Multi-Modal Multi-Agent Reinforcement Learning for Radiology Report Generation

Kaito Baba **Risa Kishikawa** **Satoshi Kodera**
Department of Cardiovascular Medicine
The University of Tokyo Hospital, Tokyo, Japan
baba-kaito662@g.ecc.u-tokyo.ac.jp

Abstract

We propose **MARL-Rad**, a multi-modal multi-agent reinforcement learning framework for radiology report generation that trains the entire agentic system on policy within its deployed radiology workflow. MARL-Rad addresses the limitation of post-hoc agentization, where fixed LLMs are organized into hand-designed agentic workflows without being optimized for their assigned roles. Our framework decomposes chest X-ray interpretation into region-specific agents and a global integrating agent, and jointly optimizes them using clinically verifiable rewards. Experiments on the MIMIC-CXR and IU X-ray datasets show that MARL-Rad consistently improves clinical efficacy metrics such as RadGraph, CheXbert, and GREEN scores, achieving state-of-the-art clinical efficacy performance. Further analyses show that MARL-Rad improves laterality consistency and produces more accurate and detailed reports. A blinded clinician evaluation further suggests that MARL-Rad produces reports clinically comparable to ground-truth reports.

1 Introduction

Recent advances in large language models (LLMs) and large vision–language models (LVLMs) have demonstrated remarkable reasoning and generation capabilities across a wide range of domains, including dialogue systems, mathematical reasoning, code synthesis, scientific discovery, and medical diagnosis [23, 31, 86, 123]. To further translate these capabilities into complex real-world tasks, agentic systems have emerged as an increasingly important paradigm for leveraging LLMs beyond single-turn generation. By decomposing tasks into manageable subtasks, agentic systems can support structured reasoning and workflow-oriented problem solving [5, 33, 37, 44, 52, 57, 59, 92, 119, 137].

Medicine is one of the most important application domains of LLMs and LVLMs. Among medical modalities, chest X-ray (CXR) is one of the most widely used diagnostic tools in clinical practice, enabling physicians to assess thoracic structures and identify conditions such as pneumonia, pneumothorax, and pleural effusion [2, 11, 12, 81]. Radiologists carefully inspect CXR images and manually compose diagnostic reports based on their interpretations. However, radiology reporting is time-consuming and labor-intensive, and the growing demand for diagnostic imaging can delay diagnosis and degrade report quality [10, 21, 26, 95]. To alleviate this burden, automated radiology report generation (RRG) using LLMs has attracted growing attention as a promising approach [1, 4, 22, 24, 55, 61, 71, 90, 103, 112, 115, 121, 122, 140], with recent studies increasingly exploring agentic-system designs to improve performance [28, 50, 78, 126, 134].

Despite these recent developments, most existing agentic systems do not train the underlying LLMs for their deployed workflows. They typically keep pretrained or domain-adapted LLMs fixed and rely on prompts and hand-designed workflows to elicit role-specific behavior [44, 52, 59, 92, 137]. As a result, role specialization and coordination are imposed only through prompt context: the workflow expects agents to act as specialized and coordinated components, while the underlying LLMs remain fixed policies that have not been optimized for their positions in the workflow. This mismatch is particularly consequential in real-world tasks that require specialized expertise, such as medicine. In

Preprint.

such settings, domain-specific fine-tuning alone is insufficient; the full agent system must be trained on policy within the deployed workflow. Indeed, in Section 4.3, we show that even MedGemma [94], a medical adaptation of Gemma 3 [106], performs poorly when its parameters are kept fixed and it is merely organized into a multi-agent workflow for RRG.

To address this limitation, we propose **MARL-Rad**, a multi-modal multi-agent reinforcement learning framework that trains the entire agentic RRG system end-to-end on policy within its deployed workflow. MARL-Rad consists of region-specific agents responsible for localized observations and a global integrating agent that synthesizes their outputs. These agents are jointly trained with clinically verifiable rewards, optimizing the agentic system within its deployed workflow. This workflow mirrors the real practice of radiologists, who meticulously examine each anatomical region before composing a comprehensive diagnostic report. Experiments on the MIMIC-CXR and IU X-ray datasets demonstrate that MARL-Rad consistently improves clinical efficacy (CE) metrics such as RadGraph F1 [43], CheXbert F1 [100], and GREEN scores [87], achieving state-of-the-art performance. Moreover, deeper analyses show that MARL-Rad improves laterality consistency and produces more detailed and clinically accurate descriptions. Finally, a small blinded clinician evaluation suggests that MARL-Rad produces clinically comparable reports to ground-truth reports.

Our key contributions are summarized as follows:

- **Workflow-aligned training beyond fixed agentification**: We show that simply organizing a fixed LLM into an agentic workflow is insufficient and can even degrade performance. MARL-Rad addresses this limitation by training the entire agentic system on policy within the same radiology-inspired workflow in which it is deployed.
- **State-of-the-art performance on CE metrics**: Experiments on the MIMIC-CXR and IU X-ray datasets demonstrate that MARL-Rad achieves state-of-the-art performance on various CE metrics, including RadGraph F1, CheXbert F1, and GREEN scores.
- **Enhanced laterality consistency and accurate, detail-informed reports**: MARL-Rad improves laterality consistency and generates more detailed and clinically accurate reports compared to single-agent RL baselines.

2 Related work

Here we review the prior work most relevant to ours; further discussion is provided in Appendix A.

LLMs for radiology report generation (RRG). Early RRG systems mostly followed encoder–decoder paradigms with Transformers, such as R2Gen [15], which established strong baselines on the MIMIC-CXR [48] and IU X-ray [25] datasets. Recent LLM-centric approaches align a medical visual encoder with a frozen or fine-tuned LLM to produce more fluent, clinically grounded reports, such as R2GenGPT [114], XrayGPT [107], MAIRA-1 [42], MAIRA-2 [7], and CheXagent [17], along with other approaches [1, 3, 8, 20, 24, 26, 30, 32, 34–36, 38–40, 45, 46, 49, 51, 55, 56, 71, 73, 74, 79, 83, 84, 90, 91, 103, 108, 111, 115–117, 120–122, 124, 125, 127, 128, 131, 133, 135, 136, 140, 141].

Agentic systems for RRG. Multi-agent frameworks are beginning to appear in RRG, where they align with clinical reasoning stages or combine retrieval-augmented generation (RAG). RadAgents [134] proposes a radiologist-like, multi-agent workflow for chest X-ray interpretation. CXRAgent [78] introduces a director-orchestrated, multi-stage agent for chest X-ray interpretation that validates tool outputs with an evidence-driven validator and coordinates diagnostic planning and team-based reasoning. Yi et al. [126] decomposes CXR reporting into retrieval, draft, refinement, vision, and synthesis agents aligned with stepwise clinical reasoning. Elboardy et al. [28] proposes a model-agnostic, ten-agent framework that unifies radiology report generation and evaluation, coordinated by an orchestrator and including an LLM-as-a-judge. However, these agentic RRG systems are mostly training-free, relying on pretrained models without end-to-end optimization of entire systems.

Reinforcement learning for RRG. Recent studies have incorporated reinforcement learning (RL) to enhance clinical accuracy in RRG. LM-RRG [140] integrates clinical-quality RL by directly optimizing RadCliQ [129] as a reward signal. BoxMed-RL [47] couples chain-of-thought supervision with spatially verifiable RL that ties textual findings to bounding-box evidence. DeepMedix-R1 [65] employs a three-stage pipeline: instruction-fine-tuning, synthetic reasoning sample exposure, and online RL. OraPO [18] proposes an oracle-educated group relative policy optimization (GRPO) for RRG that leverages fact-level rewards. Med-R1 [54] applies GRPO-based RL to medical VLMs across eight imaging modalities and five VQA task types. However, these RL approaches primarily focus on optimizing a single model rather than optimizing entire multi-agent systems.

3 Method

3.1 Preliminaries

In this section, we introduce the notations necessary for this paper, and as background, briefly describe Group Sequence Policy Optimization (GSPO) [139], which is a sequence-level variant of GRPO [96] that improved performance.

Let $\mathcal{D}$ denote the data distribution over query–answer pairs (q, a). Like GRPO, GSPO samples a group of G responses $(\{x_i\}_{i=1}^{G} \sim \pi_{\theta_{\text{old}}}(\cdot \mid q))$ for each query q, where $\pi_{\theta_{\text{old}}}$ denotes the policy used to generate outputs before updating the current policy. Each sampled response x_i is then evaluated by a verifiable reward $r(x_i, a)$, and group-relative advantage is computed as follows:

$$\hat{A}_i \coloneqq \frac{r(x_i, a) - \text{mean}(\{r(x_i, a)\}_{i=1}^{G})}{\text{std}(\{r(x_i, a)\}_{i=1}^{G})}. \tag{1}$$

GSPO then performs policy optimization by maximizing the following objective:

$$\mathcal{J}_{\text{GSPO}}(\theta) \coloneqq \mathbb{E}_{(q,a)\sim\mathcal{D},\{x_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|q)}\left[\frac{1}{G}\sum_{i=1}^{G}\min\Big(s_i(\theta)\hat{A}_i, \text{clip}(s_i(\theta), 1-\varepsilon_{\text{low}}, 1+\varepsilon_{\text{high}})\hat{A}_i\Big)\right], \tag{2}$$

where

$$s_i(\theta) \coloneqq \left(\frac{\pi_\theta(x_i \mid q)}{\pi_{\theta_{\text{old}}}(x_i \mid q)}\right)^{1/|x_i|} = \left(\frac{\prod_{t=1}^{|x_i|}\pi_\theta(x_{i,t} \mid q, x_{i,<t})}{\prod_{t=1}^{|x_i|}\pi_{\theta_{\text{old}}}(x_{i,t} \mid q, x_{i,<t})}\right)^{1/|x_i|} \tag{3}$$

is the importance ratio based on sequence likelihoods, and $|x_i|$ denotes the number of tokens of response x_i. Here, $x_{i,t}$ and $x_{i,<t} \coloneqq (x_{i,1}, \ldots, x_{i,t-1})$ denote the t-th tokens and the preceding tokens of the response x_i, respectively.

3.2 Multi-agent GSPO

In this section, we introduce a simple yet effective multi-agent formulation of GSPO. Although the formulation is a simple extension of GSPO, it enables the entire agentic system to be optimized end-to-end in an on-policy manner, encouraging coordination among agents within the deployed workflow. The adaptation to RRG is discussed in Section 3.3.

Let there be K agents with policies $\{\pi_{\theta^{(k)}}\}_{k=1}^{K}$, where the indices $k = 1, \ldots, K$ follow the order in which the agents are activated. For each $(q, a) \sim \mathcal{D}$, we sample a *group of G joint rollouts*:

$$\{x_i\}_{i=1}^{G} \sim \pi_{\theta_{\text{old}}}^{\text{agent}}(\cdot \mid q) \coloneqq \prod_{k=1}^{K}\pi_{\theta_{\text{old}}^{(k)}}(\cdot \mid c_i^{(k)}), \tag{4}$$

where $x_i \coloneqq (x_i^{(1)}, \ldots, x_i^{(K)})$. Here, $c_i^{(k)}$ denotes the context observed by agent k when generating $x_i^{(k)}$. This context may include the query q or the outputs of preceding agents.

Each joint rollout is evaluated by a system-level verifiable reward $r_i \coloneqq r(x_i^{(K)}, a)$, and the resulting group-relative advantage computed by Eq. (1) is shared across all agents.

