Linguistically-Aligned and Visually-Grounded Preference Optimization for Clinically-Augmented Medical Report Generation

Qiang Hu[†], Yuxuan Luo[†], Yingjie Guo, Hao Wang, Qimei Wang, Qiang Li, and Zhiwei Wang[(✉)]

Huazhong University of Science and Technology
{huqiang77,zwwang}@hust.edu.cn

Abstract. Despite significant advances in Medical Report Generation (MRG), the reliability remains constrained by the prevalence of factual errors. While Direct Preference Optimization (DPO) has emerged as a promising post-training paradigm to enhance the performance of Supervised Fine-Tuned (SFT) MRG models, existing DPO-based MRG methods typically adopt a naive preference construction that directly pairs model-generated reports with ground truth reports. This strategy inadvertently entangles critical clinical findings with clinically irrelevant linguistic characteristics, and fundamentally lacks explicit vision-language alignment. To address these challenges, we propose DPO-Clin, a novel post-training framework that focuses preference optimization on clinical findings and cross-modal alignment. First, we introduce the Entity-level Clinical Diagnostic (ECD) module to perform a precise entity-level factual diagnosis. ECD guides the generation of linguistically-aligned report preference pairs, isolating clinical discrepancies from linguistic variations. Second, to achieve fine-grained cross-modal alignment, we develop M^2DPO, a retrieval-augmented multi-modal DPO variant that enforces textual preference inversion triggered by visual context switches. Third, we locate correct yet highly uncertain predicted entities and apply counterfactual modifications to construct targeted preference data for latent risk mitigation, thereby further enhancing the model reliability. Extensive experiments on two public chest X-ray datasets (MIMIC-CXR and IU X-Ray) and an in-house endoscopy dataset demonstrate that DPO-Clin significantly improves the SFT baselines on clinical-aware metrics. Furthermore, it achieves superior performance over existing DPO-based MRG methods, exhibiting robust generalizability across distinct baseline architectures and diverse medical imaging modalities.

Keywords: Medical Report Generation · Direct Preference Optimization · Multi-Modality

[†] Equal contribution; ✉ Corresponding author.

1 Introduction

Automatic Medical Report Generation (MRG) aims to translate medical images into comprehensive textual reports that summarize clinically meaningful findings. By enabling scalable and standardized interpretation of medical imaging data, MRG holds significant potential to improve diagnostic efficiency, reduce clinician workload, and enhance consistency in clinical decision-making.

Early researches in MRG primarily focused on architectural innovations tailored to vision-language modeling. Representative efforts explored multi-modal attention mechanisms [5, 24], region-aware feature localization [6, 33], and the incorporation of structured medical knowledge [21, 45]. Instead of designing specialized architectures from scratch, current state-of-the-art approaches increasingly adapt general-purpose Large Language Models (LLMs) to medical report generation [11, 17, 37]. However, since they are trained on broad-domain data, effective deployment in the medical domain requires domain-specific adaptation.

The most straightforward adaptation strategy is Supervised Fine-Tuning (SFT), where models are optimized to reproduce reference reports using token-level supervision (*e.g.*, cross-entropy loss). Despite its simplicity and technical maturity, SFT fundamentally optimizes imitation rather than clinical preference understanding. It fails to capture why one report is clinically superior to another and struggles when multiple valid descriptions coexist but differ in diagnostic utility, a challenge commonly characterized as the lack of preference alignment.

To address this limitation, recent studies adopt post-training strategies following SFT. A dominant paradigm is Reinforcement Learning from Human Feedback (RLHF) [9, 35, 47], which introduces an auxiliary reward model to evaluate generated reports. The generation model is then optimized to maximize expected rewards through reinforcement learning algorithms such as PPO [29] or GRPO [30]. Although effective, RLHF suffers from substantial computational overhead, training instability, and potential reward mis-specification, which may lead to reward hacking and clinically unreliable optimization.

Recently, Direct Preference Optimization (DPO) [28] has emerged as a simpler and more stable alternative to RLHF by eliminating explicit reward modeling and policy-gradient optimization. Given pairwise preference data consisting of a prompt, a preferred response, and a dispreferred response, DPO directly encourages the model to assign higher likelihood to the preferred response, enabling efficient one-step preference alignment.

Existing DPO-based MRG approaches [48] typically construct preference pairs by treating ground-truth (GT) reports as preferred responses and model-generated outputs as dispreferred ones, followed by uniform optimization over all tokens. However, as shown in Fig. 1(a), this paradigm overlooks the fundamental nature of medical reports: critical findings, such as disease entities and their assertion status, are tightly interwoven with clinically irrelevant linguistic characteristics, including connective phrases, syntactic variations, and word ordering. Assigning equal importance to all tokens dilutes the alignment signal and weakens supervision on diagnostically relevant information, encouraging the model to favor linguistic similarity over clinical correctness.

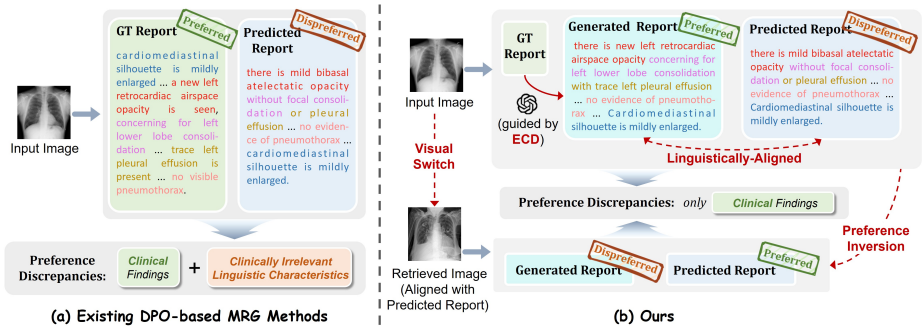

Fig. 1: Comparison between existing DPO-based MRG methods and our DPO-Clin. We utilize distinct colors to indicate descriptions corresponding to different clinical findings. (a) Existing methods construct report preference pairs directly from raw GT and predicted reports, inadvertently entangling critical clinical findings with clinically irrelevant linguistic variations. (b) Guided by the proposed Entity-level Clinical Diagnostic (ECD) module, our DPO-Clin isolates clinical discrepancies by generating linguistically-aligned preference reports. Furthermore, DPO-Clin enforces fine-grained vision-language alignment through a visual-context-triggered preference inversion mechanism. For visual clarity, the mechanism for mitigating latent risks is omitted.

Moreover, existing methods expose the model only to the positive image corresponding to the GT report during preference alignment, without introducing contrastive visual references for the same report. In clinical practice, diagnostic conclusions often depend on subtle and localized visual differences across images [**?**, 14]. Without contrasting image-level preferences, models struggle to associate textual descriptions with discriminative visual evidence, which may ultimately lead to clinically significant errors such as false positives, false negatives.

