VARIATIONAL FINDING-ANCHORED DIRECT PREFERENCE OPTIMIZATION FOR RADIOLOGY REPORT GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Radiology Report Generation (RRG) aims to generate diagnostic reports from medical radiology images and has achieved substantial progress on benchmark datasets. Nonetheless, existing RRG methods may produce clinically inconsistent reports, with omitted findings and imperfect entity-relation representations. Direct Preference Optimization (DPO) and its variants offer a way to improve report generation through preference alignment. However, adapting DPO to finding-specific preference alignment remains challenging. In this work, we propose a Variational Finding-Anchored Direct Preference Optimization (VF-DPO) approach that introduces finding-aware Question Answer pairs as latent variables to anchor the generation process. Specifically, we approximate the conditional report generation probability with finding anchors as conditions via its Evidence Lower BOund (ELBO), with which we derive a novel DPO formulation in a unified preference optimization objective that jointly models preference in latent finding-state alignment and high-fidelity report generation. To support this bi-level alignment, we construct a corresponding bi-level preference dataset comprising both finding-anchored latent variables and report sequences for training VF-DPO. At test time, we use the trained model to directly generate the radiology report, where the finding-anchored latent variables are implicitly accounted for in training, without being explicitly generated in the testing phase. Extensive experiments on MIMIC-CXR (X-ray) and CT-RATE (CT) demonstrate that VF-DPO achieves the highest F1 and RadGraph scores across all backbone comparisons, while obtaining the best RadCliQ and RaTEScore results on most backbones. These results indicate that VF-DPO enhances overall clinical consistency and outperforms state-of-the-art preference optimization methods for RRG.

1 INTRODUCTION

Automated Radiology Report Generation (RRG) has shown great promise in clinical decision support by translating complex medical radiographs, such as 2D chest X-rays and 3D CT scans, into coherent clinical text. Automated RRG aims to improve diagnostic efficiency and reduce the workload of medical professionals. However, the current RRG methods still suffer from hallucination, especially the finding-related key information may be hallucinated and lack clinical consistency with the input radiology image. This poses a challenge in balancing linguistic fluency and diagnostic fidelity, making it difficult for current models to meet reliability requirements of real-world clinical practice.

Various automated RRG methods have been proposed for improving the quality of the generated reports in clinical diagnosis. Researchers have advanced model structures or training strategies by introducing cross-modal memory modules (Chen et al., 2021; Shen et al., 2024), multi-modal alignment (Yang et al., 2023), and contrastive learning (Li et al., 2024). To explicitly regularize the generation process, some works incorporate external guidance through domain knowledge graphs (Huang et al., 2023; Yang et al., 2022; Yin et al., 2025), retrieved reports (Liu et al., 2024; Zhou et al., 2025; Liu et al., 2026), and finding tags (Jin et al., 2024; Wang et al., 2022; Chen et al., 2025b), where predicted finding categories are utilized as prompts to guide the language model. Despite these advancements in incorporating explicit clinical context, these existing approaches often treat every word equally for supervised regression of ground truth reports in training. As a result, the trained

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

models may generate reports that are consistent with the writing style of the training data, but not well aligned with the preference of radiologists for clinical accuracy.

To improve the clinical consistency in radiology report generation, recent works have introduced preference optimization to RRG, transitioning from reinforcement learning approaches (Xiao et al., 2025) to more stable DPO paradigms (Liu et al., 2025; Zhu et al., 2025). Existing DPO-based methods for RRG typically construct preferred and dispreferred report pairs to explicitly penalize diagnostic errors at the report level. However, despite their empirical success, these methods align the preference at the coarse report level, which may still generate incorrect finding descriptions. To achieve finer-grained control, token-level DPO variants (Li et al., 2025; Yang et al., 2026) have been explored in the general NLP domain to re-weight specific words and provide localized penalties. However, directly applying these token-level operations to the medical domain is suboptimal. While these methods offer finer granularity, manipulating token-level statistics may be insufficient to capture the complex clinical semantics of finding states.

To tackle this challenge, we propose a Variational Finding-Anchored Direct Preference Optimization (VF-DPO) approach to improve overall clinical consistency of generated radiology reports. Different from standard DPO that aligns the generation process at the sequence level, VF-DPO introduces interpretable finding-aware Question–Answer (QA) pairs as latent variables to anchor report generation during training. Specifically, we use an Evidence Lower BOund (ELBO) as a surrogate for the marginal report log-likelihood and derive a bi-level preference objective that couples finding-state alignment with anchor-conditioned report alignment. To support this bi-level alignment, we construct a preference dataset containing both the finding-anchored latent level and the report sequence level. By training on preference pairs constructed from latent finding anchors and reports, our model leverages mathematically grounded alignment to improve clinical finding coverage and finding-state consistency. During inference, we utilize our trained model to directly generate the full report for the given input image, without explicitly generating latent finding-anchors.

The main contributions of our work are summarized as follows:

- We propose a Variational Finding-Anchored Direct Preference Optimization (VF-DPO) framework, which models finding-aware QA pairs as latent variables to explicitly anchor the report generation process, thereby improving the overall quality of clinical finding and state descriptions.
- We derive a bi-level preference optimization objective via ELBO, which jointly allows the model to concurrently align preferences for both latent finding-state alignment and the full report generation.
- Extensive experiments on MIMIC-CXR (X-ray) and CT-RATE (CT) datasets demonstrate the favorable capability of our VF-DPO method in improving clinical finding coverage and overall clinical consistency with reference reports compared to state-of-the-art preference alignment techniques.

2 Related work

Radiology report generation. RRG aims to generate comprehensive clinical narratives from radiology medical images. To bridge the semantic gap between visual modalities and clinical text, existing methods have predominantly relied on injecting explicit structural priors, utilizing external constraints, or designing auxiliary alignment objectives. These methods include knowledge graph injection (Huang et al., 2023; Yang et al., 2022; Yin et al., 2025), cross-modal contrastive learning (Li et al., 2024), finding tag-augmented generation (Jin et al., 2024; Wang et al., 2022; Chen et al., 2025b), retrieval-augmented generation (Liu et al., 2024; Zhou et al., 2025), and memory-driven networks (Chen et al., 2020). Recently, Large Language Models (LLMs) (Touvron et al., 2023; Grattafiori et al., 2024) have achieved great success in reasoning and generative tasks. By extending LLMs to process visual inputs, recent methods have made great advances in radiology report generation (Chen et al., 2025b; Wang et al., 2023; Cheng & Subramanian, 2026; Chen et al., 2025a). Despite these architectural and task-specific advancements, these approaches often treat every word equally for supervised regression of ground truth reports in training. By strictly maximizing the autoregressive token likelihood, this training paradigm fits the linguistic distribution of the training data, but may fail to align with the diagnostic preferences.

Preference optimization in foundation models. Aligning generative models with human preferences has increasingly shifted from computationally heavy Reinforcement Learning from Human Feedback (RLHF) (Bai et al., 2022) to more efficient alternatives like GRPO (Shao et al., 2024) and DPO (Rafailov et al., 2023), with the latter further circumventing the instabilities of reward modeling. This reward-free paradigm DPO has been broadly applied to enhance Large Language Models (LLMs) in logical reasoning (Lai et al., 2024), strict instruction-following (Huang et al., 2025), and code generation (Zhang et al., 2025). Beyond text, DPO is actively deployed in multimodal domains to mitigate cross-modal hallucinations (Xie et al., 2024; Yang et al., 2025) and improve generative alignment (Wallace et al., 2024). However, standard DPO operates at the sequence level, assigning equal optimization weight to all tokens and overlooking fine-grained critical information. To address this, token-level DPO variants (Li et al., 2025; Yang et al., 2026; Zeng et al., 2024) execute optimization directly at the lexical level. Recent methods achieve this through token-level KL constraints (Zeng et al., 2024), optimal transport-based dynamic weighting (Li et al., 2025), or hybrid gradient attribution (Yang et al., 2026) for fine-grained semantic control. Concurrently, the latent-level DPO method LPC (Gong et al., 2025) aims to capture implicit preferences by employing discrete latent codes to infer conflicting factors, but this approach leaves the latent codes without clear structural and semantic grounding.

Preference optimization in radiology report generation. Recent studies explore preference alignment to improve clinical consistency for RRG (Xiao et al., 2025; Liu et al., 2025; Zhu et al., 2025; Banerjee et al., 2024; Hein et al., 2024). MPO (Xiao et al., 2025) employs reinforcement learning to optimize multi-dimensional clinical rewards. Consequently, DPO-based RRG preference optimization methods have been adopted to suppress clinical hallucinations via specialized data curation. Banerjee et al. (Banerjee et al., 2024) leverage GPT-4 to curate preference pairs and introduce a weighted DPO objective, explicitly designed to penalize hallucinated comparisons to previous imaging studies. Hein et al. (Hein et al., 2024) introduce automated pipelines to construct preferred and dispreferred data utilizing LLM-based metrics like GREEN for DPO. To address modality misalignment, MMedPO (Zhu et al., 2025) constructs dispreferred samples via both textual hallucination injection and localized lesion-noising. It then dynamically weights the DPO objective using a Clinical Relevance Score (CRS) derived from multi-agent Med-LLM consensus and visual tool confidence. RRG-DPO (Liu et al., 2025) improves alignment by retrieving hard negatives with significant diagnostic discrepancies measured by Jaccard distance between abnormality labels and introduces sub-preference optimization to preserve clinically equivalent content by deleting only conflicting sentences. Although these methods effectively penalize incorrect report outputs, they realize the preference optimization at the output report level without explicit constraints at the finding state diagnosis.

Different from standard DPO (Rafailov et al., 2023), its token-level variants (Li et al., 2025; Yang et al., 2026), and existing RRG preference alignment methods (Liu et al., 2025; Zhu et al., 2025; Banerjee et al., 2024; Hein et al., 2024), our VF-DPO derives a bi-level preference optimization objective utilizing variational approximation, which jointly models preference in the latent finding-state alignment level and the high-fidelity report generation level. Different from the latent preference method that utilizes latent codes to formulate the diversity of human preference (Gong et al., 2025), we introduce interpretable finding-aware Question-Answer (QA) pairs as latent variables to anchor the report generation process for improving the clinical fidelity of finding diagnosis description. During inference, the trained model directly generates the full report for the given input image, without requiring the explicit generation of the latent finding anchors.

3 METHOD

In this paper, we focus on the preference optimization for radiology report generation. It aims to make the model generate a preferred report y_w and reject a dispreferred report y_l for a given radiology image x. We propose Variational Finding-Anchored Direct Preference Optimization (VF-DPO) to improve finding-level diagnostic consistency in Radiology Report Generation (RRG). Different from standard DPO, which aligns preferences at the sequence level via $\pi(y|x)$ without explicit clinical constraints shown in Fig. 1(a), VF-DPO utilizes finding-aware QA pairs as latent variables b, which are extracted from reports (Fig. 1(b)), to anchor generation. Through variational inference, we design a bi-level objective for RRG preference optimization shown in Fig. 1(c). Specifically, the latent level $\pi(b|x)$ encourages extracting correct finding states b_w over incorrect finding states b_l. Concurrently,

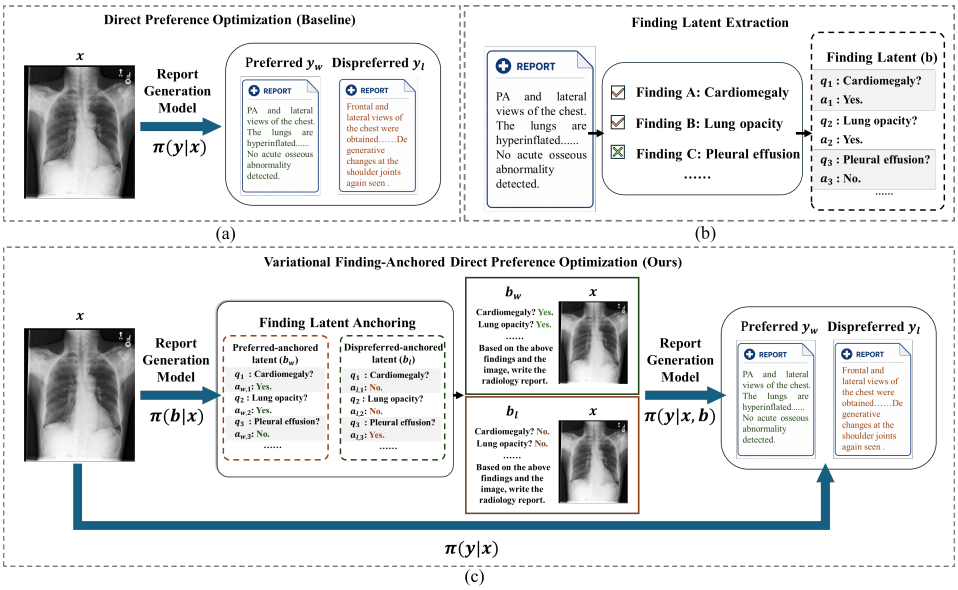


Figure 1: Illustration of our proposed variational finding-anchored direct preference optimization for radiology report generation.

the report level $\pi(y|x, b)$ promotes generating preferred reports y_w conditioned on b_w and x, while penalizing dispreferred y_l given b_l and x. Inference requires only the input image without explicit anchor generation. Next, we detail our formulation and data construction.