Then, we define the *multi-agent GSPO (MA-GSPO)* objective as follows:

$$\begin{aligned}&\mathcal{J}_{\text{MA-GSPO}}(\theta_1, \ldots, \theta_K)\\ &= \mathbb{E}_{(q,a)\sim\mathcal{D},\{x_i\}\sim\pi_{\theta_{\text{old}}}^{\text{agent}}(\cdot|q)}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{1}{G}\sum_{i=1}^{G}\min\Big(s_i^{(k)}(\theta_k)\hat{A}_i, \text{clip}\big(s_i^{(k)}(\theta_k), 1-\varepsilon_{\text{low}}, 1+\varepsilon_{\text{high}}\big)\hat{A}_i\Big)\right],\end{aligned} \tag{5}$$

where the importance ratio for agent k is:

$$s_i^{(k)}(\theta_k) = \left(\frac{\pi_{\theta_k}\big(x_i^{(k)} \mid c_i^{(k)}\big)}{\pi_{\theta_k^{\text{old}}}\big(x_i^{(k)} \mid c_i^{(k)}\big)}\right)^{1/|x_i^{(k)}|}. \tag{6}$$

This reduces to standard GSPO when $K = 1$. Although the advantage is shared, each agent is updated through its own sequence-level importance ratio, so the objective optimizes the joint agentic system while preserving agent-specific policy updates.

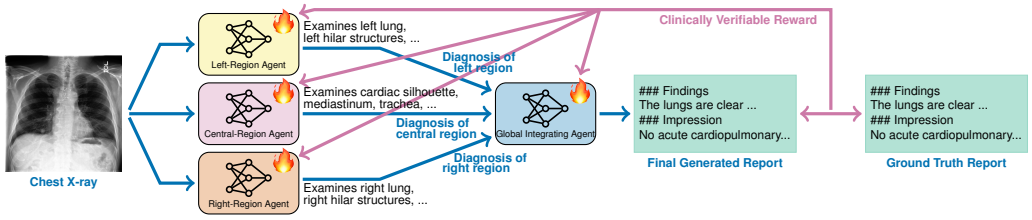

Figure 1: Overview of the proposed multi-agent RL framework. Region-specific agents and the global integrating agent collaboratively generate the radiology report, and the entire agent system is jointly optimized through RL based on clinically verifiable rewards.

3.3 Multi-agent RL on RRG

The overall architecture of MARL-Rad is illustrated in Fig. 1. Our system consists of three region-specific agents and one global integrating agent that collaborate to generate the final diagnostic report. Each region-specific agent focuses on a distinct anatomical region of the chest X-ray: a *left-region agent* examines structures such as the left lung, left hilar structures, left costophrenic angle, and left clavicle; a *right-region agent* examines structures such as the right lung, right hilar structures, right costophrenic angle, and right clavicle; and a *central-region agent* examines structures such as the cardiac silhouette, mediastinum, aortic arch, trachea, and spine. This design follows the actual diagnostic workflow of radiologists, who meticulously examine each anatomical region of a chest X-ray before composing the complete report. The *global integrating agent* then receives the outputs from the three region-specific agents, incorporates their diagnoses, and produces a final comprehensive report reflecting the overall condition of the chest X-ray.

For the reinforcement learning, we employ clinically verifiable rewards using the CheXbert [100] accuracy, which measures the correctness of predicted clinical findings based on automatically labeled disease categories, and the RadGraph F1 [43], which evaluates factual and relational consistency. Additionally, we incorporate the ROUGE-L [64] to encourage lexical alignment with reference reports. The final reward is defined as the unweighted sum of these three components, and the entire agent system is jointly optimized as we described in Section 3.2.

4 Experiments

4.1 Experimental setup

In this section, we describe the overview of the experimental setup. Further details are provided in Appendix B.

Datasets We train our agents using the MIMIC-CXR [48] dataset. Following prior studies [4, 15, 41, 55, 72, 85], we exclude samples without the corresponding reports. We adopt the official split of MIMIC-CXR and use the training set for training our agents. Evaluation is performed on the official test split of MIMIC-CXR and the IU X-ray [25] dataset. For IU X-ray, we follow the standard 70%/20%/10% split used in prior work [15, 16, 16, 36, 39, 41, 67, 97, 113, 120] and use only the test split for cross-dataset evaluation.

Evaluation metrics Following previous studies [1, 15, 16, 16, 36, 39, 41, 55, 58, 65, 66, 70–72, 75, 85, 90, 97, 104, 109, 110, 113, 120–122, 134, 140], we evaluate the generated reports using both natural language generation (NLG) metrics and CE metrics. For NLG metrics, we report BLEU-1, BLEU-4 [88], METEOR [6], and ROUGE-L [64]. For CE metrics, we report RadGraph F1 [43], CheXbert F1 [100], and GREEN scores [87]. CheXbert is a 14-label classification models that assess the presence of common thoracic findings (e.g., pneumonia, atelectasis, edema) from generated reports. RadGraph measures the correctness of clinical entity and relation extraction, reflecting the structural and factual consistency of the report. The GREEN score is a LLM-based metric that evaluates clinical efficacy by grading report-level clinical correctness beyond surface-level lexical overlap. Consistent with prior work, evaluations are performed on the Findings section alone [16, 36, 55, 71, 85] and on both the Findings and Impression sections combined [55, 71, 105].

Used models We use MedGemma-4B [94] as the base model for all agents, including the three region-specific agents and the global integrating agent. All agents are trained jointly end-to-end from the same base checkpoint, but they do not share parameters, yielding distinct parameter sets.

Table 1: Comparison with previous state-of-the-art methods on the MIMIC-CXR dataset [48]. The best score for each metric within each section is highlighted in **bold**. Our approach achieves the best performance across all clinical efficacy (CE) metrics, which are considered to align more closely with clinicians' assessments [9, 69, 105, 129].

Method	NLG Metrics ↑				CE Metrics ↑		
	BLEU-1	BLEU-4	METEOR	ROUGE-L	RadGraph F1	CheXbert F1	GREEN
MIMIC-CXR Findings							
R2Gen [15]	0.353	0.103	0.142	0.277	-	0.276	-
R2GenCMN [16]	0.353	0.106	0.142	0.278	-	0.278	-
PPKED [68]	0.360	0.106	0.149	0.284	-	-	-
CMCL [67]	0.344	0.097	0.133	0.281	-	-	-
SA [122]	-	-	-	-	0.228	-	-
RGRG [104]	0.373	0.126	0.168	0.264	-	0.447	-
METransformer [113]	0.386	0.124	0.152	0.291	-	0.311	-
KiUT [41]	0.393	0.113	0.160	0.285	-	0.321	-
CoFE [58]	-	0.125	0.176	0.304	-	0.405	-
MAN [97]	0.396	0.115	0.151	0.274	-	0.389	-
Med-LMM [72]	-	0.128	0.161	0.289	-	0.395	-
SEI [70]	-	0.135	0.158	0.299	0.249	0.460	-
FMVP [75]	0.389	0.108	0.150	0.284	-	0.336	-
HERGen [109]	0.395	0.122	0.156	0.285	-	0.317	-
CMN [16]	0.353	0.106	0.142	0.278	-	0.278	-
CXRMate [85]	-	0.079	-	0.262	0.272	0.357	-
I3+C2FD [66]	0.402	0.128	0.175	0.291	-	0.473	-
MLRG [71]	0.411	0.158	0.176	0.320	0.291	0.505	0.353
DART [90]	0.437	0.137	0.175	0.310	-	0.533	-
LM-RRG [140]	-	0.122	0.165	0.296	-	0.484	-
DeepMedix-R1 [65]	0.340	0.105	0.289	**0.329**	0.238	0.239	-
SRRG [36]	0.416	0.128	0.294	0.162	-	0.486	-
MPO [120]	0.416	0.139	0.162	0.309	-	0.353	-
DAMPER [39]	0.402	**0.193**	0.289	0.301	-	0.507	-
OISA [121]	0.428	0.129	-	-	0.244	0.486	0.322
MedGemma [94]	0.285	0.014	0.238	0.203	0.195	0.494	0.361
MARL-Rad (Ours)	**0.533**	0.056	**0.290**	0.275	**0.294**	**0.544**	**0.396**
MIMIC-CXR Findings + Impression							
MedRegA [110]	0.405	0.126	0.319	0.276	-	-	-
CXRMate [85]	-	0.074	0.158	0.255	-	0.378	-
MLRG [71]	0.402	**0.152**	0.172	**0.327**	0.289	0.509	-
Flamingo-CXR [105]	-	0.101	-	0.297	0.205	0.519	-
MedGemma [94]	0.333	0.052	0.338	0.197	0.209	0.523	0.388
MARL-Rad (Ours)	**0.598**	0.142	**0.397**	0.292	**0.316**	**0.556**	**0.398**

4.2 Main results: Comparison with previous state-of-the-art methods

The results are shown in Tables 1 and 2. Our method outperforms the baselines on all CE metrics. As demonstrated in Tanno et al. [105], where evaluations were conducted by 27 board-certified radiologists from two regions (the United States and India), traditional NLG metrics based solely on linguistic similarity do not align with clinicians' assessments, whereas CE metrics are more consistent with clinical judgments. Multiple studies have similarly reported that conventional NLG metrics fail to reflect clinicians' evaluations, underscoring the need to prioritize CE metrics [9, 69, 129]. Considering this, it is reasonable that our method does not align with some NLG metrics, while the strong performance on CE metrics suggests improved clinical relevance.

The consistent improvement on both the MIMIC-CXR and IU X-ray datasets demonstrates the effectiveness across different evaluation settings. Notably, despite being trained exclusively on MIMIC-CXR training set, our agent attains high performance on IU X-ray, suggesting promising robustness and cross-dataset generalization.

We also report the VRAM usage and inference throughput of MARL-Rad in Appendix C, showing that despite the additional computation introduced by agentization, it maintains a practical inference speed for real-world radiology reporting workflows.

Table 2: Comparison with previous state-of-the-art methods on the IU X-ray dataset [25]. The best score for each metric within each section is highlighted in **bold**. Our approach achieves the best performance across all clinical efficacy (CE) metrics, which are considered to align more closely with clinicians' assessments [9, 69, 105, 129].

Method	NLG Metrics ↑				CE Metrics ↑		
	BLEU-1	BLEU-4	METEOR	ROUGE-L	RadGraph F1	CheXbert F1	GREEN
IU X-ray Findings							
R2Gen [15]	0.470	0.165	0.187	0.371	-	-	-
R2GenCMN [16]	0.475	0.170	0.191	0.376	-	-	-
PPKED [68]	0.483	0.168	0.190	0.376	-	-	-
METransformer [113]	0.483	0.172	0.192	0.380	-	-	-
KiUT [41]	0.525	0.185	0.242	0.409	-	-	-
CoFE [58]	-	0.175	0.202	**0.438**	-	-	-
CMCL [67]	0.473	0.162	0.186	0.378	-	-	-
MAN [97]	0.501	0.170	0.213	0.386	-	-	-
CMN [16]	0.475	0.170	0.191	0.375	-	-	-
Med-LMM [72]	-	0.168	**0.381**	-	-	-	-
FMVP [75]	0.485	0.169	0.201	0.398	-	-	-
LM-RRG [140]	-	0.208	0.216	0.387	-	-	-
CXRMate [85]	-	0.046	-	0.282	0.291	0.277	-
I3+C2FD [66]	0.499	0.184	0.208	0.390	-	-	-
SRRG [36]	0.533	0.218	0.219	0.418	-	-	-
MPO [120]	**0.548**	0.209	0.224	0.415	-	-	-
DAMPER [39]	0.520	**0.225**	0.284	0.397	-	-	-
Multi-Agent [126]	-	0.047	0.362	0.247	-	-	-
OISA [121]	0.431	0.131	-	-	0.282	0.219	0.481
MedGemma [94]	0.179	0.012	0.287	0.215	0.270	0.427	0.632
MARL-Rad (Ours)	0.468	0.046	0.347	0.292	**0.337**	**0.501**	**0.644**
IU X-ray Findings + Impression							
CXRMate [85]	-	0.046	-	0.282	0.291	0.277	-
MedGemma [94]	0.244	0.058	0.432	0.212	0.265	0.469	0.653
MARL-Rad (Ours)	**0.591**	**0.182**	**0.531**	**0.348**	**0.365**	**0.516**	**0.659**

4.3 Ablation study

We conduct ablation studies to examine the effects of workflow-aligned RL, regional decomposition, and reward design. The results are presented in Table 3.