To address these challenges, we propose **DPO-Clin**, a **DPO**-based framework that explicitly isolates **clin**ically relevant content and establishes multi-modal-aware optimization. As illustrated in Fig. 1(b), our core insight lies in constructing preference report pairs that contain exclusively clinical discrepancies. To this end, we introduce an **E**ntity-level **C**linical **D**iagnostic (**ECD**) module to perform precise entity-level diagnostic. The diagnostic results guide LLM (GPT-4o [1]) to generate a preferred report that is linguistically-aligned with the predicted report, while strictly adhering to the clinical facts of the GT report. Pairing this generated report with the original prediction forces the model optimization to focus on clinically relevant discrepancies. Furthermore, we propose **M^{2}DPO**, a retrieval-augmented **M**ulti-**M**odal **DPO** variant that enforces a textual preference inversion triggered by visual context switches. This design tightly couples visual discrepancies with textual factual differences, compelling the model to ground diagnostic assertions in discriminative pathological cues rather than spurious correlations. Beyond mitigating these explicit textual errors, we also emphasize eliminating latent risks—medical entities that are pre-

dicted correctly yet with high uncertainty. We localize these uncertain entities and apply counterfactual modifications to construct additional preference data, thereby further enhancing the overall clinical reliability of the MRG model.

Our main contributions are summarized as follows:

– We propose DPO-Clin, an innovative DPO-based post-training framework tailored for MRG. By explicitly directing preference optimization to focus on clinically relevant multi-modal discrepancies, DPO-Clin effectively mitigates both explicit textual errors and latent clinical risks inherent in SFT baselines, significantly improving report generation quality.

– We introduce the ECD module for precise entity-level diagnosis, guiding the construction of reliable, linguistically-aligned report preference pairs. Furthermore, we develop M^2DPO, which enforces textual preference inversion triggered by visual context switches to achieve fine-grained vision-language alignment. Additionally, we systematically localize uncertain entities and apply counterfactual modifications to construct targeted preference data for latent risk mitigation.

– We conduct comprehensive evaluations across two chest X-ray datasets (MIMIC-CXR and IU X-Ray) and an in-house endoscopy dataset. Extensive results demonstrate that DPO-Clin significantly elevates the clinical efficacy of SFT baselines and achieves superior performance compared to other DPO-based MRG methods, exhibiting remarkable generalizability across diverse baseline architectures (R2GenGPT and RADAR) and distinct medical imaging modalities.

2 Related Works

2.1 Medical Report Generation

Medical Report Generation (MRG) aims to automatically generate coherent and accurate diagnostic reports from medical images. Early works primarily focused on enhancing report quality through architectural innovations. These include designing advanced attention mechanisms to capture complex cross-modal interactions [5, 36], introducing region-aware localization priors to explicitly ground text in specific pathological regions [6, 15, 33], and integrating structured medical knowledge [21, 40, 45] or auxiliary classification networks [4, 8, 13, 18] to guide decoding and rectify factual deviations. Recently, an advanced paradigm has gained increasing attention, which adapts general-purpose Large Language Models (LLMs) to MRG via Supervised Fine-Tuning (SFT) [11, 17, 37]. Benefiting from massive pre-training corpora, this paradigm preserves fluent text generation capabilities and achieves leading performance. Nevertheless, SFT inherently applies uniform token-level supervision, preventing the model from focusing exclusively on clinically significant content. As a result, the generation process remains plagued by prevalent factual errors, undermining the reliability.

2.2 Direct Preference Optimization

Direct Preference Optimization (DPO) has emerged as an efficient post-training technique, widely adopted to enhance the capabilities of the SFT baseline. Based

on the preference data curation strategy, these approaches can generally be categorized into two types: those relying on manually annotated preference pairs [10, 39, 42], and those utilizing heuristic rules or automatic evaluation models to generate large-scale preference data [10, 34, 43, 44]. Inspired by the success in general domains, recent studies [2, 22, 32] have explored the adaptability of DPO in medical tasks, including MRG. For example, RRG-DPO [25] automatically retrieves paired preference reports from the training database. MMedPO [48] adjusts optimization weights by quantifying the clinical relevance scores of preference reports, and [9] employs report-exclusive evaluation metrics to guide the preference prioritization among multiple model predictions. However, these methods overlook the intrinsic characteristic of medical reports, where critical clinical findings are heavily intertwined with clinically irrelevant linguistic variations. Directly applying DPO to preference reports fails to capture clinically relevant textual discrepancies. Furthermore, they lack sufficient exploration in leveraging DPO to drive fine-grained vision-language alignment, primarily executing a text-centric optimization.

3 Preliminary on DPO-based MRG

Direct Preference Optimization (DPO) [28] has emerged as a simple yet effective paradigm for aligning generative models with human preferences in a stable and reward-free manner. It directly optimizes the policy to favor preferred completions over dispreferred ones relative to a frozen reference policy. Given a query medical image x, current DPO-based MRG approaches [25] typically treat incorrect model predictions y^{pre} as dispreferred reports and ground-truth (GT) reports y^{GT} as preferred ones, maximizing the log-odds margin between them under the target policy π_θ after subtracting the corresponding log-odds under a reference policy π_{ref}:

$$\mathcal{L}_{\text{DPO}}(x, y^{\text{GT}}, y^{\text{pre}}) = -\log \sigma \left(\beta \left[\log \frac{\pi_\theta(y^{\text{GT}}|x)}{\pi_{\text{ref}}(y^{\text{GT}}|x)} - \log \frac{\pi_\theta(y^{\text{pre}}|x)}{\pi_{\text{ref}}(y^{\text{pre}}|x)} \right] \right), \quad (1)$$

where β is a temperature hyperparameter controlling the strength of preference alignment and $\sigma(\cdot)$ denotes the sigmoid function.

However, this paradigm suffers from three limitations that are particularly critical in MRG. **First**, the preference discrepancies between y^{pre} and y^{GT} entangle critical clinical findings with clinically irrelevant linguistic variations. Such uniform, report-level preference assignments prevent the optimization process from isolating clinically relevant textual discrepancies. **Second**, enforcing textual preference supervision conditioned solely on a static image fails to compel the model to ground textual discrepancies in specific pathological visual regions, thereby hindering fine-grained vision-language alignment. **Third**, it exclusively focuses on explicit textual factual errors, ignoring the optimization of 'latent risks'—correct yet highly uncertain predictions.