3.1 Variational Finding-Anchored Direct Preference Optimization

In this section, we formally introduce the proposed VF-DPO framework. Given a radiology image x, our objective is to generate a clinically preferred report y_w grounded in accurate finding descriptions. To explicitly capture the finding diagnosis, we introduce the latent variable b, representing a structured sequence of finding-aware QA pairs, into the report generation model. The conditional probability of generating the report y given x is defined by marginalizing over all possible latent variables:

$$\pi(y|x) = \mathbb{E}_{b\sim\pi(b|x)}[\pi(y|b, x)] \tag{1}$$

This formulation naturally indicates that generating a high-quality report requires the model to jointly optimize two coupled components $\pi(b|x)$ and $\pi(y|b, x)$. The item $\pi(b|x)$ is to make the model generate the correct finding diagnostic states for the given image, and $\pi(y|b, x)$ aims to make the model generate an accurate report conditioned on both the image and these diagnostic states.

Directly maximizing the marginal conditional log-likelihood $\log \pi(y|x)$ is computationally intractable, as it requires marginalizing over all possible latent states of b, making the true posterior $\pi(b|y, x)$ inaccessible. To bypass this computational bottleneck, we resort to variational inference. We introduce a variational distribution $\tilde{\pi}(b|y, \hat{y}, x)$ (detailed in Sect. 3.2) to approximate the true posterior. Letting $\hat{y}$ denote the preference counterpart of y (*i.e.*, $\hat{y} = y_l$ if $y = y_w$, and vice versa), we can derive the Evidence Lower Bound (ELBO) for $\log \pi(y|x)$, denoted as $\mathcal{E}(y, \hat{y}, x; \pi)$:

$$\log \pi(y|x) \geq \mathcal{E}(y, \hat{y}, x; \pi) \triangleq \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)}[\log \pi(y|b, x)] - \mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi(b|x)). \tag{2}$$

The proof of Eq. (2) is given in Appendix A. Let π_θ and π_{ref} denote the trainable policy and fixed reference policy, respectively. For a preferred–dispreferred report pair (y_w, y_l) conditioned on x, standard DPO (Rafailov et al., 2023) uses the reward margin

$$\Delta R_{\text{DPO}} = \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}. \tag{3}$$

We approximate the intractable conditional log-likelihoods $\log \pi_\theta(y|x)$ and $\log \pi_{\text{ref}}(y|x)$ via their respective ELBOs defined above in Eq. (2). By employing $\mathcal{E}$ as a tractable surrogate for the exact

log-likelihoods in the standard DPO implicit reward in Eq. (3), we define our variational reward margin as:

$$\Delta R_{\text{VF-DPO}} \triangleq [\mathcal{E}(y_w, y_l, x; \pi_\theta) - \mathcal{E}(y_w, y_l, x; \pi_{\text{ref}})] - [\mathcal{E}(y_l, y_w, x; \pi_\theta) - \mathcal{E}(y_l, y_w, x; \pi_{\text{ref}})]. \quad (4)$$

By applying this definition of the ELBO into this margin, we can derive our variational reward margin $\Delta R_{\text{VF-DPO}}$ as follows.

Proposition 1 *For a pair of preferred and dispreferred reports (y_w, y_l), we respectively represent their corresponding latent variables as b_w and b_l. The variational reward margin $\Delta R_{\text{VF-DPO}}$ defined in Eq.* (4) *takes the following formulation:*

$$\begin{aligned}\Delta R_{\text{VF-DPO}} = \mathbb{E}_{b_w \sim \tilde{\pi}(b_w|y_w, y_l, x)} \left[\log \frac{\pi_\theta(y_w|b_w, x)}{\pi_{\text{ref}}(y_w|b_w, x)} + \log \frac{\pi_\theta(b_w|x)}{\pi_{\text{ref}}(b_w|x)} \right] \\ - \mathbb{E}_{b_l \sim \tilde{\pi}(b_l|y_l, y_w, x)} \left[\log \frac{\pi_\theta(y_l|b_l, x)}{\pi_{\text{ref}}(y_l|b_l, x)} + \log \frac{\pi_\theta(b_l|x)}{\pi_{\text{ref}}(b_l|x)} \right].\end{aligned} \quad (5)$$

The proof is given in Appendix A. The formulation in Proposition 1 decouples the reward margin into the constraints of conditional report generation and finding-aware latent generation. To compute the margin $\Delta R_{\text{VF-DPO}}$ in Eq. (5), we parameterize the latent variables as structured sequences of independent finding-aware QA pairs. By the design of our shared variational distribution $\tilde{\pi}(b|y, \hat{y}, x)$ (detailed in Sect. 3.2), the latent variables for the preference pairs are constructed as $b_w = \{(q_i, a_{i,w})\}_{i \in \mathcal{C}}$ and $b_l = \{(q_i, a_{i,l})\}_{i \in \mathcal{C}}$, where $\mathcal{C}$ denotes the set of finding indices. b_w and b_l share an identical set of finding inquiries $\{q_i\}_{i \in \mathcal{C}}$ anchored to the input image x.

Under the conditional independence assumption, the joint probability of the latent variables factorizes as $\pi(b_w|x) = \prod_{i \in \mathcal{C}} \pi(q_i|x)\pi(a_{i,w}|x, q_i)$, and similarly for b_l. By substituting these factorizations into the reward margin $\Delta R_{\text{VF-DPO}}$ in Eq. (5), the shared inquiry generation terms $\log \frac{\pi_\theta(q_i|x)}{\pi_{\text{ref}}(q_i|x)}$ cancel out during the subtraction. This simplifies the optimization objective to focus on the finding state diagnoses generation and conditional report generation, yielding the following margin:

$$\begin{aligned}\Delta R_{\text{VF-DPO}} = \mathbb{E}_{b_w \sim \tilde{\pi}(b_w|y_w, y_l, x)} \left(\log \frac{\pi_\theta(y_w|b_w, x)}{\pi_{\text{ref}}(y_w|b_w, x)} + \sum_{i \in \mathcal{C}} \log \frac{\pi_\theta(a_{i,w}|x, q_i)}{\pi_{\text{ref}}(a_{i,w}|x, q_i)} \right) \\ - \mathbb{E}_{b_l \sim \tilde{\pi}(b_l|y_l, y_w, x)} \left(\log \frac{\pi_\theta(y_l|b_l, x)}{\pi_{\text{ref}}(y_l|b_l, x)} + \sum_{i \in \mathcal{C}} \log \frac{\pi_\theta(a_{i,l}|x, q_i)}{\pi_{\text{ref}}(a_{i,l}|x, q_i)} \right).\end{aligned} \quad (6)$$

Our VF-DPO alignment strategy guides the model to generate the finding state diagnoses alongside the final reports. Specifically, given the input image, it promotes the likelihood of predicting the correct latent finding states. Concurrently, it encourages the generation of clinically preferred reports y_w conditioned on the input image and preferred anchors b_w, while reducing the likelihood of generating dispreferred reports y_l conditioned on the same image and dispreferred anchors b_l.

To preserve generation capability, we incorporate the standard DPO reward ΔR_{DPO} in Eq. (3) as a marginal regularizer to align the marginal report generation directly conditioned on the input x. Finally, we sum the VF-DPO reward margin with the standard DPO reward margin, and construct our final joint loss function as:

$$\mathcal{L}_{\text{Final}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma \Big(\beta_1 \Delta R_{\text{VF-DPO}} + \beta_2 \Delta R_{\text{DPO}} \Big) \right], \quad (7)$$

where $\mathcal{D}$ is the preference dataset of preference triples (x, y_w, y_l), σ is the sigmoid function. β_1 and β_2 are hyperparameters used to balance the two reward components. By optimizing this objective, the model is guided towards correct finding state diagnosis description while preserving its inherent capability to generate linguistically coherent medical reports. During the training phase, we iteratively sample preference data (x, y_w, y_l) from the dataset $\mathcal{D}$ to construct the finding-aware latent anchors and compute the joint loss in Eq. (7) for optimization. The detailed training algorithm of VF-DPO is provided in Appendix B.

3.2 Preference data construction

In this section, we detail the preference data construction that supports the bi-level alignment in VF-DPO. For each radiology image x, the preference dataset consists of two coupled components: the report level and finding-anchored latent level.

Report level preference data construction. For a given radiology image x, we adopt its corresponding ground-truth report from the dataset as the preferred report y_w. To construct a clinically relevant hard negative, we retrieve a dispreferred report y_l that is linguistically analogous yet clinically contradictory to the ground-truth preferred report y_w following (Liu et al., 2025). Specifically, we first retrieve K candidate reports from the training corpus with high cosine similarity to y_w via a text encoder (e.g., BiomedVLP (Boecking et al., 2022)). Next, we obtain a multi-dimensional finding state vector (e.g., using CheXbert (Smit et al., 2020) for X-ray) for each candidate. We then designate the candidate that maximizes the label Jaccard distance from y_w as y_l.

Finding-anchored latent level preference data construction. To realize our VF-DPO, we introduce the variational distribution $\tilde{\pi}(b|y, \hat{y}, x)$ to construct the finding-anchored latent variable. We identify the finding state diagnostic differences between y and $\hat{y}$ to construct the finding inquiry set $\{q_i\}_{i\in\mathcal{C}}$. Based on each finding inquiry q_i, we obtain the finding state answer a_i based on y. The preference counterparts y_w and y_l share the same finding inquiry set $\{q_i\}_{i\in\mathcal{C}}$. The answers $a_{i,w}$ and $a_{i,l}$ for each finding are respectively extracted from y_w and y_l. Specifically, we first extract the state of N predefined findings to form state vectors $s_w, s_l \in \mathcal{S}^N$. We instantiate the difference set $\mathcal{C} = \{i \mid s_w^i \neq s_l^i, i \in \{1, \ldots, N\}\}$ in Eq. (6), where $|\mathcal{C}|$ represents the number of finding-aware QA pairs. For each finding index $i \in \mathcal{C}$, we obtain its category name and sample a clinical inquiry q_i, along with its respective answers $a_{i,w}$ and $a_{i,l}$, from predefined templates (detailed in Appendix C). Under this construction, $\tilde{\pi}(b|y, \hat{y}, x)$ concentrates its entire probability mass on these identified differences. The latent preference pairs used to evaluate the expectations in Eq. (6) are deterministically instantiated as:

$$b_w = \{(q_i, a_{i,w})\}_{i\in\mathcal{C}}, \quad b_l = \{(q_i, a_{i,l})\}_{i\in\mathcal{C}}. \tag{8}$$

By utilizing such constructed preference pairs at the finding-anchored latent level to optimize Eq. (7), our method simultaneously increases the probability of generating correct finding states and decreases the probability of producing erroneous finding state descriptions.

4 EXPERIMENTS

Datasets. We evaluate our models on a chest X-ray (CXR) data set MIMIC-CXR (Johnson et al., 2019) and a CT dataset CT-RATE (Hamamci et al.). MIMIC-CXR (Johnson et al., 2019) is a large dataset of CXR images and paired reports, which has 270,790, 2,130, and 3,858 samples for training, validation, and testing, respectively. CT-RATE (Hamamci et al.) includes 25,692 non-contrast chest CT volumes with corresponding reports of 21,304 patients. We standardize all CT volumes with the processing following (Chen et al., 2025b;a). We use the official training set (24,128 volumes/20,000 patients) for training, and a testing set (1,564 volumes/1,304 patients).

Evaluation metrics. We evaluate our method across three dimensions: (1) NLG metrics (BLEU, METEOR, ROUGE-L) to assess linguistic fidelity; (2) Clinical Efficacy (CE) metrics (Precision, Recall, F1) derived via CheXbert (Smit et al., 2020) for X-ray and RadBERT (Yan et al., 2022) for CT scans, which quantify diagnostic consistency of the generated report with the ground truth; and (3) RRG-exclusive metrics for comprehensive domain-specific assessment, namely evaluating clinical entity relationships (RadGraph (Jain et al., 2021)), factual correctness (RaTEScore (Zhao et al., 2024)), and clinical error estimation (RadCliQ (Yu et al., 2023)).

