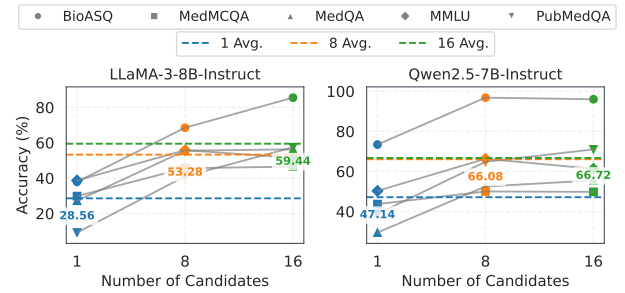


Figure 3: 在不同候选数量的 FedCoT 上的性能提升。不同形状代表不同的数据集。

和 17.8M，比起所比较的现有 SFT 方法更为高效，显示了其在低资源环境中的效率。

我们进一步研究了较小尺寸的 LLMs 的性能，结果展示在图 1 中。我们可以观察到，我们的 FedCoT 在较小模型上显示了适度但一致的提升。具体来说，FedCoT 在经过我们联邦学习框架的特定训练后显著优于 CoT 方法，并在 LLaMA-3.2-3B-Instruct 上平均提高了 20.8 %，在 Qwen2.5-3B-Instruct 上提高了 19.00 %。这些结果强调了我们方法在不同模型尺寸上的强大泛化性和适应性。

在候选生成中，候选的采样数量对模型性能具有敏感性。我们的实验分析如图 3 所示，表明将候选样本从 8 增加到 16 可以持续提高所有模型的性能。这种改善在 LLaMa-3-8B-Instruct 上尤为显著，其平均准确率从 52.32 % 增加到 59.44 %。值得注意的是，在 BioASQ 数据集上的准确率显著上升，从 65.30 % 增加到 85.50 %。

虽然 Qwen2.5-7B-Instruct 也表现出改进的增益（从 66.08 % 到 66.72 %），但这种边际改善表明模型对候选数量的敏感性是依赖于模型的。特别是，对于所有基础模型来说，8 个和 16 个候选之间的性能差距显著小于 1 个和 8 个候选之间的差距，这表明 8 个候选足以展示方法的有效性。

Method	BioASQ	MedMCQA	MedQA	MMLU	PubMedQA	Avg.(%)
CoT	37.90	29.80	27.20	38.70	9.20	28.56
FedCoT (r=8,8,8,8,8)	64.50	45.30	55.20	53.40	41.00	51.88
FedCoT (r=4,8,16,8,4)	64.50	45.40	55.40	52.10	41.00	51.68
FedCoT (r=4,32,32,16,4)	65.30	45.20	56.10	54.00	41.00	52.32

Table 3: 在不同 LoRA 配置下 FedCoT 的不同表现。“r” 代表不同客户端的 LoRA 秩，对应顺序为数据集 BioASQ、MedMCQA、MedQA、MMLU 和 PubMedQA 的客户端。最佳结果以粗体显示。

Max Length	Actor Model	Avg.(%)
512	LLaMA-3-8B-Instruct	5.93
	Qwen2.5-7B-Instruct	26.28
1024	LLaMA-3-8B-Instruct	0.02
	Qwen2.5-7B-Instruct	0.04

Table 4: 在最大生成长度下，所有数据集的平均截断率。

不同 LoRA 设置的分析

为了验证在异构环境下的鲁棒性，我们对 LoRA 秩配置进行消融研究，并在表格 3 中显示结果。我们可以看到我们的 FedCoT 在不同配置下始终表现出色，通过模块化叠加展示了对客户端异构性的内在适应性。关键的是，我们观察到战略性秩分配对于异构优化至关重要。具体而言，统一秩设置（所有客户端 $r=8$ ）达到 51.88 % 的平均准确率，而简单的数据比例划分 ($r=4,8,16,8,4$) 导致轻微下降 (51.68 %)。这些结果表明，任务复杂性和计算能力超越数据量，需要明确的调整。优化分配 ($r=4,32,32,16,4$) 将性能提升至 52.32 %，比同质基线高出 0.44 %。总体结果展示了我们的框架能够根据多维限制动态调整资源分配，同时保持竞争力的结果。

为了解决在 512 个标记限制下的截断效果可能带来的问题（例如，在生成完整答案之前过早终止），我们将最大生成长度扩展到了 1024 个标记。此措施有效地消除了截断问题，如表格 4 中所显示的几乎为零的截断率所证实。