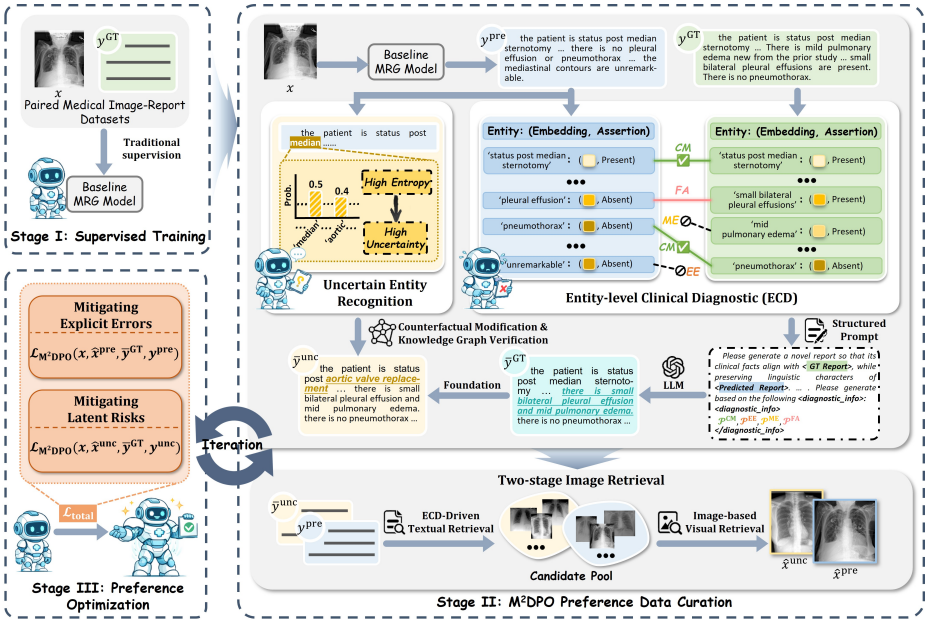

Fig. 2: Overview of the proposed DPO-Clin. DPO-Clin is a DPO-based post-training framework designed to optimize the SFT MRG baseline. It leverages an ECD module to assist LLM in generating linguistically-aligned preferred reports $\overline{y}^{\mathrm{GT}}$, and applies uncertain entity recognition to obtain counterfactual reports $\overline{y}^{\mathrm{unc}}$. By incorporating a two-stage image retrieval strategy, it constructs multi-modal preference data for M^{2}DPO, which ultimately achieves robust clinical preference alignment.

4 Methodology

4.1 Overview

As illustrated in Fig. 2, we propose DPO-Clin, a post-training framework tailored for direct clinical preference alignment. DPO-Clin introduces three core components to tackle the aforementioned limitations. First, we leverage an Entity-level Clinical Diagnostic (ECD) module to assist the Large Language Model (LLM) in precisely eliminating clinically irrelevant linguistic variations from the preference report pairs. Second, we propose M^{2}DPO, a novel multi-modal variant of DPO, to enforce a textual preference inversion triggered by visual context switches, thereby achieving fine-grained vision-language alignment. Third, we localize correct yet highly uncertain predicted clinical entities and apply counterfactual modifications to construct the dispreferred report, optimizing the model's reliability against latent risks. The following sections elaborate on each component and the overall training pipeline.

4.2 ECD-Assisted Linguistically-Aligned Preference Optimization

Medical reports intrinsically intertwine two types of content: *clinical findings*, *i.e.*, medical entities and their assertion status (Present/Absent), and *clinically irrelevant linguistic characteristics*, *i.e.*, connective phrases, syntactic variations, and word ordering. To construct linguistically-aligned preference report pairs, an intuitive approach is to prompt an LLM to generate a new preferred report that preserves the linguistic characteristics of the prediction report, while adhering to the clinical facts of the GT report. However, existing general-purpose LLMs typically exhibit limited sensitivity to medical entities and struggle with medical-specific semantic comprehension, rendering their direct generation unreliable.

To this end, we introduce Entity-level Clinical Diagnostic (ECD) module that performs fine-grained entity-level analysis. Specifically, ECD employs the RaTE-NER model [46] to parse both the predicted report y^{pre} and the GT report y^{GT}, extracting medical entities $\mathbf{e}$ along with their corresponding semantic embeddings $\mathbf{E}$ and assertion statuses $\mathbf{A} \in \{\text{Present}, \text{Absent}\}$. This yields two structured sets: $\mathcal{S}^{\text{pre}} = \{\mathbf{e}_i^{\text{pre}} : (\mathbf{E}_i^{\text{pre}}, \mathbf{A}^{\text{pre}}i)\}_{i=1}^{N}$ and $\mathcal{S}^{\text{GT}} = \{\mathbf{e}_j^{\text{GT}} : (\mathbf{E}_j^{\text{GT}}, \mathbf{A}_j^{\text{GT}})\}_{j=1}^{M}$, where N and M denote entity counts in the predicted and GT reports, respectively.

Following this, to establish fine-grained correspondences, we formulate entity matching as a threshold-constrained linear assignment problem. The matching cost between the i-th predicted entity and j-th GT entity is defined as their cosine distance: $\text{cost}_{i,j} = 1 - \cos(\mathbf{E}_i^{\text{pre}}, \mathbf{E}_j^{\text{GT}})$. The optimal assignment matrix $\mathbf{M}^*$ minimizes the global matching cost:

$$\mathbf{M}^* = \underset{\mathbf{M} \in \{0,1\}^{N \times M}}{\arg\min} \sum_{i=1}^{N} \sum_{j=1}^{M} \text{cost}_{i,j} \cdot \mathbf{M}_{i,j}, \quad \text{s.t.} \sum_{j} \mathbf{M}_{i,j} = \{0,1\}, \sum_{i} \mathbf{M}_{i,j} = \{0,1\}. \tag{2}$$

A similarity threshold τ further constrains the matching, requiring $\text{cost}_{i,j} < 1-\tau$ for any matched pair $(\mathbf{e}_i^{\text{pre}}, \mathbf{e}_j^{\text{GT}})$. Based on $\mathbf{M}^*$, we categorize entities into four clinically meaningful subsets, each corresponding to a distinct error type that will guide subsequent preference construction:

– **Correct Matches** ($\mathcal{P}^{\text{CM}}$): Matched entity pairs $(\mathbf{e}_i^{\text{pre}}, \mathbf{e}_j^{\text{GT}})$ with identical assertion statuses ($\mathbf{M}_{i,j}^* = 1$ and $\mathbf{A}_i = \mathbf{A}_j$), denoting accurate predictions.

– **Extraneous Entities** ($\mathcal{P}^{\text{EE}}$): Unmatched predicted entities $\mathbf{e}_i^{\text{pre}}$ ($\sum_j \mathbf{M}_{i,j}^* = 0$), indicating fabricated findings that are absent in the GT report.

– **Missing Entities** ($\mathcal{P}^{\text{ME}}$): Unmatched GT entities $\mathbf{e}_j^{\text{GT}}$ ($\sum_i \mathbf{M}_{i,j}^* = 0$), representing clinically significant findings that the model ignored.

– **False Assertions** ($\mathcal{P}^{\text{FA}}$): Matched entity pairs $(\mathbf{e}_i^{\text{pre}}, \mathbf{e}_j^{\text{GT}})$ with conflicting assertion statuses ($\mathbf{M}_{i,j}^* = 1$ and $\mathbf{A}_i \neq \mathbf{A}_j$), capturing instances where the model correctly identifies an anatomical entity but misattributes its clinical state.