Implementation details. The experiments are conducted with PyTorch on an NVIDIA Tesla A100 GPU with 80GB memory. To initialize π_{ref} and π_θ, we obtained the Supervised Fine-Tuning (SFT) parameters for the RRG models either by reimplement using the publicly available training codes (Zhou et al., 2025; Chen et al., 2025b; 2020; 2025a) or by directly utilizing the checkpoints (Jin et al., 2024; Bu et al., 2024) released by the respective authors. As for LLaVA-Med v1.5 (Li et al., 2023), we obtained its SFT parameters for RRG by fine-tuning the publicly available generalist model via Low-Rank Adaptation (LoRA) for 10 epochs using a standard next-token prediction objective. For preference optimization, the optimizer is AdamW with the learning rate and weight decay set to 10^{-7}. The model is trained for one epoch with our training objective. The batch size is 16 for CXRs and 1 for CT. We only finetune the text decoders for preference optimization following (Liu et al., 2025). The number of retrieved reports is empirically set to 20 following the RRG-DPO (Liu et al., 2025). For the weights in Eq. (7), we set $\beta_1 = 0.01$ and $\beta_2 = 0.1$. We use CheXbert (Smit et al., 2020) to extract finding state extraction for 14 findings following (Liu et al., 2025) in X-ray and use the

Table 1: Report generation performance on 2D MIMIC-CXR. The best and second best are respectively bolded and underlined. * denotes a significant improvement over the marked method under the same backbone, determined by the Wilcoxon signed-rank test with $p < 0.05$.

Methods	NLG metrics						CE metrics			RRG-Exclusive metrics		
	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
R2Gen (Chen et al., 2020)	0.3352*	0.2096*	0.1424*	0.1028*	0.1350*	0.2777*	0.3872*	0.2519*	0.2828*	0.1757*	1.2033*	0.5349*
+ DPO (Rafailov et al., 2023)	0.3847*	0.2371*	0.1593*	0.1134*	0.1566*	0.2815*	0.4218*	0.3494*	0.3572*	0.1832*	1.1896*	0.5345*
+ OTPO (Li et al., 2025)	0.3943*	0.2450*	0.1657*	0.1179*	0.1543*	0.2800*	0.4321*	0.3803*	0.3791*	0.1902*	1.1603*	0.5407*
+ TI-DPO (Yang et al., 2026)	0.3882*	0.2394*	0.1599*	0.1131*	0.1478*	0.2798*	0.4179*	0.3355*	0.3479*	0.1805*	1.1890*	0.5395*
+ MMedPO (Zhu et al., 2025)	0.3436*	0.2151*	0.1464*	0.1059*	0.1369*	0.2791*	0.3728*	0.2412*	0.2718*	0.1741*	1.2063*	0.5341*
+ RRG-DPO (Liu et al., 2025)	0.4021*	0.2500*	0.1696*	0.1216*	0.1542*	0.2842*	0.4092*	0.3539*	0.3563*	0.1889*	1.1590*	0.5420*
+ Ours	**0.4046**	**0.2562**	**0.1757**	**0.1258**	**0.1608**	**0.2956**	**0.4489**	**0.3999**	**0.3968**	**0.2097**	**1.0832**	**0.5675**
PromptMRG (Jin et al., 2024)	0.3941	0.2370	0.1544	0.1074	**0.1555**	0.2673*	**0.5024**	0.5078*	0.4751*	0.1908	1.1725*	0.5262
+ DPO (Rafailov et al., 2023)	0.3956	0.2381	0.1547	0.1064	0.1550	0.2688	0.4968	0.5136*	0.4749*	0.1913	1.1626	0.5262*
+ OTPO (Li et al., 2025)	0.3937	0.2372	0.1541	0.1059	0.1554	0.2679	0.4980	0.5138*	0.4759	0.1914	1.1645	0.5261*
+ TI-DPO (Yang et al., 2026)	0.3957	0.2385	0.1553	0.1070	0.1541	0.2690	0.4971	0.5147*	0.4760*	0.1906	1.1636	0.5254*
+ MMedPO (Zhu et al., 2025)	0.3865*	0.2302*	0.1480*	0.1006*	0.1447*	0.2593*	0.4829*	0.4310*	0.4261*	0.1727*	1.2430*	0.5181*
+ RRG-DPO (Liu et al., 2025)	**0.3979**	**0.2405**	**0.1567**	**0.1080**	0.1542	**0.2695**	0.5011	0.5143*	0.4774	0.1913	**1.1607**	0.5262*
+ Ours	0.3978	0.2396	0.1558	0.1073	0.1545	0.2689	0.4982	**0.5203**	**0.4795**	**0.1920**	1.1612	**0.5279**
EKAGen (Bu et al., 2024)	**0.4180**	**0.2571**	0.1685	0.1176	0.1553*	0.2853	0.4607*	0.3720*	0.3848*	0.1994*	1.1315*	0.5308*
+ DPO (Rafailov et al., 2023)	0.3797*	0.2370	0.1577	0.1106	0.1698	0.2848	0.4578*	0.4338*	0.4175*	0.2001*	1.1312*	0.5399*
+ OTPO (Li et al., 2025)	0.3684*	0.2291	0.1518	0.1059	**0.1719**	0.2825	0.4526*	0.4594*	0.4249*	0.2007*	1.1395*	0.5421*
+ TI-DPO (Yang et al., 2026)	0.4088	0.2530	0.1669	0.1168	0.1623	**0.2870**	0.4719*	0.4120*	0.4110*	0.2058*	1.1058	0.5389*
+ MMedPO (Zhu et al., 2025)	0.4173	**0.2571**	**0.1686**	**0.1177**	0.1552*	0.2856	0.4640*	0.3704*	0.3855*	0.1983*	1.1353*	0.5304*
+ RRG-DPO (Liu et al., 2025)	0.3903	0.2408	0.1582	0.1099	0.1688	0.2836	0.4653*	0.4831*	0.4412*	0.2097*	1.1084*	0.5480
+ Ours	0.4087	0.2509	0.1637	0.1130	0.1627	0.2812	**0.4867**	**0.4911**	**0.4582**	**0.2112**	**1.1002**	**0.5484**
REVTAF (Zhou et al., 2025)	**0.4655**	**0.3186**	**0.2356**	**0.1837**	**0.1997**	**0.3345**	0.6173*	0.6163*	0.5900*	0.2880	0.7511*	**0.5889**
+ DPO (Rafailov et al., 2023)	0.4558*	0.3098*	0.2269*	0.1747*	0.1959*	0.3280*	0.6108*	0.6266*	0.5901*	0.2812*	0.7720*	0.5824*
+ OTPO (Li et al., 2025)	0.4614*	0.3158*	0.2325*	0.1797*	0.1972*	0.3318*	0.6169*	0.6294*	0.5955*	0.2866*	0.7528*	0.5842
+ TI-DPO (Yang et al., 2026)	0.4607*	0.3150*	0.2317*	0.1792	0.1975*	0.3317*	0.6170*	0.6259*	0.5941*	0.2863*	0.7543*	0.5839
+ MMedPO (Zhu et al., 2025)	0.4482*	0.3005*	0.2170*	0.1650*	0.1869*	0.3168*	0.6113*	0.5672*	0.5605*	0.2627*	0.8414*	0.5721*
+ RRG-DPO (Liu et al., 2025)	0.4600*	0.3146*	0.2315*	0.1789	0.1969*	0.3315*	0.6136*	0.6279*	0.5934*	0.2850*	0.7572*	0.5840
+ Ours	0.4641	**0.3186**	0.2349	0.1818	0.1981	0.3341	**0.6295**	**0.6350**	**0.6045**	**0.2890**	**0.7438**	0.5852
LLaVA-Med v1.5 (Li et al., 2023)	0.2374*	0.1328*	0.0825*	0.0554*	0.1336*	0.2083*	0.3982*	0.3596*	0.3494*	0.1603*	1.2791*	0.5306*
+ DPO (Rafailov et al., 2023)	0.2717*	0.1519*	0.0934	0.0618	0.1458*	0.2148*	0.3979*	0.3895*	0.3650*	0.1664*	1.2558*	0.5323*
+ OTPO (Li et al., 2025)	0.2743*	0.1538*	0.0946	0.0625	0.1470*	0.2156	0.4069	0.4053*	0.3778*	0.1672*	1.2465	0.5344*
+ TI-DPO (Yang et al., 2026)	0.2513*	0.1408*	0.0870*	0.0581*	0.1387*	0.2111*	0.3965*	0.3712*	0.3558*	0.1606*	1.2667*	0.5308*
+ MMedPO (Zhu et al., 2025)	0.2564*	0.1387*	0.0817*	0.0512*	0.1383*	0.2031*	0.4026	0.3830*	0.3632*	0.1614*	1.2787*	0.5298*
+ RRG-DPO (Liu et al., 2025)	0.2833*	0.1566*	0.0951*	0.0620	0.1490*	0.2151	0.3891*	0.3838*	0.3575*	0.1652*	1.2706*	0.5323*
+ Ours	**0.2878**	**0.1606**	**0.0982**	**0.0643**	**0.1521**	**0.2159**	**0.4114**	**0.4378**	**0.3940**	**0.1691**	**1.2457**	**0.5365**

finding state provided by CT-RATE (Hamamci et al.) for 18 findings in CT for our preference data construction. More training details are shown in Appendix C.

Compared preference alignment methods. To comprehensively evaluate our method, we select a strong set of baselines spanning both general and medical-specific preference optimization techniques. For general methods, we compare against the standard DPO (Rafailov et al., 2023) along with token-level variants, OTPO (Li et al., 2025), and TI-DPO (Yang et al., 2026). Furthermore, we compare against methods MMedPO (Zhu et al., 2025) and state-of-the-art RRG-DPO (Liu et al., 2025), which are specifically tailored for radiology report generation. We evaluate MMedPO (Zhu et al., 2025) exclusively on 2D X-rays, as its local lesion-noising mechanism relies inherently on a 2D pretrained visual grounding model MedKLIP (Wu et al., 2023) to detect finding regions; it does not natively generalize to 3D CT volumes.

4.1 Main results

Our VF-DPO acts as a preference alignment method for different RRG models with SFT for radiology report generation. To compare different preference alignment methods, we apply them to different RRG methods as follows. For 2D CXR images, the experimented methods include classical method R2Gen (Chen et al., 2020), EKAGen (Bu et al., 2024), PromptMRG (Jin et al., 2024), open-sourced state-of-the-art REVTAF (Zhou et al., 2025), and LLM-based LLaVA-Med v1.5 7B (Li et al., 2023). For 3D CT volumes, we compare different preference alignment methods based on the ViSD-Boost (Cao et al., 2025) and the state-of-the-art LLaMA2-7B (Touvron et al., 2023) based Dia-LLaMA (Chen et al., 2025b) and Reg2RG (Chen et al., 2025a).

As shown in Tab. 1, we compare our proposed method against various preference alignment methods on the X-ray dataset. Our proposed VF-DPO achieves the best improvement in most evaluation CE metrics and RRG-Exclusive metrics across all five evaluated backbones compared to other preference alignment baselines. The remarkable F1 improvement reveals that our VF-DPO enhances the capacity of the model to generate reports with more consistent finding state descriptions than other compared preference alignment methods. Our method achieves the best RadGraph score across all five evaluated backbones (*e.g.*, reaching 0.2097 on R2Gen and 0.2112 on EKAGen) compared to other alignment methods, which implies that training with VF-DPO enables the model to generate more accurate entities and relations. Notably, for SFT models PromptMRG (Jin et al., 2024) and REVTAF (Zhou

Table 2: Report generation performance on 3D CT-RATE. The best and second best are respectively bolded and underlined. * denotes a significant improvement over the marked method under the same backbone, determined by the Wilcoxon signed-rank test with $p < 0.05$.

Methods	NLG metrics						CE metrics			RRG-Exclusive metrics		
	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
Reg2RG (Chen et al., 2025a)	0.4955*	0.3757*	0.2995*	0.2492	0.4222*	0.3547	0.4036	0.1754*	0.2330*	0.3219	0.7614	0.6554
+ DPO (Rafailov et al., 2023)	0.5011	0.3808	0.3040	0.2533	0.4254*	0.3562	0.3998	0.1830*	0.2394*	0.3240	0.7559	0.6574
+ OTPO (Li et al., 2025)	0.5035	0.3831	0.3065	0.2556	0.4265*	0.3568	0.4098	0.1823*	0.2403*	0.3257	0.7445	0.6572
+ TI-DPO (Yang et al., 2026)	0.5020	0.3836	0.3085	0.2587	0.4284*	**0.3605**	0.4014	0.1805*	0.2396*	0.3272	0.7448	0.6577
+ RRG-DPO (Liu et al., 2025)	0.5028	0.3840	0.3083	0.2581	0.4278	0.3593	**0.4158**	0.1856*	0.2460*	0.3259	0.7439	**0.6580**
+ Ours	**0.5132**	**0.3914**	**0.3140**	**0.2627**	**0.4340**	0.3580	0.4081	**0.2040**	**0.2576**	**0.3274**	**0.7417**	0.6577
Dia-LLaMA (Chen et al., 2025b)	0.5339*	0.4147*	0.3382*	0.2874*	0.4530*	0.3861	**0.4360**	0.3951*	0.3630*	0.3602*	0.6730*	0.6757
+ DPO (Rafailov et al., 2023)	0.5347*	0.4131*	0.3348*	0.2828*	0.4491*	0.3790*	0.4311	0.4078	0.3703	0.3549*	0.6897*	0.6744*
+ OTPO (Li et al., 2025)	0.5360*	0.4148*	0.3369*	0.2852*	0.4509*	0.3821*	0.4315	0.4086	0.3706	0.3550*	0.6891*	0.6744*
+ TI-DPO (Yang et al., 2026)	0.5343*	0.4133*	0.3356*	0.2840*	0.4488*	0.3799*	0.4197	0.4111*	0.3698*	0.3551*	0.6900*	0.6732*
+ RRG-DPO (Liu et al., 2025)	0.5381*	0.4186*	0.3416*	0.2905*	0.4544*	0.3868	0.4251	0.4103*	0.3715*	0.3611	0.6702	0.6760
+ Ours	**0.5429**	**0.4230**	**0.3457**	**0.2941**	**0.4602**	**0.3883**	0.4187	**0.4231**	**0.3766**	**0.3630**	**0.6632**	**0.6777**
ViSD-Boost (Cao et al., 2025)	0.5167*	0.3919*	0.3142	0.2628	0.4377*	0.3630	0.4996*	0.3008*	0.3304*	0.3464	0.7080	0.6795
+ DPO (Rafailov et al., 2023)	0.5169	0.3920	0.3143	0.2629	0.4378*	**0.3633**	**0.5221**	0.3035*	0.3332*	0.3465	0.7070	0.6796
+ OTPO (Li et al., 2025)	**0.5200**	0.3934*	0.3145	0.2622	0.4377	0.3599	0.5112	0.3190*	0.3425*	0.3446*	0.7151	0.6782*
+ TI-DPO (Yang et al., 2026)	0.5161	0.3913*	0.3135*	0.2619*	0.4368*	0.3626	0.5196	0.3208*	0.3485*	0.3465*	0.7093	0.6796
+ RRG-DPO (Liu et al., 2025)	0.5167*	0.3919*	0.3142	0.2628	0.4375*	0.3626	0.5020	0.3078*	0.3352*	0.3463*	0.7100	0.6793*
+ Ours	0.5196	**0.3940**	**0.3154**	**0.2633**	**0.4387**	0.3624	0.5000	**0.3484**	**0.3644**	**0.3475**	**0.7057**	**0.6800**

