arXiv:2603.01124v1 [cs.CV] 1 Mar 2026

ClinCoT：面向医学视觉语言模型的临床感知视觉思维链

Xiwei Liu[1], Yulong Li[1], Xinlin Zhuang[1,2], Xuhui Li[1], Jianxu Chen[1], Haolin Yang[1],Imran Razzak[1],和 Yutong Xie[1,†]

[1] 穆罕默德·本·扎耶德人工智能大学[2] 华东师范大学

[†]通讯作者

摘要。 医学视觉语言模型在临床决策支持方面展现出良好潜力，但由于缺乏对局部病理证据的充分事实锚定，仍容易产生事实性幻觉。现有医学对齐方法主要通过偏好优化在响应层面进行处理，虽然提升了输出的正确性，却使中间推理与视觉区域之间的联系仍然薄弱。尽管思维链（CoT）增强了多模态推理，但其仍主要以文本为中心，限制了临床视觉线索的有效整合。为弥补这一缺口，我们提出 ClinCoT，一种临床感知视觉思维链框架，将偏好优化从响应层面的纠正转变为视觉驱动推理。我们引入了一种自动数据生成流程，通过基于假设驱动的区域提议进行推理，构建具有临床依据的偏好对。多个 Med-LLM 评估器对每个响应进行排序并分配分数，这些排序结果被用作监督信号来训练目标模型。我们进一步提出一种基于评分的间隔感知优化策略，将偏好排序与分数差异相结合，以优化区域级推理轨迹。为在训练过程中随着模型策略的演变维持对齐，我们采用迭代学习方案，动态重新生成偏好数据。在三个医学视觉问答和报告生成基准测试上的大量实验表明，与现有基于偏好的对齐方法相比，ClinCoT 持续提升了事实锚定能力，并取得了更优的性能。我们的代码已发布于 https://github.com/Xiwei-web/ClinCoT

关键词： Med-VLM · 思维链 · 偏好优化。

1 引言

医学视觉语言模型（Med-VLMs）在辅助临床决策方面展现出可观的潜力，包括医学视觉问答（Med-VQA）和放射学报告生成 [14][17][18][35]。然而，尽管近期取得了进展，Med-VLMs 仍面临一个根本性局限：视觉证据与生成的临床结论之间对齐不足 [36][34]。具体而言，模态对齐能力较差的模型可能会过度依赖

预训练语言先验，而未能充分利用局部病理证据，从而产生幻觉性发现或与临床无关的回答 [4][11]。

为提高事实一致性，近期的医学对齐方法采用偏好优化策略 [1][9]。MMedPO 等临床感知偏好学习框架 [36] 构建具有医学意义的偏好对，并引入加权优化目标，以突出临床相关的回答。其他对齐方法则通过自我奖励机制 [32],、借助更强评估器进行自训练 [25],或模态对齐目标 [34]来优化模型输出。尽管这些方法提升了回答层面的事实准确性并减少了明显的幻觉，但它们主要在输出层面发挥作用，将每个回答视为一个整体。因此，它们没有明确建模局部病变区域如何影响中间推理步骤。缺乏区域感知推理会限制医学视觉语言模型的可解释性和病理依据，尤其是在复杂诊断场景中。

除了偏好优化之外，思维链（CoT）推理 [27][7][29]已成为一种有效策略，能够提升语言模型和视觉语言模型的复杂推理与逐步决策能力。在医学领域，结构化诊断提示和推理增强训练策略已被证明能够提高解释质量和逻辑一致性 [23][24]。然而，现有的大多数 CoT 方法仍以文本为中心 [7]：它们引导模型生成序列化的推理标记，却没有明确地重构视觉注意力。这实际上假设视觉编码器能够均匀捕捉所有临床相关信息，而这在医学影像中往往并不现实。放射学诊断推理从根本上依赖于检测和检查局部异常，例如微小结节、细微实变或局灶性骨折 [30]。