4.3.1 Training-free agentic workflow is insufficient

First, we evaluate a training-free agentic workflow in which MedGemma is kept fixed and executed with the same agentization as MARL-Rad. The results are presented as "MedGemma (agent)" in Table 3. This variant performs worse than the vanilla MedGemma model, despite following the same multi-agent workflow as our full method. This suggests that simply organizing a fixed LLM into an agentic workflow is insufficient and can even degrade performance, likely because the underlying model has not been optimized for its assigned roles or for coordinated reasoning within the deployed workflow. This finding highlights the importance of jointly optimizing the entire agent system, rather than relying on naive agentification without learning.

Furthermore, our method also achieves higher scores than GSPO training without agentization (presented as "MedGemma + RL"), indicating that our regional multi-agent workflow itself is effective when the agent system is properly optimized for the deployed workflow. We conduct a more in-depth analysis of this advantage in Section 4.4.

4.3.2 Shared versus counterfactual rewards

A natural concern with using a shared system-level reward is that it may suffer from a credit assignment problem: since all agents receive the same final reward, it may be unclear which agent contributed to the improvement or degradation of the final report. To examine this issue, we consider a naive credit-assignment variant based on counterfactual rewards. Let r_i denote the reward of the original final report for the i-th joint rollout. For a regional agent k, let $r_i^{(\setminus k)}$ denote the reward obtained after regenerating the final report with the global integrating agent while removing agent k's intermediate output from its input context. We then assign agent k the difference reward

Table 3: Ablation study results. "MedGemma (vanilla)" represents the unmodified model without reinforcement learning or agentization. "MedGemma (agent)" employs the same agentic workflow as MARL-Rad but without any RL optimization. "MedGemma + RL" applies GSPO to MedGemma in a single-agent setting. "Counterfactual Reward" uses agent-specific rewards based on counterfactual removal of each regional agent's output from the global agent's context. The best and worst scores for each metric within each dataset are highlighted in **bold** and underline, respectively.

Method	NLG Metrics ↑				CE Metrics ↑		
	BLEU-1	**BLEU-4**	**METEOR**	**ROUGE-L**	**RadGraph F1**	**CheXbert F1**	**GREEN**
MIMIC-CXR Findings + Impression							
MedGemma (vanilla)	0.333	0.052	0.338	0.197	0.209	0.523	0.388
MedGemma (agent)	<u>0.252</u>	<u>0.044</u>	<u>0.267</u>	<u>0.144</u>	<u>0.106</u>	<u>0.309</u>	<u>0.344</u>
MedGemma + RL	0.465	0.116	0.373	0.283	0.299	0.502	0.377
Counterfactual Reward	0.575	0.136	0.375	0.272	0.292	0.518	0.369
MARL-Rad (Ours)	**0.598**	**0.142**	**0.397**	**0.292**	**0.316**	**0.556**	**0.398**
IU X-ray Findings + Impression							
MedGemma (vanilla)	0.244	0.058	0.432	0.212	0.265	0.469	0.653
MedGemma (agent)	<u>0.210</u>	<u>0.053</u>	<u>0.380</u>	<u>0.163</u>	<u>0.141</u>	<u>0.348</u>	0.649
MedGemma + RL	**0.598**	0.181	0.529	0.339	0.358	0.495	0.629
Counterfactual Reward	0.561	0.166	0.513	0.323	0.349	0.478	<u>0.614</u>
MARL-Rad (Ours)	0.591	**0.182**	**0.531**	**0.348**	**0.365**	**0.516**	**0.659**

$r_i^{(k)} := r_i - r_i^{(\setminus k)}$. That is, each regional agent is rewarded according to how much the final report reward decreases when its regional output is removed from the integrating agent's context. For the global integrating agent, we keep the original system-level reward r_i. We report this variant as "Counterfactual Reward" in Table 3.

As shown in Table 3, this counterfactual reward variant does not improve over our shared-reward formulation and instead yields slightly worse clinical metrics. This suggests that explicitly decomposing the final reward in this naive manner is not necessarily beneficial for RRG. One reason is that the counterfactual report is generated from an ablated context that differs from the deployed workflow. Thus, the resulting reward difference reflects not only the contribution of the removed regional agent, but also the global integrating agent's ability to compensate under an out-of-distribution context. In contrast, the success of our simple shared-reward formulation shows that explicit counterfactual reward decomposition is not necessary for this setting. For structured short-horizon agentic RRG workflows, optimizing the complete system with a shared clinically grounded reward is sufficient and preferable to noisy counterfactual reward decomposition.

4.4 Analysis of laterality consistency using RadGraph

To assess the effect of decomposing the task into region-specific agents, we conduct a detailed analysis of the RadGraph outputs to examine region-level consistency. RadGraph extracts entities classified as either "Anatomy" or "Observation" along with the relations between them, enabling the identification of clinically meaningful region–finding pairs, such as "left–lung–pneumonia." Among region-level attributes, laterality (left vs. right) is the one most directly handled by our left/right agents, and it carries critical clinical importance: errors in left–right identification are strictly unacceptable in real clinical practice. Therefore, we extract representative laterality-related entities from the RadGraph output such as "left," "right," "left lung," and "right lung," along with the entities connected to them via relations. Based on these subsets, we define a laterality-specific RadGraph F1 score by ap-

Table 4: Laterality-specific RadGraph F1 scores. "Left Agent" and "Right Agent" denote the pre-integration outputs of the corresponding regional agents. Higher scores indicate better laterality consistency. The best score among the final-report methods, i.e., MedGemma w/ RL and MARL-Rad (Ours), is highlighted in **bold**.

Method	**Left**	**Right**	**Left lung**	**Right lung**
MIMIC-CXR Findings + Impression				
MedGemma w/ RL	0.144	0.142	0.219	0.102
MARL-Rad (Ours)	**0.184**	**0.277**	**0.406**	**0.146**
Left Agent	0.214	0.011	0.369	0.066
Right Agent	0.001	0.257	0.000	0.156
IU X-ray Findings + Impression				
MedGemma w/ RL	0.063	0.044	**0.035**	0.057
MARL-Rad (Ours)	**0.082**	**0.166**	0.032	**0.168**
Left Agent	0.130	0.005	0.155	0.082
Right Agent	0.000	0.169	0.000	0.175

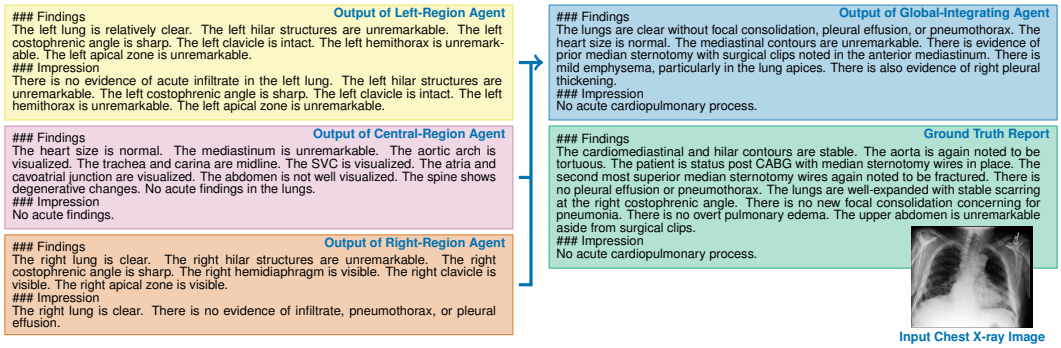

Figure 2: Example output from MARL-Rad. Region-specific agents focus on their assigned regions and generate regional diagnoses. The global integrating agent synthesizes these drafts into a concise and coherent report while adding relevant global findings. The resulting final report is detailed and aligns well with the ground truth, correctly concluding with "No acute cardiopulmonary process."

plying the standard RadGraph evaluation procedure only to these laterality-focused subgraphs and use it for our analysis. The detailed procedure for computing the laterality-specific RadGraph F1 score is described in Appendix D.

The results are shown in Table 4. We observe that MARL-Rad captures laterality more accurately than the single-agent RL baseline. This improvement may be attributed to a limitation identified in recent studies, which report that standard vision–language models tend to behave as global image parsers and struggle to reason about spatial relations at the region level [13, 19]. By introducing region-specific agents—including dedicated left/right agents—our approach decomposes the interpretation of image regions and enables each agent to focus on its corresponding region, thereby enhancing the model's ability to capture laterality.

To further verify that the improvement is not merely driven by the global integrating agent, we also report the laterality-specific RadGraph F1 scores of the left and right agents in Table 4. These scores evaluate each regional agent's output before global integration. As shown in the table, each left/right agent achieves a score comparable to that of the final integrated report on its assigned side, while its score on the opposite side remains much lower. This indicates that MARL-Rad does not simply rely on the global agent to correct or infer regional information; rather, the regional agents themselves learn to produce clinically meaningful localized diagnoses.

4.5 Case study

Fig. 2 shows an example case from our multi-agent RL system. In this example, the left, right, and central region agents each focus on their respective anatomical responsibilities and consistently produce findings and impression texts that are restricted to their assigned regions. This behavior indicates that the intended task decomposition is functioning as designed. Furthermore, the global integrating agent leverages these region-specific drafts and produces a concise and coherent final report, rather than simply concatenating the regional outputs. For instance, given the left and right agents' statements "The left lung is relatively clear" and "The right lung is clear," the integrating agent appropriately compresses these into "The lungs are clear." Similarly, central findings such as "The heart size is normal" are preserved and incorporated into the final report. In addition, the integrating agent also adds global findings, such as post-surgical changes ("There is evidence of prior median sternotomy with surgical clips noted in the anterior mediastinum"), which are not tied to any specific region. As a result, the final report accurately captures both the absence of acute pulmonary abnormalities and the patient's status post cardiac surgery, aligning with the ground-truth conclusion of "No acute cardiopulmonary process."

For comparison, Fig. 3 presents an example generated by the single-agent RL model on the same image, where MedGemma is optimized with RL without agentization. Although the content of the report and its final conclusion are broadly consistent with the ground truth, the report lacks detailed analysis and

Findings
The lungs are clear. There is no focal consolidation, pleural effusion, or pneumothorax. The heart size is normal. The mediastinal and hilar contours are unremarkable.
Impression
No acute cardiopulmonary process.

Figure 3: An example case generated by the single-agent RL model (MedGemma w/ RL) using the same input image as in Fig. 2.