Finally, these fine-grained diagnostic details are compiled into a structured prompt, guiding the LLM (GPT-4o [1]) to generate $\overline{y}^{\text{GT}}$, which strictly adheres to the clinical facts of y^{GT} while preserving the linguistic characteristics of y^{pre}. By pairing $\overline{y}^{\text{GT}}$ and y^{pre} to construct the preference data, we ensure that the optimization process focuses exclusively on clinically relevant textual

discrepancies. This text-centric preference optimization objective is formulated as $\mathcal{L}_{\text{DPO}}(x, \overline{y}^{\text{GT}}, y^{\text{pre}})$.

4.3 Retrieval-Augmented M^{2}DPO for Vision-Language Alignment

Recognizing the inherent multi-modal nature of MRG, it is crucial to optimize not only the textual output but also the model's visual grounding capabilities. To this end, we propose **M^{2}DPO**, a novel multi-modal variant of DPO. Operating on a quadruplet (x_1, x_2, y_1, y_2) where y_1 and y_2 are factually aligned with x_1 and x_2 respectively, M^{2}DPO enforces a textual preference inversion triggered by visual context switches. Specifically, y_1 serves as the preferred report over y_2 under the visual context of x_1, whereas this preference order is inverted when the visual context switches to x_2. Building upon the standard DPO objective defined in Eq. (1), the M^{2}DPO objective is formulated as follows:

$$\mathcal{L}_{\text{M}^2\text{DPO}}(x_1, x_2, y_1, y_2) = \mathcal{L}_{\text{DPO}}(x_1, y_1, y_2) + \mathcal{L}_{\text{DPO}}(x_2, y_2, y_1). \tag{3}$$

This objective enforces the model to ground clinical textual discrepancies in their corresponding local pathological visual discrepancies, thereby achieving fine-grained vision-language alignment.

To implement the M^{2}DPO objective, it is necessary to augment the preference triplets $(x, \overline{y}^{\text{GT}}, y^{\text{pre}})$ constructed in Sec. 4.2 into quadruplets. To this end, we employ a two-stage strategy to retrieve an image $\hat{x}^{\text{pre}}$ from the training database that is factually aligned with the predicted report y^{pre}. The detailed retrieval protocol is outlined below:

1. ECD-Driven Textual Retrieval: We leverage the proposed ECD module to evaluate the factual alignment between y^{pre} and candidate reports in the training set. Specifically, we construct a candidate pool comprising images whose paired reports yield a valid set of Correct Matches ($|\mathcal{P}^{\text{CM}}| > 0$) while strictly maintaining empty error subsets ($\mathcal{P}^{\text{EE}} \cup \mathcal{P}^{\text{ME}} \cup \mathcal{P}^{\text{FA}} = \varnothing$) when diagnosed against y^{pre}. This entity-level filtering guarantees absolute clinical equivalence.

2. Image-based Visual Retrieval: Within the candidate pool, we extract spatial embeddings using MedKLIP [38], and select the image with the highest cosine similarity to the query image x as the final retrieved image $\hat{x}^{\text{pre}}$.

This two-stage protocol efficiently ensures that $\hat{x}^{\text{pre}}$ is clinically congruent with y^{pre} while minimizing clinically irrelevant visual shifts from x. Thus, we construct the preference quadruplet $(x, \hat{x}^{\text{pre}}, \overline{y}^{\text{GT}}, y^{\text{pre}})$, and then augment the text-centric foundation $\mathcal{L}_{\text{DPO}}(x, \overline{y}^{\text{GT}}, y^{\text{pre}})$ in Sec. 4.2 into the multi-modal objective $\mathcal{L}_{\text{M}^2\text{DPO}}(x, \hat{x}^{\text{pre}}, \overline{y}^{\text{GT}}, y^{\text{pre}})$.

4.4 Counterfactual Data Construction for Latent Risk Mitigation

Beyond optimizing for explicit textual errors, it is also crucial to mitigate latent risks, defined as clinically relevant predictions that are factually correct yet highly uncertain. To localize these risks, we propose an uncertain entity recognition strategy. It quantifies the uncertainty of an entity by identifying the

maximum token-level entropy among its constituent tokens. Assuming an entity $\mathbf{e}$ consists of K tokens, its uncertainty score is defined as:

$$\mathcal{H}(\mathbf{e}|x) = \max_{k=1}^{K} \big(-\sum_{w \in \mathcal{V}} \pi_{\text{ref}}(w|x, \mathcal{T}_{<k}) \log \pi_{\text{ref}}(w|x, \mathcal{T}_{<k}) \big), \tag{4}$$

where w denotes a candidate token from the vocabulary $\mathcal{V}$, and $\mathcal{T}< k$ represents the complete decoding context preceding the k-th entity token. If a predicted entity belongs to the Correct Matches set ($\mathbf{e} \in \mathcal{P}^{\text{CM}}$) and its uncertainty score exceeds a predefined threshold θ, it is flagged as an *uncertain entity*.

To suppress these latent risks, we construct an additional preference quadruplet $(x, \hat{x}^{\text{unc}}, \overline{y}^{\text{GT}}, \overline{y}^{\text{unc}})$ to execute M^{2}DPO. Here, the report $\overline{y}^{\text{unc}}$ is derived via counterfactual modifications founded on $\overline{y}^{\text{GT}}$. Specifically, we replace each uncertain entity with the second most probable candidate entity from the predictive distribution. To ensure clinical plausibility, we introduce a knowledge graph verification based on RadGraph [16]: if the sentence containing the substitute entity includes other medical entities, the substitute must share valid connection edges with them within the RadGraph schema. x^{unc} is then acquired using the retrieval protocol detailed in Sec. 4.3. Finally, the optimization objective for mitigating these latent risks is formulated as $\mathcal{L}_{\text{M}^2\text{DPO}}(x, \hat{x}^{\text{unc}}, \overline{y}^{\text{GT}}, \overline{y}^{\text{unc}})$.

4.5 Training Pipeline

The overall training pipeline of DPO-Clin optimizes a joint multi-modal objective that simultaneously mitigates both explicit errors and latent risks. To ensure stable convergence and strict clinical plausibility, we introduce dynamic binary indicators $\mathbb{1}_{\text{EE}}$ and $\mathbb{1}_{\text{LR}}$ to selectively mask invalid training instances. The total training objective is formulated as:

$$\mathcal{L}_{\text{total}} = \underbrace{\mathbb{1}_{\text{EE}} \mathcal{L}_{\text{M}^2\text{DPO}}(x, \hat{x}^{\text{pre}}, \overline{y}^{\text{GT}}, y^{\text{pre}})}_{\text{mitigating explicit errors (Sec. 4.3)}} + \underbrace{\mathbb{1}_{\text{LR}} \mathcal{L}_{\text{M}^2\text{DPO}}(x, \hat{x}^{\text{unc}}, \overline{y}^{\text{GT}}, \overline{y}^{\text{unc}})}_{\text{mitigating latent risks (Sec. 4.4)}} \tag{5}$$

Here, $\mathbb{1}_{\text{EE}} = 0$ if either the ECD module identifies no factual errors in y^{pre} or the candidate pool obtained in the first-stage retrieval is empty. Conversely, $\mathbb{1}_{\text{LR}} = 0$ if the counterfactual modifications are rejected by the knowledge graph verification. Otherwise, these indicators are set to 1. This dynamic masking mechanism ensures the model exclusively learns from rigorously validated, high-quality multi-modal preference quadruplets.