Table 3: Ablation study results on MIMIC-CXR (Johnson et al., 2019) using the LLaVA-Med v1.5 model (Li et al., 2023).

ΔR_{DPO}	$\Delta R_{\text{VF-DPO}}$	NLG metrics						CE metrics			RRG-Exclusive metrics		
		Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
LLaVA-Med v1.5 (Li et al., 2023)		0.2374	0.1328	0.0825	0.0554	0.1336	0.2083	0.3982	0.3596	0.3494	0.1603	1.2791	0.5306
✓		0.2717	0.1519	0.0934	0.0618	0.1458	0.2148	0.3979	0.3895	0.3650	0.1664	1.2558	0.5323
	✓	0.2786	0.1553	0.0954	0.0634	0.1491	0.2146	0.3947	0.4294	0.3830	0.1673	1.2512	0.5309
✓	✓	**0.2878**	**0.1606**	**0.0982**	**0.0643**	**0.1521**	**0.2159**	**0.4114**	**0.4378**	**0.3940**	**0.1691**	**1.2457**	**0.5365**

et al., 2025), which rely on image-based finding state prediction for report prompting, our method further enhances CE metrics Recall and F1 (*e.g.*, improves F1 of REVTAF from 0.5900 to 0.6045). These results indicate that VF-DPO increases the ability of the model to generate better finding status descriptions. When applying our method to the classical SFT method R2Gen (Chen et al., 2020) and multi-modal models LLaVA-Med v1.5 (Li et al., 2023), our method achieves the best improvement for all the evaluated metrics. These results show that VF-DPO generates finding descriptions more consistent with the ground truth while preserving comparable linguistic fluency.

We compare our method with other preference alignment methods on the CT-RATE (Hamamci et al.) shown in Tab. 2. Our method achieves the highest scores across most evaluation metrics. Compared with the other methods, our VF-DPO alignment strategy consistently obtains the highest F1 scores. It increases the F1 from 0.2330 to 0.2576 for Reg2RG (Chen et al., 2025a). Dia-LLaMA (Chen et al., 2025b) incorporates finding diagnostic information as guidance prompts, based on which, the original DPO improves the F1 from 0.3630 to 0.3703, but degrades performance across all three RRG-Exclusive metrics. Our VF-DPO improves F1 score of Dia-LLaMA from 0.3630 to 0.3766. It also improves the score of three RRG-Exclusive metrics. Moreover, on the third CT backbone, ViSD-Boost (Cao et al., 2025), our method achieves strong performance across most evaluation metrics. In particular, VF-DPO outperforms the other preference-alignment methods on this backbone, achieving the highest Recall of 0.3484, F1 score of 0.3644, RadGraph score of 0.3475, and RaTEScore of 0.6800. These performance gains stem from the intrinsic design of our VF-DPO. Specifically, our method not only encourages the generation of full reports conditioned on correct finding state descriptions but also penalizes the generation of dispreferred reports when prompted with incorrect dispreferred finding state anchors. These results demonstrate that VF-DPO is effective at improving overall clinical consistency with reference reports.

4.2 Ablation Study

Ablation of each component of our method. We evaluate the contribution of each component of our training loss in Eq. (7) shown in Tab. 3. Standard DPO (ΔR_{DPO} only) improves linguistic fluency and yields gains in most clinical metrics over the LLaVA-Med v1.5 (Li et al., 2023). In contrast, our finding-anchored latent preference alignment ($\Delta R_{\text{VF-DPO}}$ only) outperforms standard DPO across most metrics, *e.g.*, improving the F1 score of standard DPO from 0.3650 to 0.3830, decreasing the RadCliQ from 1.2558 to 1.2512. This indicates that our finding-anchored latent preference alignment

Table 4: Ablation study on SFT and preference alignment tasks using LLaVA-Med v1.5 model (Li et al., 2023) on MIMIC-CXR (Johnson et al., 2019). RG: report generation; VQA: visual question answering for finding diagnostics. Alignment targets include standard report generation $\pi(y|x)$, finding-aware latent generation $\pi(b|x)$, and latent-anchored report generation $\pi(y|b,x)$.

SFT Tasks		Preference Alignment Targets			NLG metrics						CE metrics			RRG-Exclusive metrics		
RG	VQA	$\pi(y\|x)$	$\pi(b\|x)$	$\pi(y\|b,x)$	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
✓					0.2374	0.1328	0.0825	0.0554	0.1336	0.2083	0.3982	0.3596	0.3494	0.1603	1.2791	0.5306
✓	✓				0.2195	0.1222	0.0760	0.0512	0.1293	0.2043	0.4049	0.3628	0.3542	0.1610	1.2788	0.5282
✓		✓			0.2717	0.1519	0.0934	0.0618	0.1458	0.2148	0.3979	0.3895	0.3650	0.1664	1.2558	0.5323
✓	✓	✓			0.2713	0.1522	0.0943	0.0631	0.1457	0.2154	0.4018	0.4061	0.3741	0.1648	1.2538	0.5316
✓		✓	✓		0.2728	0.1514	0.0916	0.0592	0.1452	0.2138	0.4016	0.3981	0.3707	0.1670	1.2565	0.5363
✓		✓	✓	✓	**0.2878**	**0.1606**	**0.0982**	**0.0643**	**0.1521**	**0.2159**	**0.4114**	**0.4378**	**0.3940**	**0.1691**	**1.2457**	**0.5365**

enhances the diagnostic fidelity of the model. The joint final objective, utilizing both ΔR_{DPO} and $\Delta R_{\text{VF-DPO}}$ achieves the best performance across all metrics. These results show that our complete method enhances finding state prediction while maintaining linguistic fluency.

Ablation on training strategies for preference optimization. As shown in Tab. 4, we explicitly explore the role of finding state diagnostics VQA in enhancing accurate clinical descriptions for report generation. The results in the second row demonstrate that introducing the VQA task during the SFT phase improves clinical effectiveness (*e.g.*, F1 increases from 0.3494 to 0.3542) compared to the RG-only SFT baseline, but this degrades the NLG metrics. Applying standard DPO to the model trained with multi-task SFT in the fourth row achieves a better F1 score than applying it to the RG-only SFT baseline shown in the third row. However, it still leads to slight drops in metrics like RaTEScore and RadGraph. Instead of multi-tasking during SFT, we also investigate the impact of $\pi(b|x)$ for preference alignment. The fifth row shows that jointly yet independently aligning preferences for report generation $\pi(y|x)$ and finding-aware latent generation $\pi(b|x)$ boosts the performance of CE metrics, RadGraph, and RaTEScore metrics compared to DPO in the third row. However, optimizing these targets in parallel fails to explicitly enforce consistency between the diagnostic states and the generated text. By further incorporating the latent-anchored alignment target $\pi(y|b,x)$, our full VF-DPO (last row) constrains the generation process to be conditioned on the finding states. Our VF-DPO strategy achieves the best CE, RRG-exclusive, and NLG scores among the evaluated variants. As shown in Tab. 4, while integrating VQA during the SFT phase followed by standard DPO improves CE metrics, this decoupled strategy remains inferior to our proposed bi-level alignment for explicitly enforcing diagnostic consistency during the preference alignment stage.

Due to the space limit, more empirical analysis, *e.g.*, the results of generalization beyond the QA-supervised finding categories, comparison with auxiliary finding classification, sensitivity to hyperparameters, computational cost, visualization comparison of generated reports, etc., are in Appendix D.

5 Conclusion and Limitation

In this paper, we propose Variational Finding-Anchored Direct Preference Optimization (VF-DPO) to improve overall clinical consistency beyond the limitations of pure sequence-level alignment. We introduce finding-aware Question-Answer pairs as latent variables to anchor the generation process. By reformulating the preference optimization via the Evidence Lower Bound (ELBO), we derive a bi-level objective that simultaneously enforces latent diagnostic prediction and high-fidelity report generation. Experiments on MIMIC-CXR and CT-RATE demonstrate that VF-DPO improves clinical finding coverage and overall clinical consistency with reference reports.

One limitation is that our latent representation is anchored on finding-specific attributes, making the paradigm domain-specific to medical diagnosis. Extending it to other domains would require defining appropriate key attributes for anchoring. Another limitation is that our VF-DPO framework consumes around three times in training time compared with DPO, arising from the bi-level optimization objective (finding-anchored latent variables alignment and full report generation alignment). This is an overhead exclusive to the training phase, while VF-DPO generates reports with the same complexity as standard DPO in testing. In the future, we plan to extend this finding-anchored alignment paradigm to other clinical tasks and broader domains beyond medical applications.

AI USE STATEMENT

We used generative AI tools only to improve the language and readability of the manuscript. All revisions were reviewed by the authors to ensure that they preserved the intended meaning. The authors take full responsibility for the final content of the paper.

REFERENCES

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Oishi Banerjee, Hong-Yu Zhou, Subathra Adithan, Stephen Kwak, Kay Wu, and Pranav Rajpurkar. Direct preference optimization for suppressing hallucinated prior exams in radiology report generation. *arXiv preprint arXiv:2406.06496*, 2024.

Benedikt Boecking, Naoto Usuyama, Shruthi Bannur, Daniel C Castro, Anton Schwaighofer, Stephanie Hyland, Maria Wetscherek, Tristan Naumann, Aditya Nori, Javier Alvarez-Valle, et al. Making the most of text semantics to improve biomedical vision–language processing. In *ECCV*, 2022.

Shenshen Bu, Taiji Li, Yuedong Yang, and Zhiming Dai. Instance-level expert knowledge and aggregate discriminative attention for radiology report generation. In *CVPR*, 2024.

Weiwei Cao, Jianpeng Zhang, Zhongyi Shui, Sinuo Wang, Zeli Chen, Xi Li, Le Lu, Xianghua Ye, Qi Zhang, Tingbo Liang, et al. Boosting vision semantic density with anatomy normality modeling for medical vision-language pre-training. In *ICCV*, 2025.

Zhihong Chen, Yan Song, Tsung-Hui Chang, and Xiang Wan. Generating radiology reports via memory-driven transformer. In *EMNLP*, 2020.

Zhihong Chen, Yaling Shen, Yan Song, and Xiang Wan. Cross-modal memory networks for radiology report generation. In *ACL-IJCNLP*, 2021.

Zhixuan Chen, Yequan Bie, Haibo Jin, and Hao Chen. Large language model with region-guided referring and grounding for ct report generation. *IEEE Trans. MI*, 44(8):3139–3150, 2025a.

Zhixuan Chen, Luyang Luo, Yequan Bie, and Hao Chen. Dia-llama: Towards large language model-driven ct report generation. In *MICCAI*, 2025b.

Sheng Cheng and Devika Subramanian. Rethinking radiology report generation: From narrative flow to topic-guided findings. In *ICLR*, 2026.

Zhuocheng Gong, Jian Guan, Wei Wu, Huishuai Zhang, and Dongyan Zhao. Latent preference coding: Aligning large language models via discrete latent codes. In *ICML*, 2025.

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

Ibrahim Ethem Hamamci, Sezgin Er, Furkan Almas, Ayse Gulnihan Simsek, Sevval Nil Esirgun, Irem Dogan, Muhammed Furkan Dasdelen, Bastian Wittmann, Enis Simsar, Mehmet Simsar, et al. A foundation model utilizing chest ct volumes and radiology reports for supervised-level zero-shot detection of abnormalities. *arXiv preprint arXiv:2403.17834*.