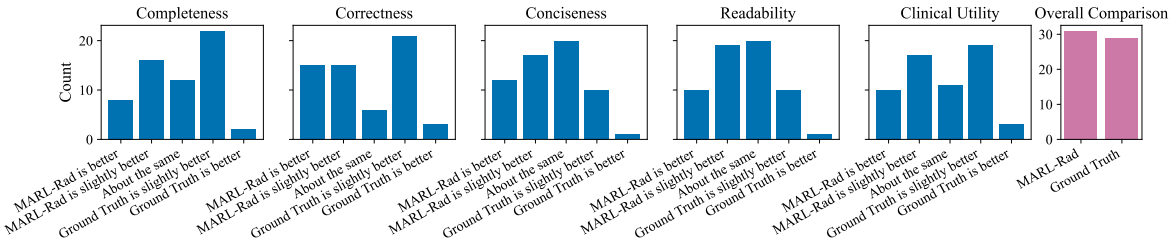

Figure 4: Physician preferences comparing MARL-Rad and the ground-truth reports. Bars show 5-point Likert ratings for five criteria; the rightmost panel shows the overall 2-choice vote. During evaluation, report sources were anonymized and randomly shuffled for each case.

remains somewhat vague, omitting several clinically relevant findings present in the ground-truth report. This contrast underscores the advantage of agentization, which explicitly examines each anatomical region and yields more accurate, detailed reports.

Additionally, as illustrated in this example, agentization improves interpretability: each regional agent produces an explicit, region-focused diagnosis, making it clear how the model assessed each anatomical region. This transparency is particularly valuable for clinicians or end-users who may not be familiar with the internal behavior of LVLMs. In contrast, a single-agent model provides only a monolithic summary, making it difficult to verify whether the underlying regional findings were adequately considered.

5 Clinician evaluation

Automated metrics may not fully capture clinical quality, and improvements on such metrics may not always align with clinician judgment. To partially address this concern, we conduct a small blinded clinician comparison between MARL-Rad outputs and ground-truth reports.

To minimize bias, the two reports for each case were anonymized and randomly ordered. Physicians evaluated each pair together with the corresponding chest X-ray image, without knowing the source of either report. For each case, physicians rated five aspects: completeness, correctness, conciseness, readability, and clinical utility. Detailed definitions of these evaluation criteria are provided in Appendix E. Each aspect was evaluated on a 5-point comparative Likert scale with the options "Report A is better," "Report A is slightly better," "About the same," "Report B is slightly better," and "Report B is better." In addition, physicians provided an overall preference vote between reports A and B. We randomly sampled 10 images from the MIMIC-CXR test set, and six physicians completed the evaluation. The results are shown in Fig. 4.

Across all five criteria, physician ratings show that MARL-Rad outputs are clinically comparable to the ground-truth reports. The overall preference vote is nearly balanced, with a slight preference for MARL-Rad. These results complement the automated evaluation and suggest that the gains of MARL-Rad do not merely reflect overfitting to automatic metrics, but correspond to clinically comparable report quality.

6 Conclusion

In this work, we introduced MARL-Rad, a novel multi-modal multi-agent reinforcement learning framework for radiology report generation. MARL-Rad coordinates region-specific agents with a global integrating agent and jointly optimizes the entire agent system end-to-end to produce clinically consistent reports. By optimizing agents on policy within their deployed workflow, MARL-Rad overcomes the limitations of training-free agentification of fixed LLMs. Our method achieves state-of-the-art performance on CE metrics on both MIMIC-CXR and IU X-ray datasets, enhances laterality consistency, and yields more accurate, detailed reports. Although this work focuses on CXR-based RRG, the framework may be extended to other workflow-structured applications. Future work includes applying this framework to other modalities in medical AI, such as electrocardiogram (ECG) and echocardiography, as well as other real-world tasks.

References

[1] Syed Bilal Ahsan, Muhammad Ikhalas, Muhammad Muzamil Khan, Sana Ullah, and Muhammad Zaigham Zaheer. ARDGen: Augmentation regularization for domain-generalized medical report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 6526–6535, 2025.

[2] Daniel Joseph Alapat, Malavika Venu Menon, and Sharmila Ashok. A review on detection of pneumonia in chest X-ray images using neural networks. *Journal of Biomedical Physics and Engineering*, 12(6):551–558, 2022.

[3] Shahad Albastaki, Anabia Sohail, Iyyakutti Iyappan Ganapathi, Basit Alawode, Asim Khan, Sajid Javed, Naoufel Werghi, Mohammed Bennamoun, and Arif Mahmood. Multi-resolution pathology-language pre-training model with text-guided visual representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25907–25919, 2025.

[4] Kaito Baba, Ryota Yagi, Junichiro Takahashi, Risa Kishikawa, and Satoshi Kodera. JRadiEvo: A japanese radiology report generation model enhanced by evolutionary optimization of model merging. *arXiv preprint arXiv:2411.09933*, 2024.

[5] Kaito Baba, Chaoran Liu, Shuhei Kurita, and Akiyoshi Sannai. Prover Agent: An agent-based framework for formal mathematical proofs. *arXiv preprint arXiv:2506.19923*, 2025.

[6] Satanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72. Association for Computational Linguistics, 2005.

[7] Shruthi Bannur, Kenza Bouzid, Daniel C. Castro, Anton Schwaighofer, Anja Thieme, Sam Bond-Taylor, Maximilian Ilse, Fernando Pérez-García, Valentina Salvatelli, Harshita Sharma, Felix Meissen, Mercy Ranjit, Shaury Srivastav, Julia Gong, Noel C. F. Codella, Fabian Falck, Ozan Oktay, Matthew P. Lungren, Maria Teodora Wetscherek, Javier Alvarez-Valle, and Stephanie L. Hyland. MAIRA-2: Grounded radiology report generation. *arXiv preprint arXiv:2406.04449*, 2024.

[8] Xiaolei Bo, Feiyang Yang, Feilong Xu, and Xiaoli Zhang. Cross-counter-repeat attention for enhanced understanding of visual semantics in radiology report generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 4242–4250. Association for Computing Machinery, 2025.

[9] William Boag, Tzu-Ming Harry Hsu, Matthew Mcdermott, Gabriela Berner, Emily Alesentzer, and Peter Szolovits. Baselines for chest X-ray report generation. In *Proceedings of the Machine Learning for Health NeurIPS Workshop*, pages 126–140. PMLR, 2020.

[10] G. W. L. Boland, A. S. Guimaraes, and P. R. Mueller. Radiology report turnaround: expectations and solutions. *European Radiology*, 18(7):1326–1328, 2008.

[11] Joshua Broder. Imaging the chest: The chest radiograph. In *Diagnostic Imaging for the Emergency Physician*, pages 185–296. Elsevier, 2011.

[12] Sema Candemir and Sameer Antani. A review on lung boundary detection in chest X-rays. *International Journal of Computer Assisted Radiology and Surgery*, 14(4):563–576, 2019.

[13] Boyuan Chen, Zhuo Xu, Sean Kirmani, Brain Ichter, Dorsa Sadigh, Leonidas Guibas, and Fei Xia. Spatialvlm: Endowing vision-language models with spatial reasoning capabilities. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14455–14465, 2024.

[14] Mingyang Chen, Linzhuang Sun, Tianpeng Li, Haoze Sum, Yijie Zhou, Chenzheng Zhu, Haofen Wang, Jeff Z. Pan, Wen Zhang, Huajun Chen, Fan Yang, Zenan Zhou, and Weipeng Chen. ReSearch: Learning to reason with search for LLMs via reinforcement learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[15] Zhihong Chen, Yan Song, Tsung-Hui Chang, and Xiang Wan. Generating radiology reports via memory-driven transformer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1439–1449. Association for Computational Linguistics, 2020.

[16] Zhihong Chen, Yaling Shen, Yan Song, and Xiang Wan. Cross-modal memory networks for radiology report generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5904–5914. Association for Computational Linguistics, 2021.

[17] Zhihong Chen, Maya Varma, Jean-Benoit Delbrouck, Magdalini Paschali, Louis Blankemeier, Dave Van Veen, Jeya Maria Jose Valanarasu, Alaa Youssef, Joseph Paul Cohen, Eduardo Pontes Reis, Emily Tsai, Andrew Johnston, Cameron Olsen, Tanishq Mathew Abraham, Sergios Gatidis, Akshay S Chaudhari, and Curtis Langlotz. CheXagent: Towards a foundation model for chest X-ray interpretation. In *AAAI 2024 Spring Symposium on Clinical Foundation Models*, 2024.

[18] Zhuoxiao Chen, Hongyang Yu, Ying Xu, Yadan Luo, Long Duong, and Yuan-Fang Li. OraPO: Oracle-educated reinforcement learning for data-efficient and factual radiology report generation. *arXiv preprint arXiv:2509.18600*, 2025.

[19] An-Chieh Cheng, Hongxu Yin, Yang Fu, Qiushan Guo, Ruihan Yang, Jan Kautz, Xiaolong Wang, and Sifei Liu. SpatialRGPT: Grounded spatial reasoning in vision-language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[20] Noel C. F. Codella, Ying Jin, Shrey Jain, Yu Gu, Ho Hin Lee, Asma Ben Abacha, Alberto Santamaria-Pang, Will Guyman, Naiteek Sangani, Sheng Zhang, Hoifung Poon, Stephanie Hyland, Shruthi Bannur, Javier Alvarez-Valle, Xue Li, John Garrett, Alan McMillan, Gaurav Rajguru, Madhu Maddi, Nilesh Vijayrania, Rehaan Bhimai, Nick Mecklenburg, Rupal Jain, Daniel Holstein, Naveen Gaur, Vijay Aski, Jenq-Neng Hwang, Thomas Lin, Ivan Tarapov, Matthew Lungren, and Mu Wei. MedImageInsight: An open-source embedding model for general domain medical imaging. *arXiv preprint arXiv:2410.06542*, 2024.

[21] Ian A. Cowan, Sharyn L. S. MacDonald, and Richard A. Floyd. Measuring and managing radiologist workload: measuring radiologist reporting times using data from a radiology information system. *Journal of Medical Imaging and Radiation Oncology*, 57(5):558–566, 2013.

[22] Daniel Coelho de Castro, Aurelia Bustos, Shruthi Bannur, Stephanie L. Hyland, Kenza Bouzid, Maria Teodora Wetscherek, Maria Dolores Sánchez-Valverde, Lara Jaques-Pérez, Lourdes Pérez-Rodríguez, Kenji Takeda, José María Salinas-Serrano, Javier Alvarez-Valle, Joaquín Galant-Herrero, and Antonio Pertusa. PadChest-GR: A bilingual chest X-ray dataset for grounded radiology report generation. *NEJM AI*, 2(7):AIdbp2401120, 2025.

[23] DeepSeek-AI. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[24] Jean-Benoit Delbrouck, Justin Xu, Johannes Moll, Alois Thomas, Zhihong Chen, Sophie Ostmeier, Asfandyar Azhar, Kelvin Zhenghao Li, Andrew Johnston, Christian Bluethgen, Eduardo Pontes Reis, Mohamed S Muneer, Maya Varma, and Curtis Langlotz. Automated structured radiology report generation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26813–26829. Association for Computational Linguistics, 2025.

[25] Dina Demner-Fushman, Marc D. Kohli, Marc B. Rosenman, Sonya E. Shooshan, Laritza Rodriguez, Sameer Antani, George R. Thoma, and Clement J. McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2015.

[26] Fei Dong, Shouping Nie, Manling Chen, Fangfang Xu, and Qian Li. Keyword-based ai assistance in the generation of radiology reports: A pilot study. *npj Digital Medicine*, 8(1): 490, 2025.