在医疗实践中，临床医生很少对整幅图像进行均匀推理。相反，他们形成鉴别诊断假设，检查疾病相关区域，并根据局部证据迭代完善结论。这自然引出一个根本问题：

偏好优化能否从响应级纠正扩展到假设驱动的临床推理？

为应对这一挑战，我们提出了 ClinCoT——一种临床感知视觉思维链框架，在统一的推理范式下，将区域级诊断假设与边际感知偏好优化相结合。ClinCoT 不仅致力于提升最终输出，还显式建模局部病理证据如何影响中间推理轨迹，从而以过程驱动的方式促进视觉定位与临床推断更加紧密地对齐。具体而言，我们设计了一个自动化两阶段流程，用于构建具有临床依据的偏好数据：1）假设驱动的区域生成：给定一张医学图像，我们使用临床感知视觉工具生成多个由疾病条件化的区域。随后，目标 Med-VLM 通过联合处理原始图像和每个候选区域来回答问题，从而形成多条病理感知推理链。2）共识加权质量评估：多个医学大语言模型评估器对生成的回答进行评分。这些分数

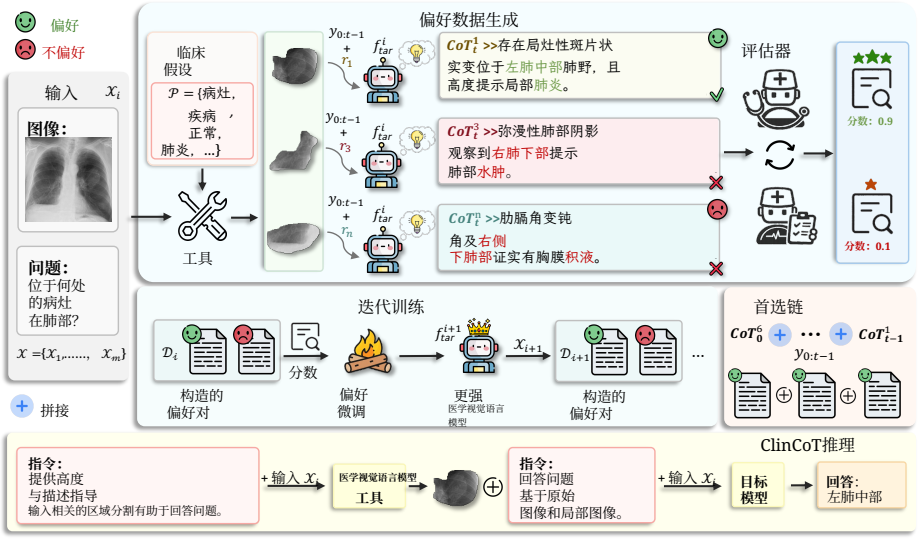


图1. ClinCoT工作流概览。给定输入对$\mathcal{X}_i \subset \{\mathcal{X}_1, \ldots, \mathcal{X}_m\}$的一个子集，ClinCoT首先使用配备预定义假设集 $\mathcal{P}$的临床感知工具来生成区域提议 $\{r_i\}_{i=1}^n$。目标模型 f_{tar}^i基于保留的偏好历史 $y_{0:t-1} = \{CoT_0^6, \ldots, CoT_{t-1}^1\}$生成区域条件推理链 $\{y_t^i = CoT_t^i\}_{i=1}^n$，同时整合原始图像和候选区域。医学大语言模型评估器为其分配分数，以构建偏好对 $\mathcal{D}_i$，区分偏好响应和非偏好响应。基于共识加权评分的优化通过迭代训练，将 f_{tar}^i更新为 f_{tar}^{i+1}。随后，将更新后的模型 f_{tar}^{i+1}应用于 $\mathcal{X}_{i+1}$，为下一轮迭代生成新的偏好对 $\mathcal{D}_{i+1}$。

可作为区域条件推理链临床有效性的代理指标，并根据评估器一致性进行调整，以确保稳健监督。

我们的贡献主要有三点： 提出一种临床假设驱动的自动化流程，用于可扩展的区域级偏好数据构建； 提出一种基于共识加权评分的偏好优化方法，并通过迭代学习实现病理感知推理对齐，从而更精细地区分关键区域，并逐步稳定推理轨迹；以及 在多个医学视觉问答和报告生成基准测试上开展了大量实验，证明该方法相较于强大的医学基线方法取得了持续改进。