表 5 中的性能对比显示，在大多数配置中，经过长度扩展后的一致准确性提升。我们的 FedCoT 方法在 LLaMA-3-8B-Instruct 和 Qwen2.5-7B-Instruct 上分别取得了 +3.04 % 和 +1.00 % 的提升，而联邦 IT 在 LLaMA-3-8B-Instruct 上显示了 +0.43 % 的改善。至于联邦 IT 在 Qwen2.5-7B-Instruct 上观察到的轻微下降 (-0.27 %)，可能是由于推理链长度延长后噪声增加所致。

至关重要的是，FedCoT 在控制截断效应时表现出强大的效果，对于 LLaMA-3-8B-Instruct 和 Qwen2.5-7B-Instruct，在 1024 的标记长度下，相较于 FedIT 基线分别实现了 +9.30 % 和 +9.16 % 的显著改进。这些增益超出了在 512 标记限制下观察到的增益，验证了我们的方法在生成长度参数上的有效性和稳健性。

细粒度过程导向歧视的讨论

虽然我们的主要框架依赖于基于结果的路径区分标签，但我们研究了过程导向的评估是否可以提供更细粒度的信号。受到强化学习中过程奖励模型 (Lightman et al.

2023) 的启发，我们设计了一种逐步的自我评估机制，其中每个客户端模型对其中间推理步骤分配置信度分数。

令人惊讶的是，如表 6 所示，模型表现出强烈的积极偏向，自我评估的步骤准确率在所有数据集中范围为 81.63 % 到 99.98 %。这种过度自信即使在最终答案准确率极低的情况下仍然存在（例如，在 PubMedQA 上仅 9.20 %），表明自我评估无法区分正确与错误的推理路径。这可能是由于模型缺乏对中间步骤的可靠内部不确定性估计，自我评估任务由于在相同数据分布上训练而继承了模型现有的偏差。

结论

在本文中，我们旨在解决在隐私保护约束和低资源消耗下优化大语言模型 (LLM) 的推理性能的问题。我们提出了 FedCoT，这是一个为联邦学习场景量身定制的推理增强框架，解决了传统联邦学习下 LLM 推理的三个核心挑战，包括推理能力不足、过高的通信开销以及严格的隐私要求。我们的 FedCoT 在推理和训练阶段使用了一个两阶段的推理增强策略，其中在推理期间使用轻量级判别模型选择最佳候选路径以提升推理能力，而在训练期间则采用 LoRA 堆叠和分类器聚合机制。实验表明，FedCoT 在五个医学数据集上超越了现有方法，提供了一个在隐私和资源限制下的高效且有效的 LLM 推理解决方案。

References

- Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; Huang, F.; et al. 2023. Qwen technical report. arXiv preprint arXiv:2309.16609.
- Beltagy, I.; Peters, M. E.; and Cohan, A. 2020. Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.
- Chen, C.; Ye, T.; Wang, L.; and Gao, M. 2022. Learning to generalize in heterogeneous federated networks. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 159–168.
- Chen, J.; Cai, Z.; Ji, K.; Wang, X.; Liu, W.; Wang, R.; Hou, J.; and Wang, B. 2024a. Huatuogpt-o1, towards medical complex reasoning with llms. arXiv preprint arXiv:2412.18925.
- Chen, S.; Mo, F.; Wang, Y.; Chen, C.; Nie, J.-Y.; Wang, C.; and Cui, J. 2023. A Customized Text Sanitization Mechanism with Differential Privacy. In Findings of the Association for Computational Linguistics: ACL 2023, 5747–5758.

Model	Method	BioASQ	MedMCQA	MedQA	MMLU	PubMedQA	Avg.(%)
LLaMA-3-8B-Instruct	FedIT-512	42.74	47.29	53.73	71.17	13.20	45.63
	FedCoT-512	65.30	45.20	56.10	54.00	41.00	<u>52.32</u>
	FedIT-1024	45.16	46.45	54.36	70.55	13.80	46.06
	FedCoT-1024	78.20	44.40	53.70	58.30	42.20	55.36
Qwen2.5-7B-Instruct	FedIT-512	82.26	48.48	44.30	68.71	47.20	58.19
	FedCoT-512	96.80	50.00	52.50	66.30	64.80	<u>66.08</u>
	FedIT-1024	82.26	48.55	43.28	68.10	47.40	57.92
	FedCoT-1024	97.60	51.60	56.50	59.50	70.20	67.08

Table 5: 在不同最大生成标记长度下的性能比较 (%)。最好的结果用粗体表示，第二高的结果用 underline 标出。

Dataset	Positive	Negative	Ratio(%)
BioASQ	5,740	145	97.54
MedMcQA	47,290	5,408	89.74
MedQA	118,719	26	99.98
MMLU	36,676	979	97.40
PubMedQA	3,288	740	81.63

Table 6: 医学问答基准中的逐步自我评估表现。正面：模型判断正确的推理步骤数量；负面：判断错误的步骤数量；比例：正确自我评估的比例。

Chen, W.; Wang, W.; Chu, Z.; Ren, K.; Zheng, Z.; and Lu, Z. 2024b. Self-Para-Consistency: Improving Reasoning Tasks at Low Cost for Large Language Models. In 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024), 14162–14167. Association for Computational Linguistics.