Dennis Hein, Zhihong Chen, Sophie Ostmeier, Justin Xu, Maya Varma, Eduardo Pontes Reis, Arne Edward Michalson, Christian Bluethgen, Hyun Joo Shin, Curtis Langlotz, et al. Chexalign: Preference fine-tuning in chest x-ray interpretation models without human feedback. *arXiv preprint arXiv:2410.07025*, 2024.

Xiang Huang, Ting-En Lin, Feiteng Fang, Yuchuan Wu, Hangyu Li, Yuzhong Qu, Fei Huang, and Yongbin Li. Reverse preference optimization for complex instruction following. In *ACL*, 2025.

Zhongzhen Huang, Xiaofan Zhang, and Shaoting Zhang. Kiut: Knowledge-injected u-transformer for radiology report generation. In *CVPR*, 2023.

Saahil Jain, Ashwin Agrawal, Adriel Saporta, Steven QH Truong, Du Nguyen Duong, Tan Bui, Pierre Chambon, Yuhao Zhang, Matthew P Lungren, Andrew Y Ng, et al. Radgraph: Extracting clinical entities and relations from radiology reports. *arXiv preprint arXiv:2106.14463*, 2021.

Haibo Jin, Haoxuan Che, Yi Lin, and Hao Chen. Promptmrg: Diagnosis-driven prompts for medical report generation. In *AAAI*, 2024.

Alistair EW 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.

Haoran Lai, Zihang Jiang, Qingsong Yao, Rongsheng Wang, Zhiyang He, Xiaodong Tao, Weifu Lv, Wei Wei, and Shaohua Kevin Zhou. Bridged semantic alignment for zero-shot 3d medical image diagnosis. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–14, 2025.

Xin Lai, Zhuotao Tian, Yukang Chen, Senqiao Yang, Xiangru Peng, and Jiaya Jia. Step-dpo: Step-wise preference optimization for long-chain reasoning of llms. *arXiv preprint arXiv:2406.18629*, 2024.

Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. In *NeurIPS*, 2023.

Meng Li, Guangda Huzhang, Haibo Zhang, Xiting Wang, and Anxiang Zeng. Optimal transport-based token weighting scheme for enhanced preference optimization. In *ACL*, 2025.

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 *ECCV*, 2024.

Mingquan Lin, Gregory Holste, Song Wang, Yiliang Zhou, Yishu Wei, Imon Banerjee, Pengyi Chen, Tianjie Dai, Yuexi Du, Nicha C Dvornek, et al. Cxr-lt 2024: A miccai challenge on long-tailed, multi-label, and zero-shot disease classification from chest x-ray. *Medical Image Analysis*, pp. 103739, 2025.

Chang Liu, Yuanhe Tian, Weidong Chen, Yan Song, and Yongdong Zhang. Bootstrapping large language models for radiology report generation. In *AAAI*, 2024.

Hong Liu, Dong Wei, Zhe Xu, Xian Wu, Yefeng Zheng, and Liansheng Wang. Rrg-dpo: Direct preference optimization for clinically accurate radiology report generation. In *MICCAI*, 2025.

Kang Liu, Zhuoqi Ma, Zikang Fang, Yunan Li, Kun Xie, and Qiguang Miao. Priorrg: Prior-guided contrastive pre-training and coarse-to-fine decoding for chest x-ray report generation. In *AAAI*, 2026.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *NeurIPS*, 2023.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Hongyu Shen, Mingtao Pei, Juncai Liu, and Zhaoxing Tian. Automatic radiology reports generation via memory alignment network. In *AAAI*, 2024.

Akshay Smit, Saahil Jain, Pranav Rajpurkar, Anuj Pareek, Andrew Y Ng, and Matthew Lungren. Combining automatic labelers and expert annotations for accurate radiology report labeling using bert. In *EMNLP*, 2020.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *CVPR*, 2024.

Jun Wang, Abhir Bhalerao, and Yulan He. Cross-modal prototype driven network for radiology report generation. In *ECCV*, 2022.

Zhanyu Wang, Lingqiao Liu, Lei Wang, and Luping Zhou. R2gengpt: Radiology report generation with frozen llms. *Meta-Radiology*, 1(3):100033, 2023.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. Medklip: Medical knowledge enhanced language-image pre-training for x-ray diagnosis. In *ICCV*, 2023.

Ting Xiao, Lei Shi, Peng Liu, Zhe Wang, and Chenjia Bai. Radiology report generation via multi-objective preference optimization. In *AAAI*, 2025.

Yuxi Xie, Guanzhen Li, Xiao Xu, and Min-Yen Kan. V-dpo: Mitigating hallucination in large vision language models via vision-guided direct preference optimization. In *EMNLP*, 2024.

An Yan, Julian McAuley, Xing Lu, Jiang Du, Eric Y Chang, Amilcare Gentili, and Chun-Nan Hsu. Radbert: adapting transformer-based language models to radiology. *Radiology: Artificial Intelligence*, 4(4):e210258, 2022.

Ning Yang, Hai Lin, Yibo Liu, Baoliang Tian, Guoqing Liu, and Haijun Zhang. Token-importance guided direct preference optimization. In *ICLR*, 2026.

Shuxin Yang, Xian Wu, Shen Ge, S Kevin Zhou, and Li Xiao. Knowledge matters: Chest radiology report generation with general and specific knowledge. *MIA*, 80:102510, 2022.

Shuxin Yang, Xian Wu, Shen Ge, Zhuozhao Zheng, S Kevin Zhou, and Li Xiao. Radiology report generation with a learned knowledge base and multi-modal alignment. *MIA*, 86:102798, 2023.

Zhihe Yang, Xufang Luo, Dongqi Han, Yunjian Xu, and Dongsheng Li. Mitigating hallucinations in large vision-language models via dpo: On-policy data hold the key. In *CVPR*, 2025.

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*, pp. 4096–4108, 2025.

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, et al. Evaluating progress in automatic chest x-ray radiology report generation. *Patterns*, 4(9), 2023.

Yongcheng Zeng, Guoqing Liu, Weiyu Ma, Ning Yang, Haifeng Zhang, and Jun Wang. Token-level direct preference optimization. In *ICML*, 2024.

Kechi Zhang, Ge Li, Jia Li, Yihong Dong, Jia Li, and Zhi Jin. Focused-DPO: Enhancing code generation through focused preference optimization on error-prone points. In *ACL*, 2025.

Weike Zhao, Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. Ratescore: A metric for radiology report generation. In *EMNLP*, 2024.

Qin Zhou, Guoyan Liang, Xindi Li, Jingyuan Chen, Zhe Wang, Chang Yao, and Sai Wu. Learnable retrieval enhanced visual-text alignment and fusion for radiology report generation. In *ICCV*, 2025.

Kangyu Zhu, Peng Xia, Yun Li, Hongtu Zhu, Sheng Wang, and Huaxiu Yao. Mmedpo: Aligning medical vision-language models with clinical-aware multimodal preference optimization. In *ICML*, 2025.

APPENDIX A PROOF

APPENDIX A.1 PROOF OF THE ELBO IN EQ. (2)

Proof. We provide the detailed derivation of the Evidence Lower Bound (ELBO) for the marginal log-likelihood $\log \pi(y|x)$ in Eq. (2). By introducing the latent variable b and a variational distribution $\tilde{\pi}(b|y, \hat{y}, x)$, the derivation proceeds as follows:

$$
\begin{aligned}
\log \pi(y|x) &= \log \int \pi(y, b|x) db \\
&= \log \int \pi(y|b, x)\pi(b|x) db \\
&= \log \int \tilde{\pi}(b|y, \hat{y}, x) \frac{\pi(y|b, x)\pi(b|x)}{\tilde{\pi}(b|y, \hat{y}, x)} db \\
&= \log \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\frac{\pi(y|b, x)\pi(b|x)}{\tilde{\pi}(b|y, \hat{y}, x)} \right] \\
&\geq \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\log \left(\frac{\pi(y|b, x)\pi(b|x)}{\tilde{\pi}(b|y, \hat{y}, x)} \right) \right] \quad \text{(By Jensen's Inequality)} \\
&= \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} [\log \pi(y|b, x)] + \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\log \frac{\pi(b|x)}{\tilde{\pi}(b|y, \hat{y}, x)} \right] \\
&= \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} [\log \pi(y|b, x)] - \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\log \frac{\tilde{\pi}(b|y, \hat{y}, x)}{\pi(b|x)} \right] \\
&= \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} [\log \pi(y|b, x)] - \mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi(b|x)) \qquad \text{(A-1)}
\end{aligned}
$$

■

APPENDIX A.2 PROOF OF PROPOSITION 1

Proposition A-1 *For a pair of preferred and dispreferred reports (y_w, y_l), we respectively represent their corresponding latent variables as b_w and b_l. The variational reward margin $\Delta R_{\text{VF-DPO}}$ defined in Eq.* (4) *takes the following formulation:*

$$
\begin{aligned}
\Delta R_{\text{VF-DPO}} = \mathbb{E}_{b_w\sim\tilde{\pi}(b_w|y_w,y_l,x)} &\left[\log \frac{\pi_\theta(y_w|b_w, x)}{\pi_{\text{ref}}(y_w|b_w, x)} + \log \frac{\pi_\theta(b_w|x)}{\pi_{\text{ref}}(b_w|x)} \right] \\
- \mathbb{E}_{b_l\sim\tilde{\pi}(b_l|y_l,y_w,x)} &\left[\log \frac{\pi_\theta(y_l|b_l, x)}{\pi_{\text{ref}}(y_l|b_l, x)} + \log \frac{\pi_\theta(b_l|x)}{\pi_{\text{ref}}(b_l|x)} \right].
\end{aligned} \qquad \text{(A-2)}
$$

Proof. The derivation begins with the definition of the Evidence Lower Bound (ELBO) $\mathcal{E}(y, \hat{y}, x; \pi)$ for a response y and its preference counterpart $\hat{y}$:

$$
\mathcal{E}(y, \hat{y}, x; \pi) \triangleq \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} [\log \pi(y|b, x)] - \mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi(b|x)). \qquad \text{(A-3)}
$$

First, we calculate the difference between the ELBO of the policy model π_θ and the reference model π_{ref} for the preference pair $(y, \hat{y})$. We use the definition in Eq. (A-3) and have:

$$
\begin{aligned}
&\mathcal{E}(y, \hat{y}, x; \pi_\theta) - \mathcal{E}(y, \hat{y}, x; \pi_{\text{ref}}) \\
=&\mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\log \frac{\pi_\theta(y|b, x)}{\pi_{\text{ref}}(y|b, x)} \right] - \big(\mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi_\theta(b|x)) - \mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi_{\text{ref}}(b|x)) \big).
\end{aligned} \qquad \text{(A-4)}
$$

Next, we calculate the difference of the KL divergence terms in Eq. (A-4) according to KL definition ($\mathrm{KL}(\tilde{\pi}||\pi) = \mathbb{E}_{\tilde{\pi}}[\log \tilde{\pi} - \log \pi]$), the $\log \tilde{\pi}(b|y, \hat{y}, x)$ terms elegantly cancel out when taking the difference. This simplifies the prior and posterior KL difference terms as:

$$
-\big(\mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi_\theta(b|x)) - \mathrm{KL}(\tilde{\pi}(b|y, \hat{y}, x)||\pi_{\text{ref}}(b|x)) \big) = \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\log \frac{\pi_\theta(b|x)}{\pi_{\text{ref}}(b|x)} \right] \qquad \text{(A-5)}
$$

Algorithm 1: Variational Finding-Anchored Direct Preference Optimization (VF-DPO)