2 方法

给定一个目标模型 f_{tar}、一组评估器模型 f_{eval}和一个图像-问题对 $x = \{x_v, x_q\}$，我们展示了如何构建 n 偏好数据点，如图1所示。假设推理链包含 T 个步骤，偏好数据会在这些 T 阶段中逐步生成。在每个时间步 t，该流程由三个阶段组成：响应生成、响应评估和

配对构建（见第 2.1 节）。不同于标准的直接偏好优化（DPO） [1]，我们采用基于共识加权评分的 DPO，该方法不仅能够对偏好数据进行排序，还能为其分配偏好分数。这种评分方式可基于分数差异实现更精确的优化。在训练过程中，通过最小化负对数似然损失，偏好数据的排序充当监督信号，而分数则定义决策边际（见第 2.2 节）。

2.1 偏好数据生成

响应生成。 本阶段旨在构建具有临床依据的区域假设，并生成以局部视觉上下文为条件的中间响应。“响应”是指模型在中间推理步骤中的任何输出，而不一定是对该问题的最终答案。我们将时间步 t 处的响应记为 y_t，并将初始输入 x 表示为 y_0。给定医学图像x_v 以及一组预定义的临床假设 $\mathcal{P} = \{p_i\}_{i=1}^n$，我们采用临床感知视觉语言模型（例如 MedKLIP [28]） $\mathcal{T}$（·）获取疾病条件激活图，以突出与临床概念 p_i相关联的图像区域：

$$h_i = \mathcal{T}(x_v, p_i), \quad i = 1, \ldots, n, \tag{1}$$

其中，每个 $h_i \in [0,1]^{H\times W}$ 随后通过阈值化和连通分量提取用于定位临床区域：$r_i = x_v \odot \Phi$（h_i）。此处，Φ（·）表示一个确定性区域提取算子，其中 $\odot$表示逐元素掩码。对于每个区域假设 r_i，目标医学视觉语言模型 f_{tar} 会生成同时以全局图像和局部区域为条件的中间响应：

$$y_t^i = f_{\mathsf{tar}}(x_v, r_i, x_q \mid y_{0:t-1}), \tag{2}$$

从而得到一组候选响应 $\{y_t^i\}_{i=1}^n$ ，它们对应于同一图像的不同区域条件视觉解释，其中 $y_{0:t-1}$表示先前各时间步中当前保留的推理链。

响应评估。 本阶段通过提示评估模型为所有生成的响应赋予0到1之间的分数来评估其质量，分数越高表示预测越好。评估模型不仅为每个单独的响应打分，还会计入该响应如何影响链中下一响应的质量。这种累积评估方法有助于量化每个有界区域对整体推理过程的影响。在时间步t，评估模型按如下方式为 y_t^i 分配分数：

$$s^i_{\mathsf{cur}} = f_{\mathsf{eval}}(y_t^i \mid y_{0:t-1}), \quad s^i_{\mathsf{nxt}} = \mathbb{E}_j[f_{\mathsf{eval}}(y_{t+1}^j \mid y_{0:t-1}, y_t^i)], \quad s^i = s^i_{\mathsf{cur}} + \gamma s^i_{\mathsf{nxt}}, \tag{3}$$

其中， s^i_{nxt} 反映对下一条响应的影响，期望值 $\mathbb{E}[\cdot]$ 通过根据当前响应 y_t^i随机采样 j 条后续响应来估计。$\gamma > 0$是一个超参数，用于结合当前响应分数和下一条响应分数，最后一步则取 $\gamma = 0$ 。为减轻单个 Med-LLM 固有的偏差并确保评估可靠，我们采用共识加权评分 策略

该策略使用两个不同的评估器。我们计算这两个评估器之间的一致性，并据此对有争议的评估进行惩罚，具体如下：

$$s^i_{\text{final}} = \left(\frac{s^i_1 + s^i_2}{2}\right) \cdot \exp\left(-|s^i_1 - s^i_2|\right) \tag{4}$$