Chen, X.; Huang, H.; Gao, Y.; Wang, Y.; Zhao, J.; and Ding, K. 2024c. Learning to maximize mutual information for chain-of-thought distillation. arXiv preprint arXiv:2403.03348.

Christiano, P. F.; Leike, J.; Brown, T.; Martic, M.; Legg, S.; and Amodei, D. 2017. Deep reinforcement learning from human preferences. Advances in neural information processing systems, 30.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 4171–4186.

Dritsas, E.; and Trigka, M. 2025. Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications. Journal of Sensor and Actuator Networks, 14(1): 9.

Fan, T.; Ma, G.; Kang, Y.; Gu, H.; Song, Y.; Fan, L.; Chen, K.; and Yang, Q. 2024. Fedmkt: Federated mutual knowledge transfer for large and small language models. arXiv preprint arXiv:2406.02224.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al.

2025a. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948.

Guo, Y.; Yang, Y.; Chen, Z.; Wang, P.; Liao, Y.; Zhang, Y.; Wang, Y.; and Wang, Y. 2025b. Dsvd: Dynamic self-verify decoding for faithful generation in large language models. arXiv preprint arXiv:2503.03149.

Havrilla, A.; Du, Y.; Raparthy, S. C.; Nalmpantis, C.; Dwivedi-Yu, J.; Zhuravinskyi, M.; Hambro, E.; Sukhbaatar, S.; and Raileanu, R. 2024. Teaching large language models to reason with reinforcement learning. arXiv preprint arXiv:2403.04642.

Hendrycks, D.; Burns, C.; Basart, S.; Zou, A.; Mazeika, M.; Song, D.; and Steinhardt, J. 2020. Measuring massive multitask language understanding. arXiv preprint arXiv:2009.03300.

Hsieh, C.-Y.; Li, C.-L.; Yeh, C.-K.; Nakhost, H.; Fujii, Y.; Ratner, A.; Krishna, R.; Lee, C.-Y.; and Pfister, T. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. arXiv preprint arXiv:2305.02301.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.; et al. 2022. Lora: Low-rank adaptation of large language models. ICLR, 1(2): 3.

Huang, C.; Tang, M.; Ma, Q.; Huang, J.; and Liu, X. 2023. Promoting collaboration in cross-silo federated learning: Challenges and opportunities. IEEE Communications Magazine, 62(4): 82–88.

Huang, Y.; Chu, L.; Zhou, Z.; Wang, L.; Liu, J.; Pei, J.; and Zhang, Y. 2021. Personalized cross-silo federated learning on non-iid data. In Proceedings of the AAAI conference on artificial intelligence, volume 35, 7865–7873.

Imteaj, A.; Mamun Ahmed, K.; Thakker, U.; Wang, S.; Li, J.; and Amini, M. H. 2022. Federated learning for resource-constrained iot devices: Panoramas and state of the art. Federated and transfer learning, 7–27.

Jin, D.; Pan, E.; Oufattole, N.; Weng, W.-H.; Fang, H.; and Szolovits, P. 2021. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. Applied Sciences, 11(14): 6421.