5 Experiments

5.1 Dataset and Evaluation Metrics

Dataset. To comprehensively evaluate the efficacy of our DPO-Clin in MRG tasks, we conduct experiments on three MRG datasets: two widely-used chest X-ray (CXR) datasets, IU X-Ray [7] and MIMIC-CXR [19], and one in-house endoscopy dataset. For the public CXR datasets, following the previous works [11,

17], we adopt the official split of MIMIC-CXR, allocating 270,790, 2,130, and 3,858 samples for training, validation, and testing, respectively. To assess the model's generalization capability, the entire IU X-Ray dataset (totaling 3,955 samples) is exclusively reserved for testing. Regarding the in-house endoscopy dataset, we divide it into 3,925 samples for training, 982 for validation, and 982 for testing.

Evaluation Metrics. Following the previous works [11,25] , we evaluate model performance using four natural language generation (NLG) metrics (BLEU-1/4 [27], METEOR [3], ROUGE-L [23]), and two clinical efficacy (CE) metrics (macro and micro F_1 scores, ^{14}Ma-F_1 and ^{14}Mi-F_1), which are calculated by converting reports into 14 abnormality classification labels using CheXbert [31]. Moreover, we employ three radiology report generation (RRG)-exclusive metrics (RadGraph [16], RadCliQ [41], RaTEScore [46]) for a more comprehensive evaluation. Notably, as CheXbert is not suitable for endoscopy reports, the CE evaluation is restricted for binary diagnoses (normal/abnormal) on the endoscopy dataset, thus we employ the CE metric ^{2}F$_1$.

5.2 Implementation Details

During the SFT phase, we train the baseline models (R2GenGPT [37] and RADAR [11]) following their official supervision configurations. During the DPO-Clin post-training phase, we employ Low-Rank Adaptation (LoRA) [12] to efficiently fine-tune all linear layer parameters. The model is optimized using the AdamW optimizer [26] with a weight decay of 0, accompanied by a cosine scheduler. Specifically, the training spans 5 epochs with an initial learning rate of 1×10^{-5} for the MIMIC-CXR dataset, and 20 epochs with an initial learning rate of 2×10^{-5} for the in-house endoscopy dataset. Regarding DPO-specific configurations, the temperature hyperparameter β is kept constant at 0.1. The empirical hyperparameters, specifically the matching threshold τ and the uncertainty threshold θ, are set to 0.4 and 0.5 for the CXR datasets, and 0.5 and 0.5 for the endoscopy dataset, respectively. Notably, when applying DPO-Clin to the in-house endoscopy dataset, the RaTE-NER model utilized in the ECD module is fine-tuned using data annotated by endoscopists, ensuring domain-accurate medical entity extraction. Moreover, we employ GastroNet-5M [20] to extract spatial vision embeddings during the image-based visual retrieval stage. All experiments are conducted on 4 NVIDIA RTX 4090 GPUs.

5.3 Main Results

Baselines. We compare our DPO-Clin with three DPO-based VLM post-training methods, including a general-domain method, SIMA [34], and two medically-tailored methods, MMedPO [48] and RRG-DPO [25]. To evaluate universality across architectures, we employ two distinct baselines: R2GenGPT [37], representing a canonical VLM architecture, and RADAR [11], a SOTA MRG method that integrates an auxiliary expert model to augment clinical observations. For

Table 1: Comparison of DPO-Clin with other DPO-based MRG methods on two CXR datasets, IU X-Ray and MIMIC-CXR. **Bold** and underlined text indicates the best performance and second-best performance, respectively.

Dataset	Methods	NLG Metrics				CE Metrics		RRG-exclusive Metrics		
		B-1 ↑	B-4 ↑	MTR ↑	R-L ↑	^{14}Ma-F$_1$ ↑	^{14}Mi-F$_1$ ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
MIMIC-CXR (*Training & Evaluation*)	R2GenGPT	0.411	0.134	0.160	0.297	0.389	0.504	0.260	2.74	0.453
	+ SIMA	0.416	0.140	0.168	0.306	0.396	0.515	0.265	2.74	0.450
	+ MMedPO	**0.426**	**0.144**	**0.174**	0.315	0.420	0.530	0.281	2.71	0.476
	+ RRG-DPO	0.417	0.138	0.165	0.302	0.415	0.522	0.275	2.71	0.470
	+ **Ours**	0.423	0.140	**0.174**	**0.321**	**0.437**	**0.545**	**0.297**	**2.67**	**0.494**
	RADAR	0.509	0.262	0.450	0.397	0.460	0.627	0.346	2.61	0.531
	+ SIMA	0.512	0.264	0.459	0.405	0.464	0.620	0.341	2.57	0.537
	+ MMedPO	0.516	0.268	0.466	0.410	0.478	0.639	0.356	2.52	0.546
	+ RRG-DPO	0.511	0.260	0.459	0.408	0.473	0.639	0.356	2.54	0.550
	+ **Ours**	**0.522**	**0.271**	**0.471**	**0.419**	**0.490**	**0.654**	**0.372**	**2.48**	**0.567**
IU X-Ray (*Evaluation*)	R2GenGPT	0.356	0.103	0.138	0.270	0.303	0.446	0.204	2.86	0.405
	+ SIMA	0.359	0.106	0.143	0.277	0.312	0.458	0.210	2.85	0.413
	+ MMedPO	**0.370**	**0.114**	0.153	0.286	0.332	0.475	0.231	2.80	0.428
	+ RRG-DPO	0.367	0.110	0.146	0.277	0.326	0.464	0.220	2.82	0.422
	+ **Ours**	0.365	0.112	**0.159**	**0.292**	**0.350**	**0.492**	**0.252**	**2.74**	**0.452**
	RADAR	0.364	0.116	0.142	0.276	0.325	0.546	0.237	2.78	0.428
	+ SIMA	0.370	0.118	0.148	0.281	0.331	0.555	0.244	2.76	0.435
	+ MMedPO	0.374	**0.126**	0.152	0.289	0.345	0.568	0.259	2.70	0.452
	+ RRG-DPO	0.372	0.124	0.148	0.285	0.338	0.560	0.251	2.72	0.444
	+ **Ours**	**0.377**	**0.126**	**0.160**	**0.298**	**0.362**	**0.580**	**0.275**	**2.66**	**0.461**

Table 2: Comparison on the in-house endoscopy report dataset.