其中， s^i_1 和 s^i_2 分别是两位 Med-LLM 评估器给出的分数。

配对构建。 在这些带分数的推理响应候选 $\{y^i_t\}^n_{i=1}$中，对于 $\{s^i_{\text{final}}\}^n_{i=1}$，我们选择 k 对，以便在每个时间步 t构建偏好比较。对于每一对响应，分数较高的响应被视为偏好响应，并与历史链 $y_{0:t-1}$ 拼接，形成“首选链” y^w_t。同样，分数较低的响应与相同的历史链拼接，形成“非偏好链” y^l_t。每对链及其对应的分数（y_w， s_w， y_l， s_l）共同构成一个偏好训练样本。所有 k 此类配对构成时间步 t 的偏好数据集 $\mathcal{D}_t$ 。为确保稳定的前向推理轨迹，只有时间步 t 中分数最高的响应会与历史链拼接，形成更新后的推理链 $y_{0:t}$。随后，该更新后的链将作为下一时间步的输入。换言之，尽管每个时间步都会创建多个偏好配对用于学习，但只有得分最高的单条链会被保留，以继续前向推理过程。

2.2 偏好微调

间隔感知优化。 给定输入 x，语言模型策略 π_θ可以生成条件分布 $\pi_\theta(y \mid x)$，其中 y 是输出文本响应。标准 DPO [1] 在 PPO 框架 [22]下，通过对奖励进行重新参数化，将奖励学习重新表述为策略优化：

$$r(x, y) = \beta \log \frac{\pi_\theta(y \mid x)}{\pi_{\text{ref}}(y \mid x)} + \beta \log Z(x). \tag{5}$$

其中， π_{ref} 表示控制策略正则化的冻结初始化参数 π_θ ， β 是温度参数， $Z(x)$是与 y无关的归一化项。在 Bradley–Terry模型 [3],下， DPO 最小化偏好响应 y_w 的排名高于不偏好响应 y_l的负对数似然：

$$\begin{aligned} P(y_w \succ y_l) &= \sigma(r(x, y_w) - r(x, y_l)), \\ \mathcal{L}_{\text{DPO}} &= -\mathbb{E}_{(x, y_w, y_l)\sim\mathcal{D}} \left[\log P(y_w \succ y_l)\right], \end{aligned} \tag{6}$$

其中， $\sigma(\cdot)$ 表示 Sigmoid函数。在图像级推理中，不同区域对最终决策的贡献并不相同，因此需要对偏好对进行更细致的区分。为捕捉这种差异性，我们引入了一个由偏好分数导出的间隔项，并构建了间隔感知目标函数：

$$\mathcal{L}_{\text{ClinCoT}}(\theta) = -\mathbb{E}_{(x, y_w, y_l)\sim\mathcal{D}} \left[\log \sigma\left(\beta \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)} - (g(s_w) - g(s_l))\right)\right]. \tag{7}$$



图 2. 生成的偏好数据可视化 ta.

其中，函数 $g(\cdot)$ 单调递增，并将偏好分数映射到 DPO 目标函数的 logit 空间，从而增强关键区域的影响。

设 $\Delta_r = g(s_w) - g(s_l)$，并定义服从 Gumbel 分布的随机变量$R_w \sim \text{Gumbel}\left(r\left(x, y_w\right), 1\right)$ 和 $R_l \sim \text{Gumbel}\left(r\left(x, y_l\right), 1\right)$。经偏好差距调整后，获胜响应受到偏好的概率为:

$$
\begin{aligned}
&P\left(R_w - R_l > \Delta_r\right) = \sigma\left(r\left(x, y_w\right) - r\left(x, y_l\right) - \Delta_r\right) \\
&= \sigma\left(\beta \log \frac{\pi_\theta\left(y_w \mid x\right)}{\pi_{\text{ref}}\left(y_w \mid x\right)} - \beta \log \frac{\pi_\theta\left(y_l \mid x\right)}{\pi_{\text{ref}}\left(y_l \mid x\right)} - \Delta_r\right).
\end{aligned} \tag{8}
$$

这一结果由 Gumbel 随机变量 [1] 的定义和 Gumbel-max 技巧 [19], 得出，类似的推导见于 ODTO [1]。Gumbel 分布对变量的极值进行建模，而 Δ_r 则量化偏好对之间差异的程度。通过最大化对数似然，我们进而得到调整后的损失函数，该函数通过区分偏好差异的顺序和大小，显式地优化偏好学习。