- Jin, Q.; Dhingra, B.; Liu, Z.; Cohen, W.; and Lu, X. 2019. PubMedQA: A Dataset for Biomedical Research Question Answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2567–2577. Hong Kong, China: Association for Computational Linguistics.
- Kim, S.; Joo, S. J.; Kim, D.; Jang, J.; Ye, S.; Shin, J.; and Seo, M. 2023. The cot collection: Improving zero-shot and few-shot learning of language models via chain-of-thought fine-tuning. *arXiv preprint arXiv:2305.14045*.
- Krouka, M.; Elgabli, A.; Issaid, C. B.; and Bennis, M. 2021. Communication-efficient and federated multi-agent reinforcement learning. *IEEE Transactions on Cognitive Communications and Networking*, 8(1): 311–320.
- Li, S.; Chen, J.; Shen, Y.; Chen, Z.; Zhang, X.; Li, Z.; Wang, H.; Qian, J.; Peng, B.; Mao, Y.; et al. 2022. Explanations from large language models make small reasoners better. *arXiv preprint arXiv:2210.06726*.
- Li, T.; Sanjabi, M.; Beirami, A.; and Smith, V. 2019. Fair resource allocation in federated learning. *arXiv preprint arXiv:1905.10497*.
- Li, Y.; Zhang, J.; Feng, S.; Yuan, P.; Wang, X.; Shi, J.; Zhang, Y.; Tan, C.; Pan, B.; Hu, Y.; et al. 2025. Revisiting self-consistency from dynamic distributional alignment perspective on answer aggregation. *arXiv preprint arXiv:2502.19830*.
- Lightman, H.; Kosaraju, V.; Burda, Y.; Edwards, H.; Baker, B.; Lee, T.; Leike, J.; Schulman, J.; Sutskever, I.; and Cobbe, K. 2023. Let’s verify step by step. In *The Twelfth International Conference on Learning Representations*.
- Liu, K.; Hu, S.; Wu, S. Z.; and Smith, V. 2022. On privacy and personalization in cross-silo federated learning. *Advances in neural information processing systems*, 35: 5925–5940.
- Magister, L. C.; Mallinson, J.; Adamek, J.; Malmi, E.; and Severyn, A. 2022. Teaching small language models to reason. *arXiv preprint arXiv:2212.08410*.
- Mao, Y.; Huang, K.; Guan, C.; Bao, G.; Mo, F.; and Xu, J. 2024. Dora: Enhancing parameter-efficient fine-tuning with dynamic rank distribution. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.
- Nair, I.; and Wang, L. 2024. MIDGARD: Self-Consistency Using Minimum Description Length for Structured Commonsense Reasoning. *arXiv preprint arXiv:2405.05189*.
- Pal, A.; Umapathi, L. K.; and Sankarasubbu, M. 2022. Medmcqa: A large-scale multi-subject multi-choice dataset for medical domain question answering. In *Conference on health, inference, and learning*, 248–260. PMLR.
- Qi, J.; Zhou, Q.; Lei, L.; and Zheng, K. 2021. Federated reinforcement learning: Techniques, applications, and open challenges. *arXiv preprint arXiv:2108.11887*.
- Qian, Y.; Rao, L.; Ma, C.; Wei, K.; Ding, M.; and Shi, L. 2024. Toward efficient and secure object detection with sparse federated training over internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 25(10): 14507–14520.
- Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.
- Shi, W.; Xu, R.; Zhuang, Y.; Yu, Y.; Sun, H.; Wu, H.; Yang, C.; and Wang, M. D. 2024. Medadapter: Efficient test-time adaptation of large language models towards medical reasoning. *arXiv preprint arXiv:2405.03000*.
- Tang, M.; and Wong, V. W. 2021. An incentive mechanism for cross-silo federated learning: A public goods perspective. In *IEEE INFOCOM 2021-IEEE conference on computer communications*, 1–10. IEEE.
- Tariq, A.; Serhani, M. A.; Sallabi, F.; Qayyum, T.; Barka, E. S.; and Shuaib, K. A. 2023. Trustworthy federated learning: A survey. *arXiv preprint arXiv:2305.11537*.
- Tariq, A.; Serhani, M. A.; Sallabi, F. M.; Barka, E. S.; Qayyum, T.; Khater, H. M.; and Shuaib, K. A. 2024. Trustworthy federated learning: A comprehensive review, architecture, key challenges, and future research prospects. *IEEE Open Journal of the Communications Society*.
- Team, K.; Du, A.; Gao, B.; Xing, B.; Jiang, C.; Chen, C.; Li, C.; Xiao, C.; Du, C.; Liao, C.; et al. 2025. Kimi k1. 5: Scaling reinforcement learning with llms. *arXiv preprint arXiv:2501.12599*.
- Tian, C.; Shi, Z.; and Li, L. 2023. Learn to select: Efficient cross-device federated learning via reinforcement learning.
- Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- Tsatsaronis, G.; Balikas, G.; Malakasiotis, P.; Partalas, I.; Zschunke, M.; Alvers, M. R.; Weissenborn, D.; Krithara, A.; Petridis, S.; Polychronopoulos, D.; et al. 2015. An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. *BMC bioinformatics*, 16: 1–28.