Input: Initial policy π_θ and frozen reference policy π_{ref} (both initialized with SFT parameters θ_0), hyperparameters β_1, β_2, preference dataset $\mathcal{D}$, total iterations N_{iter}
Output: Optimized policy model π_θ

for iter $= 1, 2, \ldots, N_{\text{iter}}$ **do**
  Sample a mini-batch $\mathcal{B}$ from $\mathcal{D}$
  // Finding-anchored latent data construction
  for each instance $(x, y_w, y_l) \in \mathcal{B}$ **do**
    Extract finding state vectors $s_w, s_l \in \mathcal{S}^N$ from reports y_w, y_l
    Obtain the diagnostic conflict set $\mathcal{C} = \{i \mid s_w^i \neq s_l^i, i \in \{1, \ldots, N\}\}$
    Construct finding-aware latent QA pairs $b_w = \{(q_i, a_{i,w})\}_{i\in\mathcal{C}}$ and $b_l = \{(q_i, a_{i,l})\}_{i\in\mathcal{C}}$
  end for
  // Reward margin computation
  for each instance $(x, y_w, y_l) \in \mathcal{B}$ **do**
    Compute variational rewards for preferred report y_w and dispreferred report y_l, alongside with their latent states b_w and b_l:
      $\mathcal{R}_w = \log \frac{\pi_\theta(y_w|b_w,x)}{\pi_{\text{ref}}(y_w|b_w,x)} + \sum_{i\in\mathcal{C}} \log \frac{\pi_\theta(a_{i,w}|x,q_i)}{\pi_{\text{ref}}(a_{i,w}|x,q_i)}$
      $\mathcal{R}_l = \log \frac{\pi_\theta(y_l|b_l,x)}{\pi_{\text{ref}}(y_l|b_l,x)} + \sum_{i\in\mathcal{C}} \log \frac{\pi_\theta(a_{i,l}|x,q_i)}{\pi_{\text{ref}}(a_{i,l}|x,q_i)}$
    Calculate variational reward margin: $\Delta R_{\text{VF-DPO}} = \mathcal{R}_w - \mathcal{R}_l$
    Compute regularized reward margin: $\Delta R_{\text{DPO}} = \log \frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \log \frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}$
    Calculate instance training loss:
      $\mathcal{L}_{(x,y_w,y_l)}(\pi_\theta; \pi_{\text{ref}}) = -\log \sigma\big(\beta_1 \Delta R_{\text{VF-DPO}} + \beta_2 \Delta R_{\text{DPO}}\big)$
  end for
  Compute final batch-level training loss:
    $\hat{\mathcal{L}}_{\text{Final}}(\pi_\theta; \pi_{\text{ref}}) = \frac{1}{|\mathcal{B}|} \sum_{(x,y_w,y_l)\in\mathcal{B}} \mathcal{L}_{(x,y_w,y_l)}(\pi_\theta; \pi_{\text{ref}})$
  Update the model π_θ with $\hat{\mathcal{L}}_{\text{Final}}(\pi_\theta; \pi_{\text{ref}})$ via gradient descent:
end for
return π_θ

Substituting this simplified KL difference in Eq. (A-5) back into the ELBO difference in Eq. (A-4), we obtain a margin for the generic response data pair as:

$$\mathcal{E}(y, \hat{y}, x; \pi_\theta) - \mathcal{E}(y, \hat{y}, x; \pi_{\text{ref}}) = \mathbb{E}_{b\sim\tilde{\pi}(b|y,\hat{y},x)} \left[\log \frac{\pi_\theta(y|b,x)}{\pi_{\text{ref}}(y|b,x)} + \log \frac{\pi_\theta(b|x)}{\pi_{\text{ref}}(b|x)} \right]. \tag{A-6}$$

Finally, by substituting the generic factorized identity from Eq. (A-6) into the definition of $\Delta R_{\text{VF-DPO}}$ in Eq. (4) for both preferred and dispreferred data, along with their corresponding latent variables, we obtain the final expanded form as:

$$\begin{aligned}
\Delta R_{\text{VF-DPO}} &\triangleq [\mathcal{E}(y_w, y_l, x; \pi_\theta) - \mathcal{E}(y_w, y_l, x; \pi_{\text{ref}})] - [\mathcal{E}(y_l, y_w, x; \pi_\theta) - \mathcal{E}(y_l, y_w, x; \pi_{\text{ref}})] \\
&= \mathbb{E}_{b_w\sim\tilde{\pi}(b_w|y_w,y_l,x)} \left[\log \frac{\pi_\theta(y_w|b_w,x)}{\pi_{\text{ref}}(y_w|b_w,x)} + \log \frac{\pi_\theta(b_w|x)}{\pi_{\text{ref}}(b_w|x)} \right] \\
&\quad - \mathbb{E}_{b_l\sim\tilde{\pi}(b_l|y_l,y_w,x)} \left[\log \frac{\pi_\theta(y_l|b_l,x)}{\pi_{\text{ref}}(y_l|b_l,x)} + \log \frac{\pi_\theta(b_l|x)}{\pi_{\text{ref}}(b_l|x)} \right].
\end{aligned} \tag{A-7}$$

■

APPENDIX B TRAINING ALGORITHM

Our training algorithm is shown in the Algorithm 1.

Table C-1: Finding categories utilized for each dataset.

Dataset	Finding Categories
MIMIC-CXR	enlarged cardiomediastinum, cardiomegaly, lung opacity, lung lesion, edema, consolidation, pneumonia, atelectasis, pneumothorax, pleural effusion, pleural other, fracture, support devices, no finding.
CT-RATE	medical material, arterial wall calcification, cardiomegaly, pericardial effusion, coronary artery wall calcification, hiatal hernia, lymphadenopathy, emphysema, atelectasis, lung nodule, lung opacity, pulmonary fibrotic sequela, pleural effusion, mosaic attenuation pattern, peribronchial thickening, consolidation, bronchiectasis, interlobular septal thickening.

Table C-2: Prompt templates for finding-anchored visual question answering.

Question Templates	
Questions	• Does the image show {finding}? • Is {finding} present in this image? • What does the image show about {finding}? • Is there any evidence of {finding}? • Are there any signs of {finding}?
Answer Templates	
Finding Status	**Templates**
Positive	• Yes, the image indicates the presence of {finding}. • There is evidence of {finding} in the image. • The image is positive for {finding}. • Yes, {finding} is observed. • The image shows {finding}.
Negative	• No, there is no evidence of {finding}. • The image is negative for {finding}. • {finding} is not seen in this image. • No, the image does not show the presence of {finding}. • {finding} is absent in this image.
Uncertain	• The image is uncertain regarding {finding}. • It is unclear if {finding} is present based on the image. • {finding} cannot be ruled out. • The evidence for {finding} is unclear. • It is uncertain whether {finding} is present.
No Mention	• The image does not contain {finding}. • There is no information regarding {finding} in this image. • {finding} is not evaluated in this image. • The image provides no information on {finding}. • No assessment of {finding} can be made.

APPENDIX C EXPERIMENTAL DETAILS

In this section, we provide the training details of our methods for constructing the finding-aware QA pairs. Table C-1 presents the finding categories we used for MIMIC-CXR (Johnson et al., 2019) and CT-RATE (Hamamci et al.). Table C-2 presents the templates used to construct the finding-aware QA pairs. We employ a unified question template across both datasets. For MIMIC-CXR, we use CheXbert (Smit et al., 2020) to extract finding labels from reports. Following the four-state formulation in PromptMRG (Jin et al., 2024), we retain Positive, Negative, Uncertain, and No Mention as separate label categories. We construct finding-specific questions and convert the extracted labels into answers using the templates in Table C-2. The resulting QA pairs are used for image-conditioned preference training, with images serving as model inputs and report labels providing the answer targets. No Mention denotes non-mention in the source report and is distinguished from Negative at

Table D-1: CE metrics for the 26-class evaluation on additional CXR findings on LLaVA-Med v1.5 backbone. The 26 classes exclude the 14 CheXbert categories used in training.

Method	Precision	Recall	F1
LLaVA-Med	0.0197	0.0198	0.0188
+ DPO	0.0206	0.0199	0.0189
+ OTPO	0.0208	0.0185	0.0184
+ TI-DPO	0.0228	0.0212	0.0210
+ RRG-DPO	0.0227	0.0228	0.0217
+ MMedPO	**0.0250**	0.0237	0.0228
+ Ours	0.0249	**0.0244**	**0.0236**

Table D-2: CE metrics on the evaluation of additional 15 CT findings on the Dia-LLaMA backbone.

Method	Precision	Recall	F1
Dia-LLaMA	**0.3590**	0.2463	0.2806
+ DPO	0.3584	0.2554	0.2882
+ OTPO	0.3494	0.2525	0.2847
+ RRG-DPO	0.3449	0.2534	0.2826
+ TI-DPO	0.3622	0.2600	0.2943
+ Ours	0.3495	**0.2651**	**0.2949**

the state level. For CT-RATE (Hamamci et al.), we use the positive and negative answer templates corresponding to its two finding states provided by CT-RATE across 18 categories per report.

APPENDIX D MORE RESULTS

In this section, we show the results of generalization beyond the QA-supervised finding categories, comparison with auxiliary finding classification, ablation of alignment targets on R2Gen, the sensitivity of hyperparameters β_1 and β_2 in Eq. (7), the sensitivity of retrieval hyperparameters K, and visualization comparison results of the generated reports.

Generalization beyond the QA-supervised finding categories. To further investigate the generalization of our method beyond a predefined finding set, we evaluate additional finding categories excluded from the QA label sets used for alignment. For CXR, we remove the 14 CheXbert categories from the 40 training categories in CXR-LT 2024 Task 1 (Lin et al., 2025), leaving 26 additional classes for evaluation. We train an independent 26-class RadBERT text classifier to evaluate reports generated by LLaVA-Med v1.5 and the compared preference-optimized methods on the MIMIC-CXR test set. As shown in Tab. D-1, VF-DPO achieves the highest recall 0.0244 and F1 0.0236, increasing F1 from 0.0188 for the SFT baseline by 25.5% in relative terms. The low absolute scores indicate that these findings remain challenging, potentially owing to their rarity and the absence of explicit QA supervision. Our method still shows better recall and F1 scores than the compared methods on these categories, which shows the effectiveness of our method. For CT, we evaluate Dia-LLaMA on 15 additional finding categories (Lai et al., 2025) beyond the 18 used for QA alignment. As shown in Tab. D-2, VF-DPO achieves the highest recall of 0.2651 and F1 of 0.2949, with a modest decrease in precision relative to SFT. The results from both modalities provide preliminary evidence that the benefits of finding-anchored preference alignment extend beyond the explicitly supervised QA categories.

Comparison with auxiliary finding classification. We compare our method with multi-task learning which uses finding classification as a post-training auxiliary loss. We trained a variant that adds a discriminative finding-classification head on the mean-pooled LLM hidden states, comprising both image and report-prefix tokens, trained with binary cross-entropy loss alongside the standard DPO report loss on LLaVA-Med v1.5, as shown in Tab. D-3. Adding the finding classifier to DPO also improves CE metrics and RRG-Exclusive metrics, raising F1 from 0.3650 to 0.3683, RadGraph from

Table D-3: The results of comparison with auxiliary finding classification on 2D MIMIC-CXR (Johnson et al., 2019). The best results are bolded.

Methods	NLG metrics						CE metrics			RRG-Exclusive metrics		
	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
LLaVA-Med v1.5	0.2374	0.1328	0.0825	0.0554	0.1336	0.2083	0.3982	0.3596	0.3494	0.1603	1.2791	0.5306
+ DPO	0.2717	0.1519	0.0934	0.0618	0.1458	0.2148	0.3979	0.3895	0.3650	0.1664	1.2558	0.5323
+ DPO + finding-classifier	0.2748	0.1534	0.0940	0.0617	0.1464	0.2156	0.4009	0.3925	0.3683	0.1688	1.2488	0.5361
+ Ours	**0.2878**	**0.1606**	**0.0982**	**0.0643**	**0.1521**	**0.2159**	**0.4114**	**0.4378**	**0.3940**	**0.1691**	**1.2457**	**0.5365**

Table D-4: Ablation study on preference alignment tasks using R2Gen (Chen et al., 2020) on MIMIC-CXR (Johnson et al., 2019). The best and second best are respectively bolded and underlined.

Preference Alignment Targets			NLG metrics						CE metrics			RRG-Exclusive metrics		
$\pi(y\|x)$	$\pi(b\|x)$	$\pi(y\|b,x)$	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
✓			0.3847	0.2371	0.1593	0.1134	0.1566	0.2815	0.4218	0.3494	0.3572	0.1832	1.1896	0.5345
✓	✓		0.4055	0.2554	0.1749	0.1254	0.1606	0.2914	0.4378	0.3766	0.3791	0.2024	1.1100	0.5605
	✓	✓	**0.4075**	**0.2582**	**0.1774**	**0.1276**	0.1566	0.2930	0.4318	0.3654	0.3714	0.2034	1.1067	0.5669
✓	✓	✓	0.4046	0.2562	0.1757	0.1258	**0.1608**	**0.2956**	**0.4489**	**0.3999**	**0.3968**	**0.2097**	**1.0832**	**0.5675**

Table D-5: Results for sensitivity of β_1 on MIMIC-CXR (Johnson et al., 2019) using the LLaVA-Med v1.5 model (Li et al., 2023).

β_1	NLG metrics						CE metrics			RRG-Exclusive metrics		
	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
0.005	0.2841	0.1591	0.0973	0.0636	0.1512	0.2163	0.4104	0.4300	0.3899	0.1707	1.2398	0.5381
0.010	0.2878	0.1606	0.0982	0.0643	0.1521	0.2159	0.4114	0.4378	0.3940	0.1691	1.2457	0.5365
0.015	0.2869	0.1601	0.0978	0.0639	0.1515	0.2169	0.4104	0.4315	0.3903	0.1675	1.2473	0.5353
0.020	0.2875	0.1607	0.0983	0.0647	0.1518	0.2160	0.4146	0.4358	0.3943	0.1676	1.2540	0.5354
0.025	0.2879	0.1608	0.0981	0.0644	0.1517	0.2154	0.4031	0.4292	0.3867	0.1674	1.2539	0.5335

0.1664 to 0.1688, and RaTEScore from 0.5323 to 0.5361. VF-DPO still achieves the best results across all metrics, which shows the effectiveness of our method.