Methods	NLG Metrics				CE Metric
	B-1	B-4	MTR	R-L	^{2}F$_1$
R2GenGPT	0.528	0.401	0.545	0.523	0.798
+ SIMA	0.546	0.418	0.574	0.541	0.825
+ MMedPO	0.558	0.424	0.585	0.557	0.834
+ RRG-DPO	0.540	0.417	0.566	0.536	0.813
+ **Ours**	**0.576**	**0.435**	**0.607**	**0.574**	**0.852**

a fair comparison, we implement a unified training pipeline, applying these preference optimization algorithms as a post-training stage following the supervised fine-tuned baseline.

Comparison on two CXR Datasets. In Table 1, we validate the in-domain performance on MIMIC-CXR and cross-domain generalization on IU X-Ray, respectively. It can be seen that all of these preference optimization algorithms consistently enhance the baseline performance on both datasets, demonstrating the effectiveness of preference optimization following supervised fine-tuning. Crucially, our DPO-Clin outperforms competing methods in almost all metrics, showing the most significant improvements in report quality, particularly in CE and RRG-exclusive metrics that emphasize clinical relevance. These results confirm the superiority of our DPO-Clin in enforcing clinical factuality and generating medically accurate reports.

Comparison on the Endoscopy Dataset. To further validate the universality of different methods beyond the radiology scenario, we conduct experiments by training models from scratch on the in-house endoscopy report dataset. R2GenGPT is selected as the baseline model as it features a streamlined, modality-

Table 3: Ablation on the effectiveness of Mitigating Explicit Errors and Latent Risks, denoted as 'MEE' and 'MLR'. 'ECD', 'Verification', and 'Masking' denote Entity-level Clinical Diagnostic module, knowledge graph verification, and dynamic masking mechanism, respectively. '$\overline{y}^{\mathrm{GT}} \rightarrow y^{\mathrm{GT}}$' indicates using the GT report y^{GT} instead of the generated linguistically-aligned report $\overline{y}^{\mathrm{GT}}$ in the total objective $\mathcal{L}_{\mathrm{total}}$.

Methods	NLG Metrics				CE Metrics		RRG-exclusive Metrics		
	B-1	B-4	MTR	R-L	^{14}Ma-F$_1$	^{14}Mi-F$_1$	RadGraph	RadCliQ	RaTEScore
RADAR (Baseline)	0.509	0.262	0.450	0.397	0.460	0.627	0.346	2.61	0.531
+ MEE (*w/o* ECD)	0.498	0.256	0.445	0.390	0.452	0.615	0.337	2.64	0.518
+ MEE	0.518	0.268	0.466	0.414	0.482	0.645	0.366	2.51	0.557
+ MLR (*w/o* Verification)	0.513	0.265	0.454	0.406	0.472	0.635	0.354	2.55	0.546
+ MLR	0.513	0.266	0.459	0.409	0.475	0.640	0.360	2.55	0.550
+ MEE&MLR (*w/o* Masking)	0.512	0.262	0.454	0.403	0.466	0.638	0.353	2.58	0.541
+ MEE&MLR ($\overline{y}^{\mathrm{GT}} \rightarrow y^{\mathrm{GT}}$)	**0.526**	**0.273**	0.466	0.417	0.482	0.642	0.360	2.52	0.552
+ MEE&MLR (DPO-Clin)	0.522	0.271	**0.471**	**0.419**	**0.490**	**0.654**	**0.372**	**2.48**	**0.567**

agnostic architecture that enables direct transfer to endoscopy scenario. As illustrated in Table 2, the comparison results on the endoscopy dataset align with those observed in the CXR benchmark. Compared to other preference optimization methods, our DPO-Clin consistently maintains a significant performance advantage. Specifically, DPO-Clin achieves the highest CE score, with its $^{2}\mathrm{F}_1$ value improving by 5.4% (0.852 *vs.* 0.798) compared to the baseline model and surpassing the second-best method MMedPO by 1.8% (0.852 *vs.* 0.834). This demonstrates DPO's robustness and effectiveness across diverse medical imaging modalities.

5.4 In-Depth Analysis

Effectiveness of Mitigating Explicit Errors and Latent Risks. In Table 3, we progressively augment the RADAR baseline with our proposed components on the MIMIC-CXR dataset. The results demonstrate that mitigating explicit textual errors and latent clinical risks independently can further improve the quality of medical reports generated by the baseline model. Moreover, the combination of both components achieves the best performance, indicating that their contributions to the baseline optimization are complementary.

Importantly, the two proposed sub-components, Entity-level Clinical Diagnostic (ECD) module and Knowledge Graph Verification mechanism, both play a pivotal role in ensuring clinical reliability. As shown in Table 3, disabling the verification mechanism yields limited improvements, while removing the ECD module results in performance degradation compared to the baseline. This discrepancy arises primarily because the ECD module guarantees that the LLM-generated reports $\overline{y}_{\mathrm{GT}}$ adhere to clinical factuality, whereas the verification ensures the clinical plausibility of the counterfactually modified reports $\overline{y}^{\mathrm{unc}}$.

Furthermore, the ablation underscores the necessity of the linguistically-aligned preference construction and the dynamic masking mechanism within the joint training framework. As shown in the last three rows of Table 3, substituting

Table 4: Performance comparison of Different DPO Variants.

Methods	MIMIC-CXR					Endoscopy	
	NLG	CE	RRG-exclusive			NLG	CE
	R-L	^{14}Ma-F$_1$	Graph	CliQ	RaTEScore	R-L	^{2}F$_1$
R2GenGPT (Baseline)	0.297	0.389	0.260	2.74	0.453	0.523	0.798
+ standard DPO	0.311	0.420	0.278	2.71	0.471	0.549	0.832
+ SimPO	0.313	0.425	0.287	2.70	0.480	0.560	0.838
+ M^{2}DPO (Ours)	**0.321**	**0.437**	**0.297**	**2.67**	**0.494**	**0.574**	**0.852**

the LLM-generated linguistically-aligned reports $\overline{y}^{\mathrm{GT}}$ with the raw GT reports y^{GT} marginally improves n-gram metrics (*i.e.*, BLEU-1 and BLEU-4) but significantly degrades CE and RRG-exclusive metrics. This indicates that while the model learns to mimic clinically irrelevant descriptions, it does so at the severe cost of clinical accuracy. Such a trade-off shows the necessity of constructing linguistically-aligned preference data to ensure the optimization focuses on capturing clinically relevant discrepancies. Similarly, disabling the dynamic masking mechanism compromises the overall framework performance. It ensures that the optimization targets failed predictions, as well as the reliability of the retrieved images $\hat{x}^{\mathrm{GT}}$ and $\hat{x}^{\mathrm{unc}}$ and the counterfactually modified reports $\overline{y}^{unc}$, thereby maintaining the stability and efficacy of the entire pipeline.