- Wan, G.; Wu, Y.; Chen, J.; and Li, S. 2024. Reasoning aware self-consistency: Leveraging reasoning paths for efficient llm sampling. arXiv preprint arXiv:2408.17017.
- Wang, H.; Jia, Y.; Zhang, M.; Hu, Q.; Ren, H.; Sun, P.; Wen, Y.; and Zhang, T. 2024a. Feddse: Distribution-aware sub-model extraction for federated learning over resource-constrained devices. In Proceedings of the ACM Web Conference 2024, 2902–2913.
- Wang, P.; Wang, Z.; Li, Z.; Gao, Y.; Yin, B.; and Ren, X. 2023. Scott: Self-consistent chain-of-thought distillation. arXiv preprint arXiv:2305.01879.
- Wang, X.; Wei, J.; Schuurmans, D.; Le, Q.; Chi, E.; Narang, S.; Chowdhery, A.; and Zhou, D. 2022. Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171.
- Wang, Z.; Shen, Z.; He, Y.; Sun, G.; Wang, H.; Lyu, L.; and Li, A. 2024b. Flora: Federated fine-tuning large language models with heterogeneous low-rank adaptations. arXiv preprint arXiv:2409.05976.
- Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.
- Wei, S.; Tong, Y.; Zhou, Z.; Xu, Y.; Gao, J.; Wei, T.; He, T.; and Lv, W. 2025. Federated reasoning LLMs: a survey. *Frontiers of Computer Science*, 19(12): 1912613.
- Wu, H.; Chen, C.; and Wang, L. 2020. A theoretical perspective on differentially private federated multi-task learning. arXiv preprint arXiv:2011.07179.
- Wu, Y.; Sun, Z.; Li, S.; Welleck, S.; and Yang, Y. 2024. Inference scaling laws: An empirical analysis of compute-optimal inference for problem-solving with language models. arXiv preprint arXiv:2408.00724.
- Wu, Y.; Tian, C.; Li, J.; Sun, H.; Tam, K.; Li, L.; and Xu, C. 2025. A survey on federated fine-tuning of large language models. arXiv preprint arXiv:2503.12016.
- Xie, Y.; Kawaguchi, K.; Zhao, Y.; Zhao, J. X.; Kan, M.-Y.; He, J.; and Xie, M. 2023. Self-evaluation guided beam search for reasoning. *Advances in Neural Information Processing Systems*, 36: 41618–41650.
- Ye, T.; Wei, S.; Cui, J.; Chen, C.; Fu, Y.; and Gao, M. 2023. Robust clustered federated learning. In *International Conference on Database Systems for Advanced Applications*, 677–692. Springer.
- Zhang, J.; Vahidian, S.; Kuo, M.; Li, C.; Zhang, R.; Yu, T.; Wang, G.; and Chen, Y. 2024a. Towards building the federatedgpt: Federated instruction tuning. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6915–6919. IEEE.
- Zhang, J.; Wu, Q.; Fan, P.; and Fan, Q. 2024b. A Comprehensive Survey on Joint Resource Allocation Strategies in Federated Edge Learning. arXiv preprint arXiv:2410.07881.
- Zhou, G.; Qiu, P.; Chen, C.; Wang, J.; Yang, Z.; Xu, J.; and Qiu, M. 2025. Reinforced mllm: A survey on rl-based reasoning in multimodal large language models. arXiv preprint arXiv:2504.21277.

算法

在这里，我们展示了整个 FedCoT 方法的伪代码，详见算法 1：

Algorithm 1: FedCoT 算法

```
Input: Total rounds  $T$ ; Local training epochs  $E$ 
      ; Client datasets  $\{\mathcal{D}_i\}_{i=1}^N$ ; Pretrained
      model  $d_\theta$ ; Number of candidates  $K$ 
for  $t = 1$  to  $T$  do
  foreach client  $i$  do
    Receive last aggregated adapter  $\mathbf{W}^{t-1}$ 
     $(\mathbf{A}_i^t, \mathbf{B}_i^t, \mathbf{W}_i^{cls}) \leftarrow$ 
      LocalUpdates( $i, \mathbf{W}^{t-1}$ );
  end
  Aggregate LoRA modules and classifier using
  Equation 6, 7 to get adapter  $\mathbf{W}^t$ ;
end
Function LocalUpdates ( $i, \mathbf{W}$ ):
  for  $e = 1$  to  $E$  do
    Setup local discriminator from  $d_\theta$  with
     $\mathbf{W}$  and apply local LoRA modules
     $\mathbf{A}_i, \mathbf{B}_i$  and classifier;
    Generate reasoning paths  $\{(\tau_{j,k}, \hat{y}_{j,k})\}_{k=1}^K$ 
    ;
    Concatenate feature vector  $\mathbf{h}_{j,k}$ ;
    Predict using Equation 8;
    Update LoRA modules and classifier
    using Equation 5;
  end
  Return updated LoRA modules and classifier
  to server;
foreach test sample  $x_j$  do
  Generate reasoning paths  $\{(\tau_{j,k}, \hat{y}_{j,k})\}_{k=1}^K$ ;
  Concatenate feature vector
   $\mathbf{h}_{j,k} = [x_j \parallel \tau_{j,k} \parallel \hat{y}_{j,k}]$ ;
  Select answer via Equations 8, 9;
end
Output: Final answer  $\{\hat{y}_j\}$ 
```