迭代学习。 标准 DPO 在静态偏好数据集上进行优化，随着模型在训练过程中不断演化，这可能会导致分布失配。为缓解这一问题，我们采用了一种受 [31], 启发的迭代偏好学习策略，其中 ClinCoT 通过逐步更新数据生成与优化过程来改进目标模型。具体而言，完整的图像-问题集 $\mathcal{X}$ 被划分为 m 个互不相交的子集 $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_m\}$，每个子集对应一次迭代。以初始目标模型 f_{tar} 和评估器 f_{eval}为起点，第 i 次迭代首先使用当前模型 f_{tar}^i 在子集 $\mathcal{X}_i$上生成偏好数据 $\mathcal{D}_i$ 。随后，通过在 $\mathcal{D}_i$上优化公式（7），将模型更新为 f_{tar}^{i+1}。该过程重复进行 m 轮，最终得到模型 f_{tar}^m。

表 1. 医学视觉问答和报告生成任务的性能比较。对于开放式问题，我们报告召回率（开放式）；对于封闭式问题，我们报告准确率（封闭式）。BLEU 分数表示 BLEU-1/2/3/4 的平均值。 +SFT 表示模型在应用相应基线方法之前，先使用监督微调进行微调。最佳结果和次佳结果分别以粗体和下划线标出。

模型	SLAKE		VQA-RAD		IU-Xray		
	Open	已关闭	Open	已关闭	BLEU	ROUGE-L	METEOR
LLaVA-Med v1.5	44.26	61.30	29.24	63.97	14.56	10.31	10.95
+ DPO	49.30	62.02	29.76	64.70	16.08	12.95	17.13
+ 自我奖励	42.63	61.30	33.29	64.17	14.20	10.38	10.52
+ STLLaVA-Med	48.65	61.75	30.17	64.38	16.11	10.58	10.51
+ POVID	52.43	70.35	31.77	65.07	20.80	24.33	30.05
+ SIMA	51.77	69.10	31.23	64.80	17.11	22.87	29.10
+ FiSAO	52.69	70.46	32.70	64.11	21.06	25.72	30.82
+ MMedPO	**53.99**	73.08	**36.36**	**66.54**	23.49	29.52	34.16
+ **ClinCoT**	53.73±0.3						
+ 监督微调	50.45	65.62	31.38	64.26	22.75	28.86	33.66
+ DPO	53.50	69.47	32.88	64.33	23.07	29.97	34.89
+ 自我奖励	50.62	65.89	32.69	65.89	22.89	28.97	33.93
+ STLLaVA-Med	52.72	66.69	33.72	64.70	22.79	28.98	34.05
+ POVID	52.18	70.67	32.95	64.97	23.95	29.75	34.63
+ SIMA	51.75	69.28	32.50	64.08	23.90	29.41	34.45
+ FiSAO	52.80	70.82	32.94	65.77	23.57	29.88	35.01
+ MMedPO	55.23	75.24	34.03	67.64	24.00	30.13	35.17
+ **ClinCoT**	**56.13**±0.5						

3 实验

3.1 实验设置

评估数据集。为验证 ClinCoT 在提升事实准确性方面的有效性，我们使用了三个医学基准：两个医学视觉问答数据集，即 VQA-RAD [13] 和 SLAKE [16], ，以及一个报告生成数据集 IU-Xray [8]。**实现细节**。 我们使用 LLaVA-Med-1.5 7B [14] 作为目标模型，并采用 LoRA微调 [10] 进行偏好优化，批量大小为 4，学习率为 1e-7，$\gamma = 0.3$、$\beta = 0.1$ ，训练 3 个周期。每条推理链包含 $T = 3$ 个时间步，每步包含 $k = 2$ 。ClinCoT 的迭代学习进行 $m = 4$ 轮。LLaMA3-Med42-7B [5]和 BioMistral-7B [12]用作医学大语言模型评估器。所有实验均使用 PyTorch 2.1.2，在两块 NVIDIA RTX A100 GPU 上完成。