Ablation of alignment targets on R2Gen. We conducted the component ablation on R2Gen to verify whether the contribution of each alignment target generalizes beyond LLaVA-Med v1.5. As shown in Tab. D-4, DPO which uses the standard report alignment $\pi(y|x)$ achieves F1 0.3572. Adding the finding-state QA term $\pi(b|x)$ to standard DPO improves F1 to 0.3791. The coupled anchor-conditioned terms $\pi(b|x) + \pi(y|b, x)$ shows better results than strategy trained independently by $\pi(y|x)$ and $\pi(b|x)$ on most evaluation metrics, which further improves BLEU and the RRG-exclusive scores. Our full method achieves the best CE and RRG-exclusive results. Some BLEU scores of the full method are lower than those of the variant without the standard DPO term.

Results for the sensitivity of β_1. As shown in Tab. D-5, we investigate the impact of β_1 on model performance. Standard NLG metrics remain relatively stable as β_1 varies from 0.005 to 0.025. Our method achieves its most competitive clinical efficacy scores for β_1 around 0.010 and 0.020. We adopt $\beta_1 = 0.010$ for our experiments. It shows the optimal balanced results, achieving the highest recall and a strong F1 score while preserving the textual fluency of the generated reports and maintaining competitive performance in RRG-exclusive metrics.

Results for the sensitivity of β_2. As shown in Tab. D-6, we evaluate the impact of the hyperparameter β_2. Setting $\beta_2 = 0$ reverts the framework to our method without the DPO reward margin ΔR_{DPO}. This setting already achieves a competitive clinical Recall and a solid F1 score. When introducing a small DPO reward ($\beta_2 = 0.05$), we observe some NLG metrics slightly drop. The performance reaches a balance at $\beta_2 = 0.10$, yielding the competitive NLG scores alongside strong RRG-exclusive metrics and the best F1 score of 0.3940. As β_2 increases to 0.15 and 0.20, most of the NLG metrics and F1 score begin to decrease. This indicates that a large weight β_2 may affect the primary generation task. Therefore, we use $\beta_2 = 0.10$ for all experiments.

Sensitivity to the Retrieval Size K. We investigate the effect of the hard-negative candidate pool size by varying the number of retrieved reports K, using R2Gen (Chen et al., 2020) on MIMIC-CXR (Johnson et al., 2019). As shown in Tab. D-7, increasing K from 10 to 20 consistently improves ROUGE-L, recall, F1, RadGraph, and RadCliQ. In particular, F1 increases from 0.3727 to 0.3968, while RadCliQ decreases from 1.1459 to 1.0832. These results suggest that moderately expanding

Table D-6: Results for sensitivity of β_2 on MIMIC-CXR (Johnson et al., 2019) using the LLaVA-Med v1.5 model (Li et al., 2023).

β_2	NLG metrics						CE metrics			RRG-Exclusive metrics		
	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
0.00	0.2786	0.1553	0.0954	0.0634	0.1491	0.2146	0.3947	0.4294	0.3830	0.1673	1.2512	0.5309
0.05	0.2805	0.1551	0.0935	0.0603	0.1552	0.2142	0.3964	0.4363	0.3855	0.1665	1.2665	0.5310
0.10	0.2878	0.1606	0.0982	0.0643	0.1521	0.2159	0.4114	0.4378	0.3940	0.1691	1.2457	0.5365
0.15	0.2811	0.1578	0.0971	0.0642	0.1495	0.2162	0.4054	0.4167	0.3821	0.1686	1.2431	0.5374
0.20	0.2781	0.1554	0.0952	0.0625	0.1483	0.2156	0.4032	0.4086	0.3769	0.1682	1.2469	0.5351

Table D-7: Effect of pool size K on MIMIC-CXR (Johnson et al., 2019) using R2Gen (Chen et al., 2020). The default K=20 follows RRG-DPO.

K	NLG metrics						CE metrics			RRG-Exclusive metrics		
	Bleu1 ↑	Bleu2 ↑	Bleu3 ↑	Bleu4 ↑	Meteor ↑	Rouge L ↑	Precision ↑	Recall ↑	F1 ↑	RadGraph ↑	RadCliQ ↓	RaTEScore ↑
10	0.4021	0.2502	0.1691	0.1210	0.1584	0.2892	0.4392	0.3637	0.3727	0.1906	1.1459	0.5454
15	0.3966	0.2491	0.1695	0.1208	0.1623	0.2925	0.4567	0.3848	0.3900	0.1997	1.1242	0.5596
20	0.4046	0.2562	0.1757	0.1258	0.1608	0.2956	0.4489	0.3999	0.3968	0.2097	1.0832	0.5675
25	0.3932	0.2493	0.1714	0.1227	0.1568	0.2903	0.4325	0.3588	0.3652	0.2103	1.1097	0.5673
30	0.3941	0.2500	0.1728	0.1243	0.1498	0.2863	0.4174	0.3305	0.3410	0.2000	1.1523	0.5680

Table D-8: Training time and GPU memory consumption based on R2Gen.

Methods	Training Time (h)	GPU Memory (GB)
+ DPO	3.01	4.43
+ Ours	7.76	12.20

the candidate pool helps identify hard negatives that are linguistically similar to the reference report but differ in clinical findings, providing more informative supervision for preference optimization. However, further increasing K to 25 or 30 degrades performance on most clinical metrics. At $K = 30$, F1 and recall drop to 0.3410 and 0.3305, respectively, while RadCliQ rises to 1.1523. A plausible explanation is that larger pools include more reports with low textual similarity to the reference. The selected negatives may therefore become easier to distinguish based on wording alone, weakening the training signal for learning clinically meaningful distinctions. We consequently set $K = 20$, which achieves the best overall performance among the evaluated settings and offers a favorable balance between negative hardness and textual similarity.

Computational Cost. We compare the training cost of VF-DPO and standard DPO using R2Gen on a single NVIDIA RTX 4090. As shown in Tab. D-8, VF-DPO requires approximately 2.6× the training time and 2.8× the GPU memory of standard DPO. The additional cost arises from evaluating log-likelihood ratios for finding-state alignment and conditional report alignment during training. At inference, VF-DPO generates reports directly from images without explicitly generating the latent finding states. It has the same inference procedure and computational complexity with that of the SFT and DPO baselines.

Qualitative Comparison of Generated Reports. We present qualitative visualizations of the generated reports for two modalities. For X-ray images, we visualize the generated results based on R2Gen (Chen et al., 2020) (Fig. D-1) and LLaVA-Med v1.5 (Li et al., 2023) (Fig. D-2, Fig. D-3). For CT scans, we provide generated examples from Dia-LLaMA (Chen et al., 2025b) (Fig. D-4) and Reg2RG (Chen et al., 2025a) (Fig. D-5, Fig. D-6). Each figure displays the input image, the ground-truth report, and the generated reports by the compared methods. We use a color-coded scheme for finding-level comparisons. Different colors represent different findings mentioned in the ground-truth report, such as cardiomegaly, stable mediastinal and hilar contours, and no pleural effusion/pneumothorax, shown in Fig. D-1. Shaded texts indicate inconsistent descriptions of related findings. Black text denotes content not mentioned in the original report.

For the X-ray reports, as shown in Fig. D-1, our method generates the reference finding "the heart is mildly enlarged" as highlighted by the green sentence and preserves the statement that "the mediastinal and hilar contours appear unchanged." While some baselines capture related information about cardiac size or contour stability, their descriptions differ from the reference in either the reported

Input Image/ Ground Truth		a picc line has been removed. the heart is mildly enlarged. the mediastinal and hilar contours appear unchanged. there is no pleural effusion or pneumothorax. the lungs appear clear.
Baseline	the lungs are clear without focal consolidation. no pleural effusion or pneumothorax is seen. the cardiac and mediastinal silhouettes are stable.	
DPO	ap portable upright view of the chest. overlying ekg leads are present. there is no focal consolidation effusion or pneumothorax. the cardiomediastinal silhouette is normal. imaged osseous structures are intact. no free air below the right hemidiaphragm is seen.	
OTPO	ap portable upright view of the chest. there is no focal consolidation effusion or pneumothorax. the cardiomediastinal silhouette is normal. imaged osseous structures are intact. no free air below the right hemidiaphragm is seen.	
TI-DPO	the lungs are clear without focal consolidation. no pleural effusion or pneumothorax is seen. the cardiac silhouette is top-normal to mildly enlarged.	
MMedPO	pa and lateral views of the chest provided. there is no focal consolidation effusion or pneumothorax. the cardiomediastinal silhouette is normal. imaged osseous structures are intact. no free air below the right hemidiaphragm is seen.	
RRG-DPO	the lungs are clear without focal consolidation. no pleural effusion or pneumothorax is seen. the cardiac and mediastinal silhouettes are stable.	
Ours	the heart is mildly enlarged. the mediastinal and hilar contours appear unchanged. there is no pleural effusion or pneumothorax. the lungs appear clear.	

Support devices Cardiomegaly Stable mediastinum/hila No effusion/pneumothorax Clear lungs

Figure D-1: Visualization comparison of X-ray report generation on the MIMIC-CXR dataset (Johnson et al., 2019) based on R2Gen (Chen et al., 2020). Different colors denote different findings in the ground-truth report, and shaded text indicates inconsistent descriptions of related findings. Black text represents the content not mentioned in the ground truth report.

status or the set of covered findings. As shown in Fig. D-2, our method captures the finding "Clips in the upper abdomen noted," whereas all compared baselines omit this observation. Fig. D-3 compares our method with standard DPO. Our method retains the reference finding "Tiny clips seen in the left upper quadrant," while DPO additionally generates "Mild degenerative changes are noted in the lower thoracic spine," a finding not supported by the reference report. These visualization comparison results show the advantages of our method over other baselines.

For the CT scans, our method produces reports that are more consistent with the reference findings, as shown in Figs. D-4, D-5, and D-6. As shown in Fig. D-4, several baselines mention nodules that are not supported by the reference report, whereas VF-DPO does not include these statements. In addition, our method provides a relatively accurate description of the sequelae changes, as highlighted by the red sentence. Additionally, our method mentions "There are degenerative changes in the bone structure", as highlighted by the brown sentence. In Fig. D-5, our method mentions "Calcific atheroma plaques" and "Mosaic attenuation pattern", which are absent from the reports generated by other methods. As shown in Fig. D-6, our method generates more accurate descriptions of emphysematous changes and sequelae compared with the standard DPO method. Overall, these qualitative comparisons demonstrate that our method achieves more accurate alignment with the ground-truth reports, reducing factual errors while preserving diagnostic details.

APPENDIX E BROADER IMPACTS

Our work presents positive societal impacts by advancing automated radiology report generation. By aligning models to prioritize clinical relevance and the factual correctness of finding state predictions, our method has the potential to reduce the heavy workload of radiologists, streamline clinical workflows, and improve the overall efficiency of medical diagnostics.

However, we also acknowledge potential negative impacts arising from the inherent limitations of generative models. The generated radiology reports are not perfectly accurate and may still contain clinical hallucinations, omissions, or misdiagnoses. If these generated reports are relied upon without proper human verification, incorrect results could lead to inappropriate medical decisions

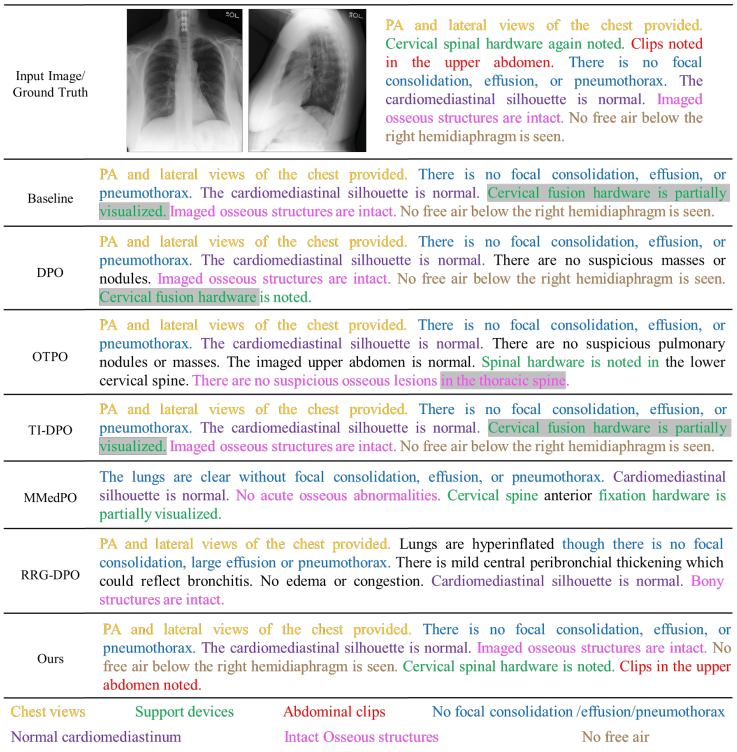


Figure D-2: Visualization comparison of X-ray report generation on the MIMIC-CXR dataset (Johnson et al., 2019) based on LLaVA-Med v1.5 (Li et al., 2023). Different colors denote different findings in the ground-truth report, and shaded text indicates inconsistent descriptions of related findings. Black text represents content not mentioned in the ground truth report.