案例研究

表格 7 中的这个案例涉及一名 29 岁的男性，他患有烧灼感的排尿（尿道炎）、急性不对称关节痛（右脚踝、左膝）、双侧结膜炎以及近期抗生素治疗后的血性腹泻。

问题要求从四个选项中选择最可能的附加发现。正确答案是 B（跟腱附着点的压痛），这表明是由肠道感染（例如志贺菌/沙门氏菌）引发的反应性关节炎。

第一代错误地优先考虑了指关节疼痛，尽管患者的脚踝疼痛是关键线索。反应性关节炎通常涉及下肢（例如，跟腱附着点炎症），而不是指关节。第二代忽略了腹泻和关节症状之间的两个星期潜伏期——这是反应性（而非化脓性）关节炎的标志。

第三代正确诊断了反应性关节炎，并将跟腱压痛（B）作为主要的额外发现，与患者的脚踝疼痛一致。

其他代生成的答案都是错误的。

判别模型正确地将最高分（0.834）赋予生成 3，因为它准确地将病理生理学联系起来：将跟腱压痛与反应性关节炎联系起来。

其他几代在似是而非但不正确的“腹部感染”理论上得分为 0.62–0.79，但模型仍然将它们排在正确答案之下。

改进潜力

表 8 显示了多重采样下，多个模型在三种不同参数水平下的测试集改进情况，其中参数小于 3B 的模型几乎无法通过多重采样获得性能提升。拥有 3B 参数的模型在多重采样下的改进显著高于参数小于 3B 的模型，并且与 7B 参数的模型相比，其改进性能相对更接近。

此外，我们可以看到，在少于 3B 的情况下，Qwen 系列模型仍然可以实现轻微的提升，而 LLaMA 系列模型基本上没有任何提升。这可能是由于预训练期间的不同处理所致。

总体而言，无论是 Qwen 系列还是 LLaMA 系列，模型参数量越大，在相应数据集条件下的基本性能越好，并且通过采样获得的改进也越大。这反映出模型的基本能力是进一步提升的一个非常关键的部分。如果基础模型的基本能力太低，那么即使进行更多的采样也无法实现显著的改进。

数据集信息

我们将在此展示用于培训阶段生成 CoT 候选项的医学数据集和测试阶段用于评估的医学数据集。

- PubMedQA (Jin et al. 2019)：一个基于 PubMed 摘要的生物医学问答数据集，包含 1000 个专家标注的多项选择（是/否/可能）生物医学问答样本。其知识来源涉及 211,300 篇 PubMed 论文。此任务需要根据摘要内容判断研究问题的答案。问题干和上下文的平均长度分别为 14.4 和 238.9 标记。
- BioASQ (Tsatsaronis et al. 2015)：一个大规模生物医学语义索引和问答数据集，包括两个任务：信息检索（任务 A）和机器阅读理解（任务 B）。本文重点研究任务 B 中近年来（2019 - 2023 年）的 618 个二元判断（是/否）阅读理解问题。问题词的平均长度为 17。与 PubMedQA 相似，该数据集通过提供标准的答案文本片段支持机器阅读理解研究。
- MMLU-Med (Hendrycks et al. 2020)：一个从多任务语言理解基准 MMLU（以下简称 MMLU）中提取的医学推理专用数据集。在本文的实验中，选择了七个与医学相关的领域：临床知识、大学生物学、大学医学、高中生物学、医学遗传学、专业医学和病毒学。主要研究模型在专业医学场景中的推理能力。
- MedMCQA (Pal, Umapathi, and Sankarasubbu 2022)：一个大规模的多选医学问答数据集，整合了来自印度的全印度医学科学研究所 (AIIMS) 和全国研究生入学考试 (NEET-PG) 的真实问题资源。该数据集包含超过 194,000 个高质量的医学问题，涵盖 2,400 个健康主题和 21 个医学学科领域，具有显著的主题多样性。问题的平均标记长度为 12.77。
- MedQA (Jin et al. 2021)：基于美国医师执照考试 (USMLE) 的临床医学题库，包含 12,723 道来自 18 本广泛使用的权威临床医学教科书的问题。这些问题涵盖了丰富多样的临床医学主题，并且需要通过

Question :A previously healthy 29-year-old man comes to the emergency department because of burning with urination for several days. He has also had pain in the right ankle for 3 days and pain and swelling in the left knee for 1 day. Two weeks ago, he had several days of fever and bloody diarrhea, for which he was treated with antibiotics. Examination shows a small left knee effusion and bilateral conjunctival injection. Which of the following is the most likely additional finding in this patient?\n A. Palpable mass in the right lower quadrant\n B. Tenderness at the insertion of the Achilles tendon\n C. Pain on passive extension of the fingers\n D. Circular erythematous rash with central clearing Ground-Truth Answer : B ;