Comparison of Different DPO Variants. To validate the effectiveness of incorporating visual constraints during preference optimization, we compare our M^{2}DPO with text-centric competitors, the standard DPO and SimPO, in Table 4. We conduct evaluation on MIMIC-CXR and the in-house endoscopy dataset, setting R2GenGPT as the baseline model. Taking the standard DPO implementation as an example, the overall objective in Eq. (5) is reformulated as $\mathcal{L}_{\mathrm{total}} = \mathbb{1}_{\mathrm{EE}}\mathcal{L}_{\mathrm{DPO}}(x, \overline{y}^{\mathrm{GT}}, y^{\mathrm{pre}}) + \mathbb{1}_{\mathrm{LR}}\mathcal{L}_{\mathrm{DPO}}(x, \overline{y}^{\mathrm{GT}}, \overline{y}^{\mathrm{unc}})$. As illustrated in Table 4, our M^{2}DPO consistently outperforms both DPO and SimPO across all evaluated metrics on both datasets. This quantitatively confirms that incorporating a multi-modal preference alignment mechanism enhances the clinical reliability of MRG models across diverse medical imaging modalities.

To further analyze the underlying mechanism, we visualize the cross-modal attention weights between the generated text tokens and the input visual tokens in the LLM decoder. As shown in Fig. 3, models optimized with the standard DPO and SimPO both exhibit scattered attention distribution that are inconsistent with the actual pathological regions (indicated by green arrows). This visual misalignment results in ambiguous descriptions (*e.g.*, the generic 'pleural effusion') or severe factual errors (*e.g.*, missing of the polyp). In contrast, M^{2}DPO yields attention maps that are more consistent with the actual lesion regions, which empowers the model to tackle challenging scenarios, such as detecting the small sessile flat polyp. These results verify that M^{2}DPO enforces fine-grained vision-language alignment of the MRG baselines.

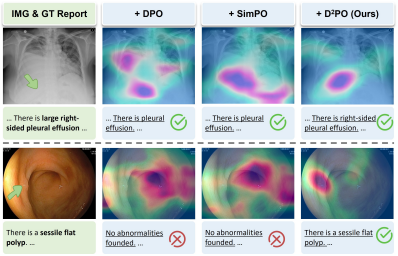

Fig. 3: Visualization of cross-modal attention maps. The attention map represents the average attention weights of all tokens within the underlined sentence.

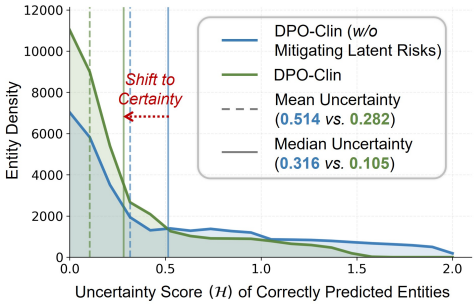

Fig. 4: Density distribution of uncertainty scores for correctly predicted entities with/without mitigating latent risks.

Effect of Latent Risks Mitigation on Entity Uncertainty. As illustrated in Fig. 4, we plot the frequency distribution curves of uncertainty scores for all correctly predicted entities on MIMIC-CXR. Using RADAR as the baseline model, the entity extraction and correctness verification are performed via the ECD module, with uncertainty scores calculated according to Eq. (4). It is seen that after mitigating latent risks, the uncertainty distribution of correct entities exhibits a shift toward lower values (mean uncertainty score: $0.514 \rightarrow 0.282$, median uncertainty score: $0.316 \rightarrow 0.105$). This 'shift to certainty' phenomenon demonstrates the model's increasing confidence in its accurate predictions. Furthermore, the expansion of the area under the curve intuitively signifies an increase in the total count of correctly predicted entities, thus validating the enhanced clinical accuracy of the generated reports.

6 Conclusion

In this paper, we propose DPO-Clin, a DPO-based post-training framework tailored for MRG. Existing DPO-based MRG methods suffer from fundamental limitations: their preference curation inadvertently entangles critical clinical findings with clinically irrelevant linguistic characteristics, and they lack explicit vision-language alignment. To address these issues, we introduce the ECD module to assist in generating linguistically-aligned preference data, forcing the optimization process to focus exclusively on clinically relevant discrepancies. Furthermore, we propose M^2DPO to achieve fine-grained vision-language grounding. By explicitly emphasizing the necessity of mitigating latent risks, DPO-Clin further elevates the diagnostic reliability of the generated reports. Extensive evaluations across diverse medical imaging modalities (chest X-rays and endoscopy) consistently demonstrate that DPO-Clin achieves SOTA performance. Furthermore, the plug-and-play generalizability of DPO-Clin has also been verified across two distinct baseline architectures (R2GenGPT and RADAR), paving the way for groundbreaking advancements in the future development of trustworthy medical

report generation. This is conducive to the further development of trustworthy medical report generation.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (Grant No.62576145), research grants from Wuhan United Imaging Healthcare Surgical Technology Co., Ltd.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Banerjee, O., Zhou, H.Y., Adithan, S., Kwak, S., Wu, K., Rajpurkar, P.: Direct preference optimization for suppressing hallucinated prior exams in radiology report generation. arXiv preprint arXiv:2406.06496 (2024)
3. Banerjee, S., Lavie, A.: 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. pp. 65–72 (2005)
4. Bu, S., Li, T., Yang, Y., Dai, Z.: Instance-level expert knowledge and aggregate discriminative attention for radiology report generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14194–14204 (2024)
5. Chen, Z., Song, Y., Chang, T.H., Wan, X.: Generating radiology reports via memory-driven transformer. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1439–1449 (2020)
6. Chen, Z., Bie, Y., Jin, H., Chen, H.: Large language model with region-guided referring and grounding for ct report generation. IEEE Transactions on Medical Imaging (2025)
7. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R., McDonald, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. Journal of the American Medical Informatics Association **23**(2), 304–310 (2015)
8. Fan, Y., Yang, Z., Liu, R., Li, M., Chang, X.: Medical report generation is a multi-label classification problem. arXiv preprint arXiv:2409.00250 (2024)
9. Hein, D., Chen, Z., Ostmeier, S., Xu, J., Varma, M., Reis, E.P., Michalson, A.E., Bluethgen, C., Shin, H.J., Langlotz, C., et al.: Preference fine-tuning for factuality in chest x-ray interpretation models without human feedback (2024)
10. Hong, I., Li, Z., Bukharin, A., Li, Y., Jiang, H., Yang, T., Zhao, T.: Adaptive preference scaling for reinforcement learning with human feedback. Advances in Neural Information Processing Systems **37**, 107249–107269 (2024)
11. Hou, W., Cheng, Y., Xu, K., Li, H., Hu, Y., Li, W., Liu, J.: Radar: Enhancing radiology report generation with supplementary knowledge injection. arXiv preprint arXiv:2505.14318 (2025)
12. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)