[27] Guanting Dong, Yifei Chen, Xiaoxi Li, Jiajie Jin, Hongjin Qian, Yutao Zhu, Hangyu Mao, Guorui Zhou, Zhicheng Dou, and Ji-Rong Wen. Tool-Star: Empowering LLM-brained multi-tool reasoner via reinforcement learning. *arXiv preprint arXiv:2505.16410*, 2025.

[28] Ahmed T. Elboardy, Ghada Khoriba, and Essam A. Rashed. Medical AI consensus: A multi-agent framework for radiology report generation and evaluation. *arXiv preprint arXiv:2509.17353*, 2025.

[29] Jiazhan Feng, Shijue Huang, Xingwei Qu, Ge Zhang, Yujia Qin, Baoquan Zhong, Chengquan Jiang, Jinxin Chi, and Wanjun Zhong. ReTool: Reinforcement learning for strategic tool use in LLMs, 2025.

[30] Anna Fink, Alexander Rau, Marco Reisert, Fabian Bamberg, and Maximilian F. Russe. Retrieval-augmented generation with large language models in radiology: From theory to practice. *Radiology: Artificial Intelligence*, 7(4):e240790, 2025.

[31] Google. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

[32] Alice Heiman, Xiaoman Zhang, Emma Chen, Sung Eun Kim, and Pranav Rajpurkar. FactCheXcker: Mitigating measurement hallucinations in chest X-ray report generation models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 30787–30796, 2025.

[33] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. MetaGPT: Meta programming for a multi-agent collaborative framework. In *The Twelfth International Conference on Learning Representations*, 2024.

[34] Wenjun Hou, Yi Cheng, Kaishuai Xu, Heng Li, Yan Hu, Wenjie Li, and Jiang Liu. RADAR: Enhancing radiology report generation with supplementary knowledge injection. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26366–26381. Association for Computational Linguistics, 2025.

[35] Xiaodi Hou, Xiaobo Li, Mingyu Lu, Simiao Wang, and Yijia Zhang. RRG-Mamba: Efficient radiology report generation with state space model. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pages 7410–7418. International Joint Conferences on Artificial Intelligence Organization, 2025.

[36] Xuege Hou, Yali Li, and Shengjin Wang. Knowledge-driven query network with adaptive cross-view attention for structured radiology report generation. In *IEEE/CVF International Conference on Computer Vision Workshops*, pages 1234–1243, 2025.

[37] Mengkang Hu, Yuhang Zhou, Wendong Fan, Yuzhou Nie, Ziyu Ye, Bowei Xia, Tao Sun, Zhaoxuan Jin, Yingru Li, Zeyu Zhang, Yifeng Wang, Qianshuo Ye, Bernard Ghanem, Ping Luo, and Guohao Li. OWL: Optimized workforce learning for general multi-agent assistance in real-world task automation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[38] Shih-Cheng Huang, Liyue Shen, Matthew P. Lungren, and Serena Yeung. GLoRIA: A multimodal global-local representation learning framework for label-efficient medical image recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3922–3931, 2021.

[39] Xiaofei Huang, Wenting Chen, Jie Liu, Qisheng Lu, Xiaoling Luo, and Linlin Shen. DAMPER: A dual-stage medical report generation framework with coarse-grained mesh alignment and fine-grained hypergraph matching. *AAAI Conference on Artificial Intelligence*, 39(4):3769–3778, 2025.

[40] Xiyang Huang, Yingjie Han, Yx L, Runzhi Li, Pengcheng Wu, and Kunli Zhang. CmEAA: Cross-modal enhancement and alignment adapter for radiology report generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8546–8556. Association for Computational Linguistics, 2025.

[41] Zhongzhen Huang, Xiaofan Zhang, and Shaoting Zhang. Kiut: Knowledge-injected u-transformer for radiology report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19809–19818, 2023.

[42] Stephanie L. Hyland, Shruthi Bannur, Kenza Bouzid, Daniel C. Castro, Mercy Ranjit, Anton Schwaighofer, Fernando Pérez-García, Valentina Salvatelli, Shaury Srivastav, Anja Thieme, Noel Codella, Matthew P. Lungren, Maria Teodora Wetscherek, Ozan Oktay, and Javier Alvarez-Valle. MAIRA-1: A specialised large multimodal model for radiology report generation. *arXiv preprint arXiv:2311.13668*, 2024.

[43] Saahil Jain, Ashwin Agrawal, Adriel Saporta, Steven Truong, Du Nguyen Duong, Tan Bui, Pierre Chambon, Yuhao Zhang, Matthew Lungren, Andrew Ng, Curtis Langlotz, Pranav Rajpurkar, and Pranav Rajpurkar. RadGraph: Extracting clinical entities and relations from radiology reports. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.

[44] Feibo Jiang, Cunhua Pan, Li Dong, Kezhi Wang, Octavia A. Dobre, and Merouane Debbah. From large AI models to agentic AI: A tutorial on future intelligent communications. *arXiv preprint arXiv:2505.22311*, 2025.

[45] Hanqi Jiang, Xixuan Hao, Yuzhou Huang, Chong Ma, Jiaxun Zhang, Yi Pan, and Ruimao Zhang. Advancing medical radiograph representation learning: A hybrid pre-training paradigm with multilevel semantic granularity. In *European Conference on Computer Vision Workshops*, pages 16–33, 2025.

[46] Yue Jiang, Jiawei Chen, Dingkang Yang, Mingcheng Li, Shunli Wang, Tong Wu, Ke Li, and Lihua Zhang. CoMT: Chain-of-medical-thought reduces hallucination in medical report generation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.

[47] Peiyuan Jing, Kinhei Lee, Zhenxuan Zhang, Huichi Zhou, Zhengqing Yuan, Zhifan Gao, Lei Zhu, Giorgos Papanastasiou, Yingying Fang, and Guang Yang. Reason like a radiologist: Chain-of-thought and reinforcement learning for verifiable report generation. *arXiv preprint arXiv:2504.18453*, 2025.

[48] Alistair E. W. Johnson, Tom J. Pollard, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren, Chih-ying Deng, Roger G. Mark, and Steven Horng. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*, 6(1): 317, 2019.

[49] Hamza Kalisch, Fabian Hörst, Jens Kleesiek, Ken Herrmann, and Constantin Seibold. CT-GRAPH: Hierarchical graph attention network for anatomy-guided CT report generation. *arXiv preprint arXiv:2508.05375*, 2025.

[50] Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik Siu Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae Won Park. MDAgents: An adaptive collaboration of LLMs for medical decision-making. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[51] Yunsoo Kim, Jinge Wu, Su Hwan Kim, Pardeep Vasudev, Jiashu Shen, and Honghan Wu. Look & mark: Leveraging radiologist eye fixations and bounding boxes in multimodal large language models for chest X-ray report generation. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 17680–17694. Association for Computational Linguistics, 2025.

[52] Anis Koubaa. From pre-trained language models to agentic AI: Evolution and architectures for autonomous intelligence. *Preprints*, 2025.

[53] Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th Symposium on Operating Systems Principles*, pages 611–626. Association for Computing Machinery, 2023.

[54] Yuxiang Lai, Jike Zhong, Ming Li, Shitian Zhao, Yuheng Li, Konstantinos Psounis, and Xiaofeng Yang. Med-R1: Reinforcement learning for generalizable medical reasoning in vision-language models. *arXiv preprint arXiv:2503.13939*, 2025.

[55] Kyeongkyu Lee, Seonghwan Yoon, and Hongki Lim. CLARIFID: Improving radiology report generation by reinforcing clinically accurate impressions and enforcing detailed findings. *arXiv preprint arXiv:2507.17234*, 2025.

[56] Seowoo Lee, Jiwon Youn, Hyungjin Kim, Mansu Kim, and Soon Ho Yoon. CXR-LLaVA: a multimodal large language model for interpreting chest X-ray images. *European Radiology*, 35(7):4374–4386, 2025.

[57] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. CAMEL: Communicative agents for "mind" exploration of large language model society. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[58] Mingjie Li, Haokun Lin, Liang Qiu, Xiaodan Liang, Ling Chen, Abdulmotaleb Elsaddik, and Xiaojun Chang. Contrastive learning with counterfactual explanations for radiology report generation. In *European Conference on Computer Vision*, pages 162–180, 2024.

[59] Xinyi Li, Sai Wang, Siqi Zeng, Yu Wu, and Yi Yang. A survey on llm-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth*, 1(1):9, 2024.

[60] Xuefeng Li, Haoyang Zou, and Pengfei Liu. ToRL: Scaling tool-integrated RL. *arXiv preprint arXiv:2503.23383*, 2025.

[61] Yingshu Li, Yunyi Liu, Zhanyu Wang, Xinyu Liang, Lingqiao Liu, Lei Wang, and Luping Zhou. S-RRG-Bench: Structured radiology report generation with fine-grained evaluation framework. *Meta-Radiology*, page 100171, 2025.

[62] Zhuofeng Li, Haoxiang Zhang, Seungju Han, Sheng Liu, Jianwen Xie, Yu Zhang, Yejin Choi, James Zou, and Pan Lu. In-the-flow agentic system optimization for effective planning and tool use. *arXiv preprint arXiv:2510.05592*, 2025.

[63] Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17889–17904. Association for Computational Linguistics, 2024.

[64] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81. Association for Computational Linguistics, 2004.

[65] Qika Lin, Yifan Zhu, Bin Pu, Ling Huang, Haoran Luo, Jingying Ma, Zhen Peng, Tianzhe Zhao, Fangzhi Xu, Jian Zhang, Kai He, Zhonghong Ou, Swapnil Mishra, and Mengling Feng. A foundation model for chest X-ray interpretation with grounded reasoning via online reinforcement learning. *arXiv preprint arXiv:2509.03906*, 2025.

[66] Chang Liu, Yuanhe Tian, Weidong Chen, Yan Song, and Yongdong Zhang. Bootstrapping large language models for radiology report generation. *AAAI Conference on Artificial Intelligence*, 38(17):18635–18643, 2024.

[67] Fenglin Liu, Shen Ge, and Xian Wu. Competence-based multimodal curriculum learning for medical report generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3001–3012. Association for Computational Linguistics, 2021.

[68] Fenglin Liu, Xian Wu, Shen Ge, Wei Fan, and Yuexian Zou. Exploring and distilling posterior and prior knowledge for radiology report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13748–13757, 2021.

[69] Guanxiong Liu, Tzu-Ming Harry Hsu, Matthew McDermott, Willie Boag, Wei-Hung Weng, Peter Szolovits, and Marzyeh Ghassemi. Clinically accurate chest X-ray report generation. In *Proceedings of the 4th Machine Learning for Healthcare Conference*, pages 249–269. PMLR, 2019.

[70] Kang Liu, Zhuoqi Ma, Xiaolu Kang, Zhusi Zhong, Zhicheng Jiao, Grayson Baird, Harrison Bai, and Qiguang Miao. Structural entities extraction and patient indications incorporation for chest X-ray report generation. In *proceedings of Medical Image Computing and Computer Assisted Intervention*. Springer Nature Switzerland, 2024.

[71] Kang Liu, Zhuoqi Ma, Xiaolu Kang, Yunan Li, Kun Xie, Zhicheng Jiao, and Qiguang Miao. Enhanced contrastive learning with multi-view longitudinal data for chest X-ray report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10348–10359, 2025.

[72] Rui Liu, Mingjie Li, Shen Zhao, Ling Chen, Xiaojun Chang, and Lina Yao. In-context learning for zero-shot medical report generation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 8721–8730. Association for Computing Machinery, 2024.