基线方法。 我们将 ClinCoT 与标准 DPO [21] 及其变体进行比较，包括自奖励方法 [33] 和 STLLaVA-Med [25]。自奖励方法根据模型生成的响应构建偏好对，而 STLLaVA-Med 借助 GPT-4o 提升偏好选择能力，以实现医学领域的对齐。我们还评估了三种最初为自然图像开发的视觉语言模型偏好微调方法：POVID [34],、SIMA [26],和 FiSAO [6], ，以及专为医学领域设计的 MMedPO [36]。为进一步评估训练兼容性,

表 2. ClinCoT 关键组件的消融研究。“w/o ClinCoT”移除中间推理并直接输出答案。“w/ naive DPO”采用标准 DPO 损失。“w/o iterative learning”在单次过程中生成所有偏好对，且仅训练一次。“w/o γ model”在响应评估阶段设置 $\gamma = 0$。“single evaluator”使用单一 Med-LLM LLaMA3-Med42-7B 作为评估器。

模型	SLAKE		VQA-RAD		IU-Xray		
	Open	关闭	Open	关闭	BLEU	ROUGE-L	METEOR
ClinCoT（无监督微调）	53.73±0.3	73.41±0.2	33.15±0.2	65.88±0.1	24.96±0.2	29.79±0.1	35.44±0.2
无 ClinCoT	41.09±0.2	56.83±0.1	24.55±0.1	51.38±0.3	17.13±0.2	23.44±0.3	26.85±0.3
使用朴素DPO	50.22±0.1	71.39±0.3	31.06±0.4	62.61±0.1	22.76±0.2	28.51±0.1	33.18±0.4
不使用迭代学习	49.97±0.5	69.83±0.4	30.23±0.4	60.75±0.3	21.84±0.6	26.19±0.3	31.91±0.5
无 γ	46.53±0.1	65.37±0.2	28.71±0.1	57.69±0.1	20.03±0.3	25.42±0.3	29.89±0.1
单一评估器	52.16±0.7	72.28±0.4	32.34±0.5	64.95±0.2	23.62±0.4	28.81±0.3	34.73±0.2

我们在监督微调（SFT）设置下评估 ClinCoT 和所有基线方法，以考察其在不同训练范式下的性能提升。

评估指标。对于医学视觉问答任务，我们使用准确率和召回率评估封闭式和开放式问题。对于报告生成任务，我们使用 BLEU分数 [20], ROUGE-L [15] 和 METEOR [2]作为评估指标。

3.2 主要结果

总体性能对比结果见表 1。与所比较的基线方法相比，ClinCoT在报告生成任务上表现最佳，但在 VQA-RAD 上不及 MMedPO。这可能表明，对于简答式视觉问答，基于回答层面的临床加权仍具竞争力，因为简洁、精准的语言表达在这类任务中更为重要。相比之下，ClinCoT强调区域条件化推理链；如果没有预先进行任务适配，中间推理步骤可能会带来不稳定性。在监督微调增强的设置下，ClinCoT取得了整体最佳表现。这表明监督微调可能提供了更稳定的领域对齐初始化，从而有助于后续的假设驱动精炼。当基础模型与领域风格对齐后，ClinCoT会进一步精炼回答和推理轨迹，从而带来更一致的提升。图 2 展示了推理过程中一些偏好数据的可视化结果。

3.3 消融研究

在表 2 中，我们展示了对关键组件的消融实验。**1）图像级思维链：**移除思维链会导致性能显著下降，证实了其必要性。这凸显了我们方法的潜力，并为未来提升区域准确性指明了方向。**（2）边际感知直接偏好优化：**结果显示，性能持续下降，表明纳入偏好分数差距对于区分推理链之间的细微差异至关重要。**（3）迭代学习：**消除迭代更新会导致明显的性能下降，这表明动态生成偏好数据有助于在模型演进过程中保持对齐