Generation 1 : Here is the step-by-step reasoning to answer the question:\n\nThe patient presents with a combination of symptoms, including burning with urination, pain in the right ankle, and pain and swelling in the left knee. These symptoms suggest a systemic infection, rather than a localized issue.\n\nThe patient's recent history of fever and bloody diarrhea, which was treated with antibiotics, is likely related to an underlying infection. The development of new symptoms, such as joint pain and swelling, suggests that the infection may have spread to other parts of the body.\n\nThe patient's conjunctival injection (inflammation of the conjunctiva) is also a significant finding, as it is a common manifestation of disseminated infection.\n\nGiven the patient's symptoms and medical history, the most likely diagnosis is reactive arthritis, which is a type of spondyloarthropathy that occurs in response to an underlying infection, often gastrointestinal.\n\nReactive arthritis typically affects the joints, eyes, and skin, and can cause symptoms such as joint pain, swelling, and inflammation. The patient's bilateral conjunctival injection and left knee effusion are consistent with this diagnosis.\n\nThe most likely additional finding in this patient would be:\n\n# # # # C. Pain on passive extension of the fingers\n\nThis symptom is a common manifestation of reactive arthritis, as the condition can cause inflammation and stiffness in the hands and fingers. Passive extension of the fingers would likely cause pain and stiffness, which would be consistent with the patient's underlying diagnosis.\n\nThe other options are not as likely:\n\nA. Palpable mass in the right lower quadrant: While the patient has symptoms in the right ankle, there is no indication of a palpable mass in this area.\n\nB. Tenderness at the insertion of the Achilles tendon: Achilles tendonitis is a possible diagnosis, but it is not the most likely explanation for the patient's symptoms given the patient's medical history and conjunctival injection.\n\nD. Circular erythematous rash with central clearing: While a rash could be a symptom of reactive arthritis, it is not the most likely finding in this patient given the patient's other symptoms and medical history.

Prediction 1 : C ; Score 1: 0.77344

Generation 2 : Here is the step-by-step reasoning to answer the question:\n\nThe patient presents with a complex set of symptoms, including burning with urination, pain in the right ankle, and pain and swelling in the left knee. The patient also reports a history of fever and bloody diarrhea two weeks ago, which was treated with antibiotics.\n\nStep 1: Identify the possible underlying causes of the patient's symptoms.\n\nThe patient's symptoms suggest a possible infection, given the presence of fever, bloody diarrhea, and pain and swelling in the joints. The antibiotics given two weeks ago may have contributed to the development of a resistant infection.\n\nStep 2: Consider the possible sources of the infection.\n\nGiven the patient's history of bloody diarrhea, it is possible that the infection originated in the gastrointestinal tract. The patient's symptoms of burning with urination and bilateral conjunctival injection suggest that the infection may have spread to other parts of the body, such as the urinary tract and eyes.\n\nStep 3: Evaluate the patient's symptoms in relation to the possible sources of infection.\n\nThe patient's symptoms of pain and swelling in the left knee and right ankle suggest that the infection may have spread to the joints. The patient's small left knee effusion and bilateral conjunctival injection support this possibility.\n\nStep 4: Consider the possible causes of the patient's symptoms in the context of the possible sources of infection.\n\nGiven the patient's history of bloody diarrhea and the presence of symptoms in the joints and eyes, it is possible that the patient has developed a septic arthritis, which is an infection of the joint that is caused by bacteria that have spread from another part of the body.\n\nStep 5: Evaluate the patient's symptoms in relation to the possible causes of septic arthritis.\n\nThe patient's symptoms of pain and swelling in the joints, as well as the presence of a small left knee effusion, are consistent with septic arthritis. The patient's bilateral conjunctival injection also suggests that the infection may have spread to the eyes.\n\nStep 6: Consider the most likely additional finding in the patient.\n\nGiven the patient's symptoms and the possible causes of septic arthritis, the most likely additional finding is a palpable mass in the right lower quadrant, which could indicate an abscess or a collection of pus in the right lower quadrant of the abdomen.\n\n# # # # A. Palpable mass in the right lower quadrant