13. Hu, Q., Wang, Q., Chen, J., Ji, X., Liu, M., Li, Q., Wang, Z.: Holistic white-light polyp classification via alignment-free dense distillation of auxiliary optical chromoendoscopy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 251–261. Springer (2025)
14. Hu, Q., Wang, Q., Guo, Y., Li, Q., Wang, Z.: Pairing-free group-level knowledge distillation for robust gastrointestinal lesion classification in white-light endoscopy. arXiv preprint arXiv:2601.09209 (2026)
15. Hu, Q., Yi, Z., Zhou, Y., Peng, F., Liu, M., Li, Q., Wang, Z.: Sali: Short-term alignment and long-term interaction network for colonoscopy video polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 531–541. Springer (2024)
16. Jain, S., Agrawal, A., Saporta, A., Truong, S.Q., Duong, D.N., Bui, T., Chambon, P., Zhang, Y., Lungren, M.P., Ng, A.Y., et al.: Radgraph: Extracting clinical entities and relations from radiology reports. arXiv preprint arXiv:2106.14463 (2021)
17. Jin, H., Che, H., Lin, Y., Chen, H.: Promptmrg: Diagnosis-driven prompts for medical report generation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 2607–2615 (2024)
18. Jing, B., Xie, P., Xing, E.: On the automatic generation of medical imaging reports. In: Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers). pp. 2577–2586 (2018)
19. Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Mark, R.G., Horng, S.: Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data **6**(1), 317 (2019)
20. Jong, M.R., Boers, T.G., Fockens, K.N., Jukema, J.B., Kusters, C.H., Jaspers, T.J., van Heslinga, R.v.E., Slooter, F.C., Struyvenberg, M.R., Bisschops, R., et al.: Gastronet-5m: A multicenter dataset for developing foundation models in gastrointestinal endoscopy. Gastroenterology (2025)
21. Li, M., Lin, B., Chen, Z., Lin, H., Liang, X., Chang, X.: Dynamic graph enhanced contrastive learning for chest x-ray report generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3334–3343 (2023)
22. Liang, X., Hu, J., Wang, D., Ma, Z., Zhao, L., Li, R., Wan, B., Wang, Q.: Chexpo: Preference optimization for chest x-ray vlms with counterfactual rationale. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 2606–2615 (2025)
23. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Text summarization branches out. pp. 74–81 (2004)
24. Liu, F., Yin, C., Wu, X., Ge, S., Zhang, P., Sun, X.: Contrastive attention for automatic chest x-ray report generation. In: Findings of the association for computational linguistics: ACL-IJCNLP 2021. pp. 269–280 (2021)
25. Liu, H., Wei, D., Xu, Z., Wu, X., Zheng, Y., Wang, L.: Rrg-dpo: Direct preference optimization for clinically accurate radiology report generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 552–562. Springer (2025)
26. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
27. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the Association for Computational Linguistics. pp. 311–318 (2002)

28. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. Advances in neural information processing systems **36**, 53728–53741 (2023)
29. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
30. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
31. Smit, A., Jain, S., Rajpurkar, P., Pareek, A., Ng, A.Y., Lungren, M.P.: Chexbert: combining automatic labelers and expert annotations for accurate radiology report labeling using bert. arXiv preprint arXiv:2004.09167 (2020)
32. Sun, G., Qin, C., Fu, H., Wang, L., Tao, Z.: Self-training large language and vision assistant for medical question answering. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. pp. 20052–20060 (2024)
33. Tanida, T., Müller, P., Kaissis, G., Rueckert, D.: Interactive and explainable region-guided radiology report generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7433–7442 (2023)
34. Wang, X., Chen, J., Wang, Z., Zhou, Y., Zhou, Y., Yao, H., Zhou, T., Goldstein, T., Bhatia, P., Kass-Hout, T., et al.: Enhancing visual-language modality alignment in large vision language models via self-improvement. In: Findings of the Association for Computational Linguistics: NAACL 2025. pp. 268–282 (2025)
35. Wang, Y., Gao, S., Liu, J., Jiang, S., Xia, H., Zhang, X., Kang, Z., Wang, Y., Liu, Z.: Beyond n-grams: A hierarchical reward learning framework for clinically-aware medical report generation. arXiv preprint arXiv:2512.02710 (2025)
36. Wang, Z., Liu, L., Wang, L., Zhou, L.: Metransformer: Radiology report generation by transformer with multiple learnable expert tokens. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11558–11567 (2023)
37. Wang, Z., Liu, L., Wang, L., Zhou, L.: R2gengpt: Radiology report generation with frozen llms. Meta-Radiology **1**(3), 100033 (2023)
38. Wu, C., Zhang, X., Zhang, Y., Wang, Y., Xie, W.: Medklip: Medical knowledge enhanced language-image pre-training for x-ray diagnosis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21372–21383 (2023)
39. Yang, K., Tao, J., Lyu, J., Ge, C., Chen, J., Shen, W., Zhu, X., Li, X.: Using human feedback to fine-tune diffusion models without any reward model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8941–8951 (2024)
40. Yang, S., Wu, X., Ge, S., Zhou, S.K., Xiao, L.: Knowledge matters: Chest radiology report generation with general and specific knowledge. Medical image analysis **80**, 102510 (2022)
41. Yu, F., Endo, M., Krishnan, R., Pan, I., Tsai, A., Reis, E.P., Fonseca, E.K.U.N., Lee, H.M.H., Abad, Z.S.H., Ng, A.Y., et al.: Evaluating progress in automatic chest x-ray radiology report generation. Patterns **4**(9) (2023)
42. Yu, T., Yao, Y., Zhang, H., He, T., Han, Y., Cui, G., Hu, J., Liu, Z., Zheng, H.T., Sun, M., et al.: Rlhf-v: Towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13807–13816 (2024)
43. Yu, T., Zhang, H., Li, Q., Xu, Q., Yao, Y., Chen, D., Lu, X., Cui, G., Dang, Y., He, T., et al.: Rlaif-v: Open-source ai feedback leads to super gpt-4v trustworthiness. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 19985–19995 (2025)

44. Zhang, M., Wu, W., Lu, Y., Song, Y., Rong, K., Yao, H., Zhao, J., Liu, F., Feng, H., Wang, J., et al.: Automated multi-level preference for mllms. Advances in Neural Information Processing Systems **37**, 26171–26194 (2024)
45. Zhang, Y., Wang, X., Xu, Z., Yu, Q., Yuille, A., Xu, D.: When radiology report generation meets knowledge graph. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 12910–12917 (2020)
46. Zhao, W., Wu, C., Zhang, X., Zhang, Y., Wang, Y., Xie, W.: Ratescore: A metric for radiology report generation. arXiv preprint arXiv:2406.16845 (2024)
47. Zhou, Z., Shi, M., Wei, M., Alabi, O., Yue, Z., Vercauteren, T.: Large model driven radiology report generation with clinical quality reinforcement learning. arXiv preprint arXiv:2403.06728 (2024)
48. Zhu, K., Xia, P., Li, Y., Zhu, H., Wang, S., Yao, H.: Mmedpo: Aligning medical vision-language models with clinical-aware multimodal preference optimization. In: Forty-second International Conference on Machine Learning (2025)