。**（4）响应评估：**移除下一步得分或使用单一评估器会降低性能，尤其是在报告生成方面，这表明稳定的共识感知评分能够提升长程推理质量。

4 结论

我们提出了 ClinCoT，将偏好优化从响应级别转向假设驱动推理。通过结合疾病条件区域提议、共识加权边际优化和迭代学习，ClinCoT 使中间推理与局部病理证据相一致。在医学视觉问答和报告生成基准测试上的实验表明，与强大的医学基线方法相比，ClinCoT 始终能够带来改进；在监督微调下还能进一步提升。这些结果表明，可将区域级临床推理融入偏好学习，从而提升医学视觉语言模型的事实锚定能力和推理稳定性。

参考文献

1. Amini, A., Vieira, T., Cotterell, R.：带偏移的直接偏好优化。arXiv预印本 arXiv:2402.10571（2024） 2. Banerjee, S., Lavie, A.：Meteor：一种用于机器翻译评估的自动指标，与人工判断具有更高的相关性。载于：机器翻译和/或摘要的内在与外在评估指标ACL研讨会论文集。第65–72页（2005） 3. Bradley, R.A., Terry, M.E.：不完全区组设计的秩分析：I. 配对比较法。Biometrika **39**（3/4），324–345（1952） 4. Chen, J., Yang, D., Wu, T., Jiang, Y., Hou, X., Li, M., Wang, S., Xiao, D., Li, K., Zhang, L.：检测和评估大型视觉语言模型中的医学幻觉。arXiv预印本 arXiv:2406.10185（2024） 5. Christophe, C., Kanithi, P.K., Raha, T., Khan, S., Pimentel, M.A.：Med42-v2：临床大语言模型套件。arXiv预印本 arXiv:2408.06142（2024） 6. Cui, C., Zhang, A., Zhou, Y., Chen, Z., Deng, G., Yao, H., Chua, T.S.：细粒度验证器：视觉-语言对齐中的偏好建模作为下一词元预测。arXiv预印本 arXiv:2410.14148（2024） 7. Dai, P., Ou, Y., Yang, Y., Jin, Z., Suzuki, K.：Goca：通过显式思维蒸馏实现可靠决策的可信多模态检索增强生成，应用于医学大型视觉语言模型。载于：医学图像计算与计算机辅助干预国际会议。第251–261页。Springer（2025） 8. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R., McDonald, C.J.：准备供分发和检索的放射学检查集合。美国医学信息学协会期刊 **23**（2），304–310（2016） 9. Hein, D., Chen, Z., Ostmeier, S., Xu, J., Varma, M., Reis, E.P., Michalson, A.E., Bluethgen, C., Shin, H.J., Langlotz, C.等：在无人工反馈的情况下，通过偏好微调提升胸部X光解读模型的事实性（2024） 10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.：LoRA：大语言模型的低秩自适应。arXiv预印本 arXiv:2106.09685（2021）