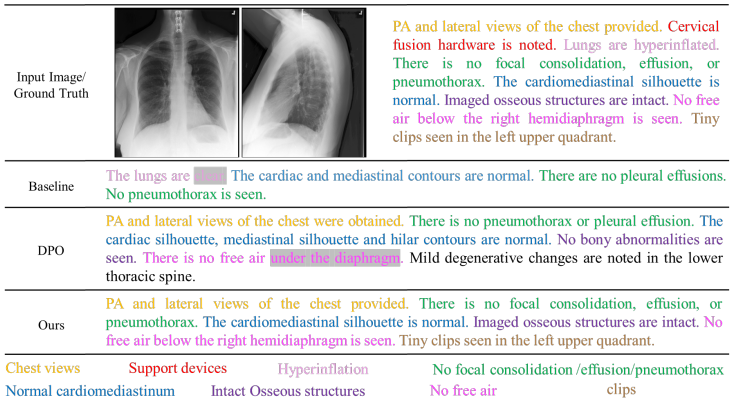
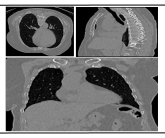


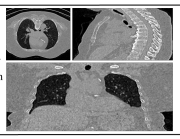
Figure D-3: Visualization comparison of X-ray report generation on the MIMIC-CXR dataset (Johnson et al., 2019) based on LLaVA-Med v1.5 (Li et al., 2023). Different colors denote different findings in the ground-truth report, and shaded text indicates inconsistent descriptions of related findings. Black text represents content not mentioned in the ground truth report.

Input Image/ Ground Truth	 The aortic arch calibration is 30 mm, slightly larger than normal. Thoracic esophagus calibration was normal and no significant tumoral wall thickening was detected. Mild hiatal hernia is observed. When examined in the lung parenchyma window; There are increases in pleuroparenchymal density evaluated in favor of sequelae in the middle lobe of the right lung. In the mediobasal level of the lower lobe of the right lung, changes that are evaluated primarily in favor of sequelae are observed. Sequelae changes are observed in the middle lobe on the right. There was no finding suggestive of active infiltration in both lungs. No pleural effusion or pneumothorax was observed. Minimal degenerative changes are observed in the bone structure.
Baseline	Trachea and both main bronchi are open. No pathological wall thickness increase was observed in the esophagus within the sections. Heart contour and size are normal. No pleural or pericardial effusion was detected. When examined in the lung parenchyma window; Ground glass areas are observed in the peripheral subpleural area in the upper lobe of the left lung, the middle lobe of the right lung and the lower lobe of both lungs. The described manifestations are the findings frequently observed in Covid-19 pneumonia. No mass was detected in both lungs. There are several millimetric nodules in both lungs. No lytic-destructive lesions were detected in the bone structures within the sections.
DPO	Trachea and both main bronchi are open. There is a sliding type hiatal hernia at the lower end of the esophagus. When examined in the lung parenchyma window; Peripheral and centrally located ground glass areas are observed in the upper and lower lobes of both lungs and the middle lobe of the right lung. The described manifestations were primarily evaluated in favor of Covid-19 pneumonia during the pandemic process. There are millimetric nodules in both lungs. No mass was detected in both lungs. No lytic-destructive lesions were detected in the bone structures within the sections.
OTPO	Trachea and both main bronchi are open. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. Sliding type hiatal hernia was observed at the lower end of the esophagus. When examined in the lung parenchyma window; Patchy ground glass densities were observed in the right lung lower lobe superior segment and right lung upper lobe anterior segment, and in the left lung upper lobe lingular segment. The appearance was evaluated in favor of Covid-19 pneumonia. It is recommended to be evaluated together with clinical and laboratory. A few millimetric nonspecific parenchymal nodules were observed in both lungs. No mass lesion with distinguishable borders was detected in both lungs. Bone structures in the study area are natural. Vertebral corpus heights are preserved.
TI-DPO	Trachea, both main bronchi are open. Mediastinal main vascular structures, heart contour, size are normal. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. When examined in the lung parenchyma window; Emphysematous changes are observed in both lungs. There are nodules measuring up to 7 mm in size in both lungs. In both lungs, there are ground glass densities in the peripheral subpleural area. Bone structures in the study area are natural. Vertebral corpus heights are preserved.
RRG-DPO	Trachea and both main bronchi are open. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. Sliding type hiatal hernia was observed at the lower end of the esophagus. When examined in the lung parenchyma window; Peripheral ground glass densities were observed in both lungs. The outlook is highly suspicious for Covid-19 pneumonia. It is recommended to be evaluated together with clinical and laboratory. A few millimetric nonspecific parenchymal nodules were observed in both lungs. No mass lesion with distinguishable borders was detected in both lungs. Bone structures in the study area are natural. Vertebral corpus heights are preserved.
Ours	Trachea and both main bronchi were in the midline and no obstructive pathology was detected in the lumen. Thoracic esophagus calibration was normal and no significant tumoral wall thickening was detected. Sliding type hiatal hernia was observed at the lower end of the esophagus. When examined in the lung parenchyma window; Pleuroparenchymal density increases were observed in the right lung lower lobe medial basal segment, left lung upper lobe lingular segment and both lung lower lobe posterobasal segments. The outlook was primarily evaluated in favor of sequelae. No mass lesion-active infiltration with distinguishable borders was detected in the lung parenchyma. There are degenerative changes in the bone structure.

Enlarged aortic arch | Normal esophageal calibration/wall | Hiatal hernia | Sequelae change

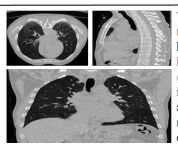
No infiltration | No pleural effusion/pneumothorax | Degenerative changes

Figure D-4: Visualization comparison of CT report generation on the CT-RATE dataset (Hamamci et al.) based on Dia-LLaMA (Chen et al., 2025b). Different colors denote different findings in the ground-truth report, and shaded text indicates inconsistent descriptions of related findings. Black text represents content not mentioned in the ground truth report.

Input Image/ Ground Truth		No occlusive pathology was detected in the trachea and lumen of both main bronchi. Calcific atheroma plaques were observed in the abdominal aorta. Sliding type hiatal hernia was observed at the lower end of the esophagus. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. Heart contour, size is normal. A mosaic attenuation pattern was observed in both lungs (small airway disease?, small vessel disease?). When examined in the lung parenchyma window; Pleuroparenchymal fibroatelectasis sequelae changes were observed in the right lung middle lobe medial and left lung upper lobe inferior lingular segment.
Baseline		No occlusive pathology was observed in the trachea and lumen of both main bronchi. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. Heart size increased. No mass lesion-active infiltration with distinguishable borders was detected in both lungs. When examined in the lung parenchyma window; Minimal emphysematous changes were observed in both lungs. No enlarged lymph nodes in prevascular, pre-paratracheal, subcarinal or bilateral hilar-axillary pathological dimensions were detected.
DPO		Trachea, both main bronchi are open. Thoracic esophagus calibration was normal and no significant tumoral wall thickening was detected. Mediastinal main vascular structures, heart contour, size are normal. When examined in the lung parenchyma window; Minimal pleural effusion is observed in both lungs. No enlarged lymph nodes in prevascular, pre-paratracheal, subcarinal or bilateral hilar-axillary pathological dimensions were detected. Millimetric calcific nodules were observed in the middle lobe of the right lung.
OTPO		Trachea, both main bronchi are open. Thoracic esophagus calibration was normal and no significant tumoral wall thickening was detected. Pericardial effusion-thickening was not observed. Mediastinal main vascular structures, heart contour, size are normal. When examined in the lung parenchyma window; Aeration of both lung parenchyma is normal and no nodular or infiltrative lesion is detected in the lung parenchyma.
TI-DPO		As far as can be seen; Trachea and lumen of both main bronchi are open. No dilatation was detected in the thoracic aorta. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. Heart contour size is natural. No lymph node was detected in mediastinal and bilateral hilar pathological size and appearance. The outlook is consistent with the frequently reported imaging features of Covid-19 pneumonia. When examined in the lung parenchyma window; Diffuse ground glass density increases were observed in both lungs. Other viral pneumonias can be considered in the differential diagnosis.
RRG-DPO		Trachea, both main bronchi are open. Thoracic esophagus calibration was normal and no significant tumoral wall thickening was detected. Mediastinal main vascular structures, heart contour, size are normal. When examined in the lung parenchyma window; Minimal pleural effusion is observed in both lungs. Millimetric calcific nodules were observed in the middle lobe of the right lung. It is recommended to be evaluated together with US. The thyroid gland has a heterogeneous appearance.
Ours		No occlusive pathology was detected in the trachea and both main bronchi. Calcific atheroma plaques were observed in the aorta and coronary arteries. Sliding type hiatal hernia was observed at the lower end of the esophagus. No pathological increase in wall thickness was observed in the esophagus. A mosaic attenuation pattern was observed in both lungs (small airway disease? small vessel disease?). There are several nonspecific nodules in both lungs, the largest of which is 4 mm in diameter. No mass or infiltrative lesion was detected in both lungs.

Normal trachea/bronchi | Calcific atheroma plaques | Hiatal hernia | Normal esophageal calibration/wall
Normal heart | Mosaic attenuation pattern | Sequelae changes

Figure D-5: Visualization comparison of CT report generation on the CT-RATE dataset (Hamamci et al.) based on Reg2RG (Chen et al., 2025a). Different colors denote different findings in the ground-truth report, and shaded text indicates the inconsistent descriptions of related findings. Black text represents the content not mentioned in the ground truth report.

Input Image/ Ground Truth		Trachea and main bronchi are open. No pathology was detected in the sections passing through the upper part of the abdomen. Pleural effusion-thickening was not detected in both hemithorax. No pathological increase in wall thickness was observed in the esophagus. The heart and mediastinal vascular structures could not be evaluated optimally due to the lack of contrast, and they have a natural appearance. Active infiltration or mass lesion is not detected in both lungs, and there are millimetric nonspecific nodules. Emphysematous changes, which are more evident in the upper lobes of both lungs, are sequelae in the right lung middle lobe medial segment, left lower lobe inferior lingular segment. No lytic or destructive lesions were detected in bone structures.
Baseline		Trachea and lumen of both main bronchi are open. No occlusive pathology was detected in the trachea and lumen of both main bronchi. Upper abdominal sections entering the examination area are natural. Bilateral gynecomastia was observed. Thoracic esophagus calibration was normal and no significant pathological wall thickening was detected. Heart contour size is natural. Pericardial thickening-effusion was not detected. No lymph node was detected in mediastinal and bilateral hilar pathological size and appearance. When examined in the lung parenchyma window; no mass nodule-infiltration was detected in both lung parenchyma. No dilatation was detected in the thoracic aorta. As far as can be observed: Calcified lymph nodes were observed in the left thyroid lobe. US control is recommended. No lytic-destructive lesion was detected in bone structures.
DPO		Trachea, both main bronchi are open. Upper abdominal organs included in the sections are normal. There is a bilateral gynecomastia appearance. Thoracic esophageal calibration was normal and no significant tumoral wall thickening was detected. Pericardial effusion-thickening was not observed. Mediastinal main vascular structures, heart contour, size are normal. When examined in the lung parenchyma window; There are mild atelectasis in the lower lobes of both lungs. Thoracic aorta diameter is normal. No enlarged lymph nodes in prevascular, pre-paratracheal, subcarinal or bilateral hilar-axillary pathological dimensions were detected. The thyroid gland is larger than normal and has a heterogeneous appearance. Bone structures in the study area are natural. Vertebral corpus heights are preserved.
Ours		Trachea, both main bronchi are open. Upper abdominal organs included in the sections are normal. Thoracic esophageal calibration was normal and no significant tumoral wall thickening was detected. Pericardial effusion-thickening was not observed. Mediastinal main vascular structures, heart contour, size are normal. When examined in the lung parenchyma window; There are minimal sequelae changes at the apical level. There are minimal emphysematous changes in both lungs. There are sequelae changes in the lingular segment of the left lung. No enlarged lymph nodes in prevascular, pre-paratracheal, subcarinal or bilateral hilar-axillary pathological dimensions were detected. The thyroid gland has a heterogeneous appearance and nodular formations are observed. Bone structures in the study area are natural. Vertebral corpus heights are preserved.

Normal trachea/bronchi | Normal abdomen | No Pleural effusion-thickening | Normal esophageal wall
Normal heart and mediastinal vascular structures | Lung nodules | Emphysematous changes | Sequelae | Normal bone structures

Figure D-6: Visualization comparison of CT report generation on the CT-RATE dataset (Hamamci et al.) based on Reg2RG (Chen et al., 2025a). Different colors denote different findings in the ground-truth report, and shaded text indicates the inconsistent descriptions of related findings. Black text represents content not mentioned in the ground truth report.

or delayed treatments. Therefore, we emphasize that our method is designed strictly as an assistive tool for qualified healthcare professionals and must not be deployed for fully autonomous clinical decision-making.