[73] Tengfei Liu, Jiapu Wang, Yongli Hu, Mingjie Li, Junfei Yi, Xiaojun Chang, Junbin Gao, and Baocai Yin. HC-LLM: Historical-constrained large language models for radiology report generation. In *AAAI Conference on Artificial Intelligence*, pages 5595–5603, 2025.

[74] Xiaohong Liu, Hao Liu, Guoxing Yang, Zeyu Jiang, Shuguang Cui, Zhaoze Zhang, Huan Wang, Liyuan Tao, Yongchang Sun, Zhu Song, Tianpei Hong, Jin Yang, Tianrun Gao, Jiangjiang Zhang, Xiaohu Li, Jing Zhang, Ye Sang, Zhao Yang, Kanmin Xue, Song Wu, Ping Zhang, Jian Yang, Chunli Song, and Guangyu Wang. A generalist medical language model for disease diagnosis assistance. *Nature Medicine*, 31(3):932–942, 2025.

[75] Zhizhe Liu, Zhenfeng Zhu, Shuai Zheng, Yawei Zhao, Kunlun He, and Yao Zhao. From observation to concept: A flexible multi-view paradigm for medical report generation. *IEEE Transactions on Multimedia*, 26:5987–5995, 2024.

[76] Zichen Liu, Changyu Chen, Wenjun Li, Penghui Qi, Tianyu Pang, Chao Du, Wee Sun Lee, and Min Lin. Understanding R1-Zero-like training: A critical perspective. In *Second Conference on Language Modeling*, 2025.

[77] Zihe Liu, Jiashun Liu, Yancheng He, Weixun Wang, Jiaheng Liu, Ling Pan, Xinyu Hu, Shaopan Xiong, Ju Huang, Jian Hu, Shengyi Huang, Johan Obando-Ceron, Siran Yang, Jiamang Wang, Wenbo Su, and Bo Zheng. Part I: Tricks or traps? a deep dive into RL for LLM reasoning. *arXiv preprint arXiv:2508.08221*, 2025.

[78] Jinhui Lou, Yan Yang, Zhou Yu, Zhenqi Fu, Weidong Han, Qingming Huang, and Jun Yu. CXRAgent: Director-orchestrated multi-stage reasoning for chest X-ray interpretation. *arXiv preprint arXiv:2510.21324*, 2025.

[79] Chong Ma, Hanqi Jiang, Wenting Chen, Yiwei Li, Zihao Wu, Xiaowei Yu, Zhengliang Liu, Lei Guo, Dajiang Zhu, Tuo Zhang, Dinggang Shen, Tianming Liu, and Xiang Li. Eye-gaze guided multi-modal alignment for medical representation learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[80] Xinji Mai, Haotian Xu, Zhong-Zhi Li, Xing W, Weinong Wang, Jian Hu, Yingying Zhang, and Wenqiang Zhang. Agent RL scaling law: Agent rl with spontaneous code execution for mathematical problem solving. *arXiv preprint arXiv:2505.07773*, 2025.

[81] Joshua P. Metlay, Grant W. Waterer, Ann C. Long, Antonio Anzueto, Jan Brozek, Kristina Crothers, Laura A. Cooley, Nathan C. Dean, Michael J. Fine, Scott A. Flanders, Marie R. Griffin, Mark L. Metersky, Daniel M. Musher, Marcos I. Restrepo, and Cynthia G. Whitney. Diagnosis and treatment of adults with community-acquired pneumonia. an official clinical practice guideline of the american thoracic society and infectious diseases society of america. *American Journal of Respiratory and Critical Care Medicine*, 200(7):e45–e67, 2019.

[82] Sumeet Ramesh Motwani, Chandler Smith, Rocktim Jyoti Das, Rafael Rafailov, Philip Torr, Ivan Laptev, Fabio Pizzati, Ronald Clark, and Christian Schroeder de Witt. MALT: Improving reasoning with multi-agent LLM training. In *Second Conference on Language Modeling*, 2025.

[83] Théo Moutakanni, Piotr Bojanowski, Guillaume Chassagnon, Céline Hudelot, Armand Joulin, Yann LeCun, Matthew Muckley, Maxime Oquab, Marie-Pierre Revel, and Maria Vakalopoulou. Advancing human-centric ai for robust X-ray analysis through holistic self-supervised learning. *arXiv preprint arXiv:2405.01469*, 2024.

[84] Vishwesh Nath, Wenqi Li, Dong Yang, Andriy Myronenko, Mingxin Zheng, Yao Lu, Zhijian Liu, Hongxu Yin, Yee Man Law, Yucheng Tang, Pengfei Guo, Can Zhao, Ziyue Xu, Yufan He, Stephanie Harmon, Benjamin Simon, Greg Heinrich, Stephen Aylward, Marc Edgar, Michael Zephyr, Pavlo Molchanov, Baris Turkbey, Holger Roth, and Daguang Xu. VILA-M3: Enhancing vision-language models with medical expert knowledge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14788–14798, 2025.

[85] Aaron Nicolson, Jason Dowling, Douglas Anderson, and Bevan Koopman. Longitudinal data and a semantic similarity reward for chest X-ray report generation. *Informatics in Medicine Unlocked*, 50:101585, 2024.

[86] OpenAI. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*, 2024.

[87] Sophie Ostmeier, Justin Xu, Zhihong Chen, Maya Varma, Louis Blankemeier, Christian Bluethgen, Arne Edward Michalson Md, Michael Moseley, Curtis Langlotz, Akshay S Chaudhari, and Jean-Benoit Delbrouck. GREEN: Generative radiology report evaluation and error notation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 374–390. Association for Computational Linguistics, 2024.

[88] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

[89] Chanwoo Park, Seungju Han, Xingzhi Guo, Asuman E. Ozdaglar, Kaiqing Zhang, and Joo-Kyung Kim. MAPoRL: Multi-agent post-co-training for collaborative large language models with reinforcement learning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 30215–30248. Association for Computational Linguistics, 2025.

[90] Sang-Jun Park, Keun-Soo Heo, Dong-Hee Shin, Young-Han Son, Ji-Hye Oh, and Tae-Eui Kam. DART: Disease-aware image-text alignment and self-correcting re-alignment for trustworthy radiology report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15580–15589, 2025.

[91] Chantal Pellegrini, Ege Özsoy, Benjamin Busam, Nassir Navab, and Matthias Keicher. RaDialog: A large vision-language model for radiology report generation and conversational assistance. *arXiv preprint arXiv:2311.18681*, 2025.

[92] Aske Plaat, Max van Duijn, Niki van Stein, Mike Preuss, Peter van der Putten, and Kees Joost Batenburg. Agentic large language models, a survey. *arXiv preprint arXiv:2503.23037*, 2025.

[93] Cheng Qian, Emre Can Acikgoz, Qi He, Hongru Wang, Xiusi Chen, Dilek Hakkani-Tür, Gokhan Tur, and Heng Ji. ToolRL: Reward is all tool learning needs. *arXiv preprint arXiv:2504.13958*, 2025.

[94] Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cían Hughes, Charles Lau, Justin Chen, Fereshteh Mahvar, Liron Yatziv, Tiffany Chen, Bram Sterling, Stefanie Anna Baby, Susanna Maria Baby, Jeremy Lai, Samuel Schmidgall, Lu Yang, Kejia Chen, Per Bjornsson, Shashir Reddy, Ryan Brush, Kenneth Philbrick, Mercy Asiedu, Ines Mezerreg, Howard Hu, Howard Yang, Richa Tiwari, Sunny Jansen, Preeti Singh, Yun Liu, Shekoofeh Azizi, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova,

Alexandre Ramé, Morgane Riviere, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Elena Buchatskaya, Jean-Baptiste Alayrac, Dmitry Lepikhin, Vlad Feinberg, Sebastian Borgeaud, Alek Andreev, Cassidy Hardin, Robert Dadashi, Léonard Hussenot, Armand Joulin, Olivier Bachem, Yossi Matias, Katherine Chou, Avinatan Hassidim, Kavi Goel, Clement Farabet, Joelle Barral, Tris Warkentin, Jonathon Shlens, David Fleet, Victor Cotruta, Omar Sanseviero, Gus Martins, Phoebe Kirk, Anand Rao, Shravya Shetty, David F. Steiner, Can Kirmizibayrak, Rory Pilgrim, Daniel Golden, and Lin Yang. MedGemma technical report. *arXiv preprint arXiv:2507.05201*, 2025.

[95] Vaibhavi Shah, Yeshwant R. Chillakuru, Alex Rybkin, Youngho Seo, Thienkhai Vu, and Jae Ho Sohn. Algorithmic prediction of delayed radiology turn-around-time during non-business hours. *Academic Radiology*, 29(5):e82–e90, 2022.

[96] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[97] Hongyu Shen, Mingtao Pei, Juncai Liu, and Zhaoxing Tian. Automatic radiology reports generation via memory alignment network. *AAAI Conference on Artificial Intelligence*, 38(5): 4776–4783, 2024.

[98] Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. HybridFlow: A flexible and efficient rlhf framework. In *Proceedings of the Twentieth European Conference on Computer Systems*, pages 1279–1297. Association for Computing Machinery, 2025.

[99] Joykirat Singh, Raghav Magazine, Yash Pandya, and Akshay Nambi. Agentic reasoning and tool integration for LLMs via reinforcement learning. *arXiv preprint arXiv:2505.01441*, 2025.

[100] Akshay Smit, Saahil Jain, Pranav Rajpurkar, Anuj Pareek, Andrew Ng, and Matthew Lungren. Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. In *Empirical Methods in Natural Language Processing*, pages 1500–1519. Association for Computational Linguistics, 2020.

[101] Huatong Song, Jinhao Jiang, Yingqian Min, Jie Chen, Zhipeng Chen, Wayne Xin Zhao, Lei Fang, and Ji-Rong Wen. R1-Searcher: Incentivizing the search capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2503.05592*, 2025.

[102] Yi Su, Dian Yu, Linfeng Song, Juntao Li, Haitao Mi, Zhaopeng Tu, Min Zhang, and Dong Yu. Crossing the reward bridge: Expanding rl with verifiable rewards across diverse domains. *arXiv preprint arXiv:2503.23829*, 2025.

[103] Iustin Sîrbu, Iulia-Renata Sîrbu, Jasmina Bogojeska, and Traian Rebedea. GIT-CXR: End-to-end transformer for chest X-ray report generation. *Information*, 16(7), 2025.

[104] Tim Tanida, Philip Müller, Georgios Kaissis, and Daniel Rueckert. Interactive and explainable region-guided radiology report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7433–7442, 2023.

[105] Ryutaro Tanno, David G. T. Barrett, Andrew Sellergren, Sumedh Ghaisas, Sumanth Dathathri, Abigail See, Johannes Welbl, Charles Lau, Tao Tu, Shekoofeh Azizi, Karan Singhal, Mike Schaekermann, Rhys May, Roy Lee, SiWai Man, Sara Mahdavi, Zahra Ahmed, Yossi Matias, Joelle Barral, S. M. Ali Eslami, Danielle Belgrave, Yun Liu, Sreenivasa Raju Kalidindi, Shravya Shetty, Vivek Natarajan, Pushmeet Kohli, Po-Sen Huang, Alan Karthikesalingam, and Ira Ktena. Collaboration between clinicians and vision–language models in radiology report generation. *Nature Medicine*, 31(2):599–608, 2025.