11. Jiang, Y., Chen, J., Yang, D., Li, M., Wang, S., Wu, T., Li, K., Zhang, L.：Med-think：通过更多思考诱导医学大规模视觉语言模型减少幻觉。arXiv预印本 arXiv:2406.11451 **1**（8）（2024） 12. Labrak, Y., Bazoge, A., Morin, E., Gourraud, P.A., Rouvier, M., Dufour, R.：Biomistral：面向医学领域的开源预训练大语言模型集合。arXiv预印本 arXiv:2402.10373（2024） 13. Lau, J.J., Gayen, S., Ben Abacha, A., Demner-Fushman, D.：关于放射影像的临床生成视觉问答数据集。Scientific Data **5**（1），1–10（2018） 14. Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.：Llava-med：用一天时间训练用于生物医学的语言与视觉助手。载于：第三十七届神经信息处理系统进展数据集与基准测试专题（2023） 15. Lin, C.Y.：Rouge：用于摘要自动评估的软件包。载于：摘要总结另辟蹊径。第74–81页（2004） 16. Liu, B., Zhan, L.M., Xu, L., Ma, L., Yang, Y., Wu, X.M.：SLAKE：用于医学视觉问答的语义标注知识增强数据集。载于：2021 IEEE第18届国际生物医学成像研讨会（ISBI）。第1650–1654页。IEEE（2021） 17. Liu, C., Wan, Z., Ouyang, C., Shah, A., Bai, W., Arcucci, R.：结合多模态学习与测试时临床知识增强的零样本ECG分类。arXiv预印本 arXiv:2403.06659（2024） 18. Liu, X., Xu, M., Ho, Q.：Fias：采用动态融合与混合注意力的特征不平衡感知医学图像分割。载于：2025 IEEE第22届国际生物医学成像研讨会（ISBI）。第1–5页。IEEE（2025） 19. Maddison, C.J., Tarlow, D.：Gumbel machinery（2017），https://cmaddis.github.io/gumbel-machinery，在线提供 20. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.：BLEU：一种机器翻译自动评估方法。载于：计算语言学协会第40届年会论文集。第311–318页（2002） 21. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S.：直接偏好优化：你的语言模型其实是一个隐藏的奖励模型。NeurIPS（2023） 22. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.：近端策略优化算法。arXiv预印本 arXiv:1707.06347（2017） 23. Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S.等：大语言模型编码临床知识。Nature **620**（7972），172–180（2023） 24. Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., Hou, L., Clark, K., Pfohl, S.R., Cole-Lewis, H.等：利用大语言模型实现专家级医学问答。Nature Medicine **31**（3），943–950（2025） 25. Sun, G., Qin, C., Fu, H., Wang, L., Tao, Z.：STLLaVA-Med：面向医学的自训练大型语言与视觉助手。arXiv预印本 arXiv:2406.19973（2024） 26. Wang, X., Chen, J., Wang, Z., Zhou, Y., Zhou, Y., Yao, H., Zhou, T., Goldstein, T., Bhatia, P., Huang, F.等：通过自我改进增强大型视觉语言模型中的视觉语言模态对齐。arXiv预印本 arXiv:2405.15973（2024） 27. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D.等：思维链提示激发大语言模型中的推理能力。神经信息处理系统进展 **35**，24824–24837（2022）

28. Wu, C., Zhang, X., Zhang, Y., Wang, Y., Xie, W.：Medklip：通过医学知识增强的语言-图像预训练实现X光诊断。载于：IEEE/CVF国际计算机视觉会议论文集，第21372–21383 页（2023） 29. Wu, M., Liu, Z., Yan, Y., Li, X., Yu, S., Zeng, Z., Gu, Y., Yu, G.：Rankcot：通过排序思维链优化检索增强生成中的知识。载于：第63届计算语言学协会年会论文集（第1卷：长篇论文），第 12857–12874 页（2025） 30. Xu, W., Chan, H.P., Li, L., Aljunied, M., Yuan, R., Wang, J., Xiao, C., Chen, G., Liu, C., Li, Z., 等：Lingshu：用于统一多模态医学理解与推理的通用基础模型。arXiv 预印本 arXiv:2506.07044（2025） 31. Yu, T., Zhang, H., Yao, Y., Dang, Y., Chen, D., Lu, X., Cui, G., He, T., Liu, Z., Chua, T.S., 等：Rlaif-v：通过开源人工智能反馈提升多模态大语言模型与超可信GPT-4V的对齐。arXiv 预印本 arXiv:2405.17220（2024） 32. Yuan, W., Pang, R.Y., Cho, K., Li, X., Sukhbaatar, S., Xu, J., Weston, J.E.：自我奖励语言模型。载于：第四十一届国际机器学习大会（2024） 33. Yuan, W., Pang, R.Y., Cho, K., Li, X., Sukhbaatar, S., Xu, J., Weston, J.E.：自我奖励语言模型。载于：第四十一届国际机器学习大会（2024） 34. Zhou, Y., Cui, C., Rafailov, R., Finn, C.：通过偏好微调实现视觉大语言模型中的模态对齐。arXiv 预印本 arXiv:2402.11411（2024） 35. Zhu, H., Liu, X., Xue, R., Zhang, Z., Xu, Y., Ergu, D., Cai, Y., Zhao, Y.：Sss：采用高效提示的半监督SAM-2，用于医学影像分割。arXiv 预印本 arXiv:2506.08949（2025） 36. Zhu, K., Xia, P., Li, Y., Zhu, H., Wang, S., Yao, H.：Mmedpo：通过临床感知多模态偏好优化实现医学视觉语言模型对齐。arXiv 预印本 arXiv:2412.06141（2024）