Prediction 2 : A ; Score 2: 0.717285

Generation 3 : Here is the step-by-step reasoning:\n\nThe patient presents with a combination of symptoms, including burning with urination, pain in the right ankle, and pain and swelling in the left knee. This suggests that the patient may have a systemic infection that is affecting multiple parts of the body.\n\nThe patient had a recent episode of bloody diarrhea, which is a common symptom of gastrointestinal infections. The fact that he was treated with antibiotics suggests that the infection was likely bacterial in nature.\n\nThe patient's symptoms are consistent with a urinary tract infection (UTI), as burning with urination is a common symptom of UTIs. The presence of a small left knee effusion and bilateral conjunctival injection suggests that the infection may have spread to the joints and eyes, respectively.\n\nThe patient's symptoms are also consistent with reactive arthritis, a condition that can occur as a complication of certain bacterial infections, such as Salmonella or Shigella. Reactive arthritis is characterized by joint pain and swelling, as well as conjunctivitis and urethritis.\n\nGiven the patient's symptoms and medical history, the most likely additional finding is:\n\nB. Tenderness at the insertion of the Achilles tendon\n\nThis is because reactive arthritis often affects the joints, including the ankles, and the Achilles tendon is a common site of tenderness in this condition.\n\nB. Tenderness at the insertion of the Achilles tendon

Prediction 3 : B ; Score 3: 0.833984

Generation 4 :

Prediction 4 : A ; Score 4: 0.621094

Generation 5 :

Prediction 5 : A ; Score 5: 0.785645

Generation 6 :

Prediction 6 : A ; Score 6: 0.781250

Generation 7 :

Prediction 7 : A ; Score 7: 0.777344

Generation 8 :

Prediction 8 : C ; Score 8: 0.717285

Table 7: 关于 MedQA 数据集的案例研究。

Model	bioasq		medmcqa		medqa		mmlu		pubmedqa		average		Δ
	p@1	p@k	p@1	p@k	p@1	p@k	p@1	p@k	p@1	p@k	p@1	p@k	
LLaMA-3-8B	37.90	70.97	29.84	68.18	27.18	70.46	38.65	71.78	9.20	42.00	28.55	64.68	36.12
LLaMA-3.1-8B	25.81	64.52	35.07	72.22	32.05	73.53	39.26	77.30	15.20	54.80	29.48	68.47	38.99
Qwen2.5-7B	73.39	98.39	43.75	70.98	29.46	66.54	50.31	78.53	38.80	73.40	47.14	77.57	30.43
Qwen2.5-3B	10.48	46.77	18.98	55.41	5.03	23.80	26.38	65.64	1.20	7.40	12.41	39.81	27.39
LLaMA3.2-3B	31.45	70.16	31.41	65.57	20.03	59.62	32.52	70.55	4.80	24.20	24.04	58.02	33.98
Qwen-1.5B	0.81	9.68	1.65	11.09	1.73	15.48	0.61	10.43	0.00	0.40	0.96	9.41	8.46
LLaMA3.2-1B	0.00	0.00	0.10	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.05
Qwen2.5-0.5B	0.00	0.81	1.63	8.92	0.47	5.89	1.23	7.98	0.60	3.20	0.78	5.36	4.57

Table 8: 生物医学测试集上模型表现（以% 表示），通过多次采样展示准确性潜力。表格显示首次采样时的通过率（p@1）以及 k 次采样后的通过率（p@k），其中 $\Delta = \text{p@k} - \text{p@1}$ 表示准确性改善潜力。

整合多源证据进行专业级别的推理。问题的平均长度为 116.6 个标记，选项的平均长度为 3.5 个标记。

提示模板

我们在图 4 中展示了一次性 CoT 提示，用于生成候选推理路径。这个精心设计的模板指导模型进行严格推理，并遵循必要的格式以便于答案提取和评估。

The Prompt Template for one-shot CoT

To answer the following question, provide a highly detailed and comprehensive step-by-step reasoning. Your explanation should break down the problem into clear, logical steps, with each step on a new line.

The goal is to construct a robust and self-contained argument that fully justifies the final conclusion. After the reasoning, present the final answer. Please model the overall structure on the provided example.

Example:

Q:

Primary Colonization of plaque is dominated by
A. Facultative Aerobic Gram+ve rods
B. Facultative Anaerobic Gram-ve rods
C. Facultative Aerobic Gram-ve cocci
D. Facultative Anaerobic Gram+ve cocci

A:

The oral cavity is a complex ecosystem where various microorganisms coexist, and plaque is a biofilm that forms on tooth surfaces.

Primary colonization of plaque is dominated by Streptococcus species, which are Gram-positive, facultative anaerobic cocci.

These Streptococcus species are able to adhere to the tooth surface and multiply, creating a matrix for other microorganisms to colonize.

D. Facultative Anaerobic Gram+ve cocci

Here is the question:

Q:

{ question }
{ options }

Figure 4: Template for Multiple-choice Questions (MCQs). Placeholders { question } and { options } denote the MCQ stem and options.