[106] Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, Gaël Liu, Francesco Visin, Kathleen Kenealy, Lucas Beyer, Xiaohai Zhai, Anton Tsitsulin, Robert Busa-Fekete, Alex Feng, Noveen Sachdeva,

Benjamin Coleman, Yi Gao, Basil Mustafa, Iain Barr, Emilio Parisotto, David Tian, Matan Eyal, Colin Cherry, Jan-Thorsten Peter, Danila Sinopalnikov, Surya Bhupatiraju, Rishabh Agarwal, Mehran Kazemi, Dan Malkin, Ravin Kumar, David Vilar, Idan Brusilovsky, Jiaming Luo, Andreas Steiner, Abe Friesen, Abhanshu Sharma, Abheesht Sharma, Adi Mayrav Gilady, Adrian Goedeckemeyer, Alaa Saade, Alex Feng, Alexander Kolesnikov, Alexei Bendebury, Alvin Abdagic, Amit Vadi, András György, André Susano Pinto, Anil Das, Ankur Bapna, Antoine Miech, Antoine Yang, Antonia Paterson, Ashish Shenoy, Ayan Chakrabarti, Bilal Piot, Bo Wu, Bobak Shahriari, Bryce Petrini, Charlie Chen, Charline Le Lan, Christopher A. Choquette-Choo, CJ Carey, Cormac Brick, Daniel Deutsch, Danielle Eisenbud, Dee Cattle, Derek Cheng, Dimitris Paparas, Divyashree Shivakumar Sreepathihalli, Doug Reid, Dustin Tran, Dustin Zelle, Eric Noland, Erwin Huizenga, Eugene Kharitonov, Frederick Liu, Gagik Amirkhanyan, Glenn Cameron, Hadi Hashemi, Hanna Klimczak-Plucińska, Harman Singh, Harsh Mehta, Harshal Tushar Lehri, Hussein Hazimeh, Ian Ballantyne, Idan Szpektor, Ivan Nardini, Jean Pouget-Abadie, Jetha Chan, Joe Stanton, John Wieting, Jonathan Lai, Jordi Orbay, Joseph Fernandez, Josh Newlan, Ju yeong Ji, Jyotinder Singh, Kat Black, Kathy Yu, Kevin Hui, Kiran Vodrahalli, Klaus Greff, Linhai Qiu, Marcella Valentine, Marina Coelho, Marvin Ritter, Matt Hoffman, Matthew Watson, Mayank Chaturvedi, Michael Moynihan, Min Ma, Nabila Babar, Natasha Noy, Nathan Byrd, Nick Roy, Nikola Momchev, Nilay Chauhan, Noveen Sachdeva, Oskar Bunyan, Pankil Botarda, Paul Caron, Paul Kishan Rubenstein, Phil Culliton, Philipp Schmid, Pier Giuseppe Sessa, Pingmei Xu, Piotr Stanczyk, Pouya Tafti, Rakesh Shivanna, Renjie Wu, Renke Pan, Reza Rokni, Rob Willoughby, Rohith Vallu, Ryan Mullins, Sammy Jerome, Sara Smoot, Sertan Girgin, Shariq Iqbal, Shashir Reddy, Shruti Sheth, Siim Põder, Sijal Bhatnagar, Sindhu Raghuram Panyam, Sivan Eiger, Susan Zhang, Tianqi Liu, Trevor Yacovone, Tyler Liechty, Uday Kalra, Utku Evci, Vedant Misra, Vincent Roseberry, Vlad Feinberg, Vlad Kolesnikov, Woohyun Han, Woosuk Kwon, Xi Chen, Yinlam Chow, Yuvein Zhu, Zichuan Wei, Zoltan Egyed, Victor Cotruta, Minh Giang, Phoebe Kirk, Anand Rao, Kat Black, Nabila Babar, Jessica Lo, Erica Moreira, Luiz Gustavo Martins, Omar Sanseviero, Lucas Gonzalez, Zach Gleicher, Tris Warkentin, Vahab Mirrokni, Evan Senter, Eli Collins, Joelle Barral, Zoubin Ghahramani, Raia Hadsell, Yossi Matias, D. Sculley, Slav Petrov, Noah Fiedel, Noam Shazeer, Oriol Vinyals, Jeff Dean, Demis Hassabis, Koray Kavukcuoglu, Clement Farabet, Elena Buchatskaya, Jean-Baptiste Alayrac, Rohan Anil, Dmitry, Lepikhin, Sebastian Borgeaud, Olivier Bachem, Armand Joulin, Alek Andreev, Cassidy Hardin, Robert Dadashi, and Léonard Hussenot. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025.

[107] Omkar Chakradhar Thawakar, Abdelrahman M. Shaker, Sahal Shaji Mullappilly, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, Jorma Laaksonen, and Fahad Khan. XrayGPT: Chest radiographs summarization using large medical vision-language models. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 440–448. Association for Computational Linguistics, 2024.

[108] Aowen Wang, Zhiwang Zhang, Dongang Wang, Fanyi Wang, Haotian Hu, Jinyang Guo, Yipeng Zhou, Chaoyi Pang, and Shiting Wen. Overcoming heterogeneous data in federated medical vision-language pre-training: A triple-embedding model selector approach. *AAAI Conference on Artificial Intelligence*, 39(7):7500–7508, 2025.

[109] Fuying Wang, Shenghui Du, and Lequan Yu. HERGen: Elevating radiology report generation with longitudinal data. In *European Conference on Computer Vision*, pages 183–200. Springer Nature Switzerland, 2025.

[110] Lehan Wang, Haonan Wang, Honglong Yang, Jiaji Mao, Zehong Yang, Jun Shen, and Xiaomeng Li. Interpretable bilingual multimodal large language model for diverse biomedical tasks. In *The Thirteenth International Conference on Learning Representations*, 2025.

[111] Siyu Wang, Cailian Chen, Xinyi Le, Qimin Xu, Lei Xu, Yanzhou Zhang, and Jie Yang. CAD-GPT: Synthesising cad construction sequence with spatial reasoning-enhanced multimodal llms. *AAAI Conference on Artificial Intelligence*, 39(8):7880–7888, 2025.

[112] Xiao Wang, Fuling Wang, Yuehang Li, Qingchuan Ma, Shiao Wang, Bo Jiang, and Jin Tang. CXPMRG-Bench: Pre-training and benchmarking for X-ray medical report generation on chexpert plus dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5123–5133, 2025.

[113] Zhanyu Wang, Lingqiao Liu, Lei Wang, and Luping Zhou. METransformer: Radiology report generation by transformer with multiple learnable expert tokens. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11558–11567, 2023.

[114] Zhanyu Wang, Lingqiao Liu, Lei Wang, and Luping Zhou. R2GenGPT: Radiology report generation with frozen LLMs. *Meta-Radiology*, 1(3):100033, 2023.

[115] Zhichuan Wang, Kinhei Lee, Qiao Deng, Tiffany Y. So, Wan Hang Chiu, Bingjing Zhou, and Edward S. Hui. Expert insight-enhanced follow-up chest X-ray summary generation. In *Artificial Intelligence in Medicine*, pages 181–193. Springer-Verlag, 2024.

[116] Zifeng Wang, Zichen Wang, Balasubramaniam Srinivasan, Vassilis N. Ioannidis, Huzefa Rangwala, and Rishita Anubhai. BioBridge: Bridging biomedical foundation models via knowledge graphs. In *International Conference on Learning Representations*, 2024.

[117] Ziao Wang, Sixing Yan, Kejing Yin, Xiaofeng Zhang, and William K. Cheung. CURV: Coherent uncertainty-aware reasoning in vision-language models for X-ray report generation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[118] Xumeng Wen, Zihan Liu, Shun Zheng, Shengyu Ye, Zhirong Wu, Yang Wang, Zhijian Xu, Xiao Liang, Junjie Li, Ziming Miao, Jiang Bian, and Mao Yang. Reinforcement learning with verifiable rewards implicitly incentivizes correct reasoning in base LLMs. *arXiv preprint arXiv:2506.14245*, 2025.

[119] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. AutoGen: Enabling next-gen LLM applications via multi-agent conversations. In *First Conference on Language Modeling*, 2024.

[120] Ting Xiao, Lei Shi, Peng Liu, Zhe Wang, and Chenjia Bai. Radiology report generation via multi-objective preference optimization. In *AAAI Conference on Artificial Intelligence*, pages 8664–8672, 2025.

[121] Ting Xiao, Lei Shi, Yang Zhang, HaoFeng Yang, Zhe Wang, and Chenjia Bai. Online iterative self-alignment for radiology report generation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 27799–27814. Association for Computational Linguistics, 2025.

[122] Benjamin Yan, Ruochen Liu, David Kuo, Subathra Adithan, Eduardo Reis, Stephen Kwak, Vasantha Venugopal, Chloe O'Connell, Agustina Saenz, Pranav Rajpurkar, and Michael Moor. Style-aware radiology report generation with RadGraph and few-shot prompting. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 14676–14688. Association for Computational Linguistics, 2023.

[123] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

[124] Jinxia Yang, Bing Su, Xin Zhao, and Ji-Rong Wen. Unlocking the power of spatial and temporal information in medical multimodal pre-training. In *Proceedings of the 41st International Conference on Machine Learning*, pages 56382–56396. PMLR, 2024.

[125] Longzhen Yang, Zhangkai Ni, Ying Wen, Yihang Liu, Lianghua He, and Heng Tao Shen. Self-supervised anatomical consistency learning for vision-grounded medical report generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 2958–2967. Association for Computing Machinery, 2025.

[126] Ziruo Yi, Ting Xiao, and Mark V. Albert. A multimodal multi-agent framework for radiology report generation. *arXiv preprint arXiv:2505.09787*, 2025.

[127] Heng Yin, Shanlin Zhou, Pandong Wang, Zirui Wu, and Yongtao Hao. KIA: Knowledge-guided implicit vision-language alignment for chest X-ray report generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 4096–4108. Association for Computational Linguistics, 2025.

[128] Kihyun You, Jawook Gu, Jiyeon Ham, Beomhee Park, Jiho Kim, Eun K. Hong, Woonhyuk Baek, and Byungseok Roh. CXR-CLIP: Toward large scale chest X-ray language-image pre-training. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 101–111. Springer Nature Switzerland, 2023.

[129] Feiyang Yu, Mark Endo, Rayan Krishnan, Ian Pan, Andy Tsai, Eduardo Pontes Reis, Eduardo Kaiser Ururahy Nunes Fonseca, Henrique Min Ho Lee, Zahra Shakeri Hossein Abad, Andrew Y. Ng, Curtis P. Langlotz, Vasantha Kumar Venugopal, and Pranav Rajpurkar. Evaluating progress in automatic chest X-ray radiology report generation. *Patterns*, 4(9):100802, 2023.

[130] Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Mu Qiao, Yonghui Wu, and Mingxuan Wang. DAPO: An open-source LLM reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*, 2025.

[131] Juan Manuel Zambrano Chaves, Shih-Cheng Huang, Yanbo Xu, Hanwen Xu, Naoto Usuyama, Sheng Zhang, Fei Wang, Yujia Xie, Mahmoud Khademi, Ziyi Yang, Hany Awadalla, Julia Gong, Houdong Hu, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Yu Gu, Cliff Wong, Mu Wei, Tristan Naumann, Muhao Chen, Matthew P. Lungren, Akshay Chaudhari, Serena Yeung-Levy, Curtis P. Langlotz, Sheng Wang, and Hoifung Poon. A clinically accessible small multimodal radiology model and evaluation metric for chest X-ray findings. *Nature Communications*, 16 (1):3108, 2025.

[132] Junjie Zhang, Jingyi Xi, Zhuoyang Song, Junyu Lu, Yuhua Ke, Ting Sun, Yukun Yang, Jiaxing Zhang, Songxin Zhang, and Zejian Xie. L0: Reinforcement learning to become general agents. *arXiv preprint arXiv:2506.23667*, 2025.

[133] Kai Zhang, Rong Zhou, Eashan Adhikarla, Zhiling Yan, Yixin Liu, Jun Yu, Zhengliang Liu, Xun Chen, Brian D. Davison, Hui Ren, Jing Huang, Chen Chen, Yuyin Zhou, Sunyang Fu, Wei Liu, Tianming Liu, Xiang Li, Yong Chen, Lifang He, James Zou, Quanzheng Li, Hongfang Liu, and Lichao Sun. A generalist vision–language foundation model for diverse biomedical tasks. *Nature Medicine*, 30(11):3129–3141, 2024.

[134] Kai Zhang, Corey D Barrett, Jangwon Kim, Lichao Sun, Tara Taghavi, and Krishnaram Kenthapadi. RadAgents: Multimodal agentic reasoning for chest X-ray interpretation with radiologist-like workflows. *arXiv preprint arXiv:2509.20490*, 2025.

[135] Xi Zhang, Zaiqiao Meng, Jake Lever, and Edmond S. L. Ho. Libra: Leveraging temporal images for biomedical radiology analysis. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 17275–17303. Association for Computational Linguistics, 2025.

[136] Xiaodan Zhang, Yanzhao Shi, Junzhong Ji, Chengxin Zheng, and Liangqiong Qu. MEPNet: Medical entity-balanced prompting network for brain ct report generation. *AAAI Conference on Artificial Intelligence*, 39(24):25940–25948, 2025.

[137] Bingxi Zhao, Lin Geng Foo, Ping Hu, Christian Theobalt, Hossein Rahmani, and Jun Liu. LLM-based agentic reasoning frameworks: A survey from methods to scenarios. *arXiv preprint arXiv:2508.17692*, 2025.

[138] Yanli Zhao, Andrew Gu, Rohan Varma, Liang Luo, Chien-Chin Huang, Min Xu, Less Wright, Hamid Shojanazeri, Myle Ott, Sam Shleifer, Alban Desmaison, Can Balioglu, Pritam Damania, Bernard Nguyen, Geeta Chauhan, Yuchen Hao, Ajit Mathews, and Shen Li. PyTorch FSDP:

Experiences on scaling fully sharded data parallel. *Proceedings of the VLDB Endowment*, 16 (12):3848–3860, 2023.

[139] Chujie Zheng, Shixuan Liu, Mingze Li, Xiong-Hui Chen, Bowen Yu, Chang Gao, Kai Dang, Yuqiong Liu, Rui Men, An Yang, Jingren Zhou, and Junyang Lin. Group sequence policy optimization. *arXiv preprint arXiv:2507.18071*, 2025.

[140] Zijian Zhou, Miaojing Shi, Meng Wei, Oluwatosin Alabi, Zijie Yue, and Tom Vercauteren. Large model driven radiology report generation with clinical quality reinforcement learning. *arXiv preprint arXiv:2403.06728*, 2024.

[141] Yuanhao Zou and Zhaozheng Yin. MVCM: Enhancing multi-view and cross-modality alignment for medical visual question answering and medical image-text retrieval. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 180–190, 2025.

A Extended related work

A.1 Reinforcement learning for LLMs

Reinforcement Learning with Verifiable Rewards (RLVR) has recently emerged as an alternative to traditional Reinforcement Learning from Human Feedback (RLHF), providing objective, outcome-based feedback through deterministic verification functions rather than human preference models. By rewarding correctness or rule-based validity, RLVR has improved reasoning capabilities in structured domains, such as mathematics and code generation [76, 77, 96, 102, 118, 130, 139]. More recently, agentic reinforcement learning has gained attention, where models interact with external tools such as Python interpreters or web search engines to improve factual accuracy [14, 27, 29, 60, 80, 93, 99, 101, 132]. However, most of these methods still focus on training a single, monolithic model, rather than jointly optimizing multiple agents. Consequently, genuine cooperative optimization among agents in real-world workflows remains largely unexplored.

A.2 Agentic systems with LLMs

Agentic systems leverage LLMs to perform goal-oriented reasoning, planning, and collaboration through structured interactions among multiple agents. Recent frameworks such as AutoGen [119], MetaGPT [33], CAMEL [57], have shown that pretrained LLMs can engage in cooperative behaviors via carefully designed prompts and workflows [5, 37, 44, 52, 59, 63, 92, 137]. However, most of these systems are training-free, relying on pre-trained models without end-to-end optimization, resulting in suboptimal coordination and limited adaptability when applied to complex, real-world workflows.

Recent works have started to explore reinforcement learning within agentic systems. For example, MALT [82] performs post-training using trajectories collected from agent executions, but its optimization remains off-policy and detached from real-world workflow interactions. MAPoRL [89] applies multi-agent reinforcement learning to improve collaboration among language models, yet the optimization of realistic, role-structured workflows in applied domains remains largely unexplored. To address this gap, AgentFlow [62] introduces on-policy reinforcement learning within an actual multi-agent workflow, but the optimization is limited to a single key planner agent rather than the entire agentic system.

B Detailed experimental setup

Section 4.1 provides an overview of the experimental setup. In this section, we describe additional details that were not included in Section 4.1.

Datasets MIMIC-CXR [48] is a large-scale publicly available chest X-ray dataset containing more than 370,000 images paired with free-text radiology reports. The dataset covers a wide range of thoracic conditions and includes both frontal and lateral views. IU X-ray (Indiana University Chest X-Ray dataset) [25] consists of chest X-ray images paired with corresponding radiology reports, including both Findings and Impression sections. Each study contains frontal and lateral views with detailed narrative annotations.

Implementation details The reinforcement learning implementation is built on top of verl [98], using vLLM [53] for rollouts and Fully Sharded Data Parallel (FSDP) [138] for parameter updates. We set the batch size to 16, and 16 rollouts are generated for each training sample. The maximum rollout length is set to 2048 tokens. We train each model for 100 reinforcement learning update steps. Following the original GSPO paper [139], we set the clipping thresholds in objectives to $\varepsilon_{\text{high}} = 0.0004$ and $\varepsilon_{\text{low}} = 0.0003$. The rollout temperature is fixed at 1.0. We use a learning rate of 1e-6, a warmup ratio of 0.05, and a weight decay of 0.1. We use AdamW as the optimizer. The batch size for parameter updates is set to 4. We use `google/medgemma-4b-it`[1] as the base checkpoint. We use verl v0.5.0, vLLM v0.10.1, Transformers v4.53.3, and PyTorch v2.7.1. All experiments are conducted on 4 $\times$ NVIDIA H100 GPUs, each with 80 GB of memory. Each training run took approximately 18 hours. For inference, we follow the default vLLM generation settings except that the maximum number of generated tokens is set to 2048.

C Computational cost and deployment considerations

To assess the practical deployment cost of MARL-Rad, we measured inference latency and GPU memory usage on H100 GPUs using 600 reports from the MIMIC-CXR test set. When the three region-specific agents were executed in parallel, MARL-Rad required 772 seconds in total, corresponding to 1.28 seconds per sample. When all agents were executed sequentially, the total inference time was 1586 seconds, corresponding to 2.64 seconds per sample. This latency is identical to that of the training-free agentic baseline used in our ablation study, i.e., "MedGemma (agent)," because both methods share the same multi-agent architecture at inference time. For reference, the non-agent baseline, "MedGemma (vanilla)," required 398 seconds in total, corresponding to 0.66 seconds per sample. The GPU memory usage was approximately 9 GB per agent.

Although MARL-Rad introduces additional computational overhead due to agentization, the region-specific agents can be executed in parallel, substantially reducing latency compared with sequential execution. The resulting inference speed remains within a practical range for real-world radiology reporting workflows.

D Details of the laterality-specific RadGraph F1 score computation

To evaluate region-level consistency focused on laterality, we compute a laterality-specific RadGraph F1 score derived from the standard RadGraph evaluation. We first construct the RadGraph for both the prediction and the ground truth using the standard RadGraph extraction procedure. From each constructed graph, we then extract the entities corresponding to laterality-related anatomical regions, which include the tokens "left", "right", "left lung", and "right lung." From these entities, we further extract the subgraph consisting of all nodes and relations connected to them, resulting in a laterality-specific subset of each RadGraph. We then calculate the laterality-specific RadGraph F1 score by applying the standard RadGraph matching procedure but restrict the comparison to this laterality-specific subgraph only.

E Clinician evaluation criteria

Physicians evaluated each report pair according to the following five criteria.

- **Completeness**: Whether the report includes all clinically relevant findings visible in the chest X-ray, without omitting important abnormalities or normal findings that should be documented.
- **Correctness**: Whether the described findings are clinically accurate and consistent with the chest X-ray, without introducing incorrect diagnoses, false findings, or contradictions.
- **Conciseness**: Whether the report is appropriately concise, avoiding unnecessary repetition, irrelevant descriptions, or overly verbose statements while preserving clinically important information.

[1] `https://huggingface.co/google/medgemma-4b-it`

- **Readability**: Whether the report is clearly written, well organized, and easy for clinicians to understand, with coherent phrasing and appropriate radiological terminology.
- **Clinical Utility**: Whether the report would be useful in clinical practice for supporting diagnosis, communication, and downstream decision-making.

Physicians were instructed to compare two anonymized reports for each chest X-ray case and select which report was better for each criterion. The report order was randomized, and the physicians were blinded to whether each report was generated by MARL-Rad or taken from the ground truth. The evaluation was conducted with approval from the affiliated hospital according to the applicable institutional requirements. Potential risks to participants were minimal, as the task involved expert assessment of anonymized reports without identifiable patient information. Physicians performed the evaluation during regular working hours under the affiliated hospital's approval, and no additional honorarium was provided specifically for this task.

F Limitations and broader impact

F.1 Limitations

Although MARL-Rad achieves strong performance on standard RRG benchmarks, this work has several limitations. First, our experiments focus on chest X-ray report generation using MIMIC-CXR and IU X-ray, and the effectiveness of the proposed framework on other imaging modalities, institutions, and clinical settings remains to be validated. Second, our reinforcement learning rewards rely on automatic clinical metrics such as CheXbert and RadGraph. While these metrics are clinically motivated and widely used, they may not fully capture all aspects of report quality and may contain annotation or evaluation noise. We partially address this concern through blinded clinician evaluation, but the evaluation is limited in scale and should be expanded in future work. Third, MARL-Rad requires multiple model invocations at inference time, increasing computational cost and GPU memory usage compared with a single-model baseline, although parallel execution mitigates the added latency as discussed in Appendix C. Finally, although our framework represents a step toward clinically useful agentic RRG systems by improving clinical efficacy metrics and laterality consistency, real-world deployment would require rigorous prospective validation, safety monitoring, and integration into radiologist-supervised workflows.

F.2 Broader impact

MARL-Rad aims to improve radiology report generation by training agentic systems within the workflows in which they are deployed. If properly validated, such systems could help reduce the burden of report writing, improve the consistency of generated reports, and support radiologists in clinical documentation.

At the same time, medical report generation is a high-stakes application. Incorrect or hallucinated findings could negatively affect clinical decision-making if used without appropriate oversight. Future deployment should require careful clinical validation, bias assessment across patient populations and institutions, privacy-preserving data handling, and mechanisms for human review and correction.