CEF-OPD: LEARNING TO CORRECT CONSEQUENTIAL ERRORS IN RADIOLOGY REPORT GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Radiology report generators are typically trained on reference prefixes but deployed on their own predictions. An early clinical contradiction may influence subsequent claims, yet its local prediction error does not directly quantify the downstream clinical consequences. We introduce a clinical error frontier on-policy distillation (CEF-OPD) framework that allocates corrective supervision according to the estimated downstream consequences of model-generated errors. CEF-OPD identifies the earliest contradiction with extracted reference facts and constructs a reference-supported minimal correction. Paired continuations from the erroneous and corrected prefixes estimate the change in downstream clinical cost, excluding the directly edited claim. This estimate determines the weight assigned to local correction and distillation, emphasizing errors with greater estimated downstream impact. Reference clinical facts provide privileged guidance for self-distillation from the pre-error state, targeting the decision that introduces the contradiction. We characterize the paired cost difference as a model-internal cost-to-go advantage and bound local clinical risk differences through propagation-weighted distillation under stated assumptions. Experiments on MIMIC-CXR and IU X-ray datasets demonstrate the effectiveness of our proposed method.

1 INTRODUCTION

Radiology report generation aims to produce clinically accurate descriptions from medical images. Existing methods have improved generation through memory-based decoding (Chen et al., 2020), visual-language alignment (Wang et al., 2023), diagnostic prompts (Jin et al., 2024), and additional clinical knowledge (Hou et al., 2025). Despite this progress, clinically significant errors remain in generated reports (Tanno et al., 2025). Since linguistic similarity does not fully reflect clinical accuracy, previous work has introduced objectives based on clinical consistency and semantic rewards (Miura et al., 2021; Delbrouck et al., 2022), alongside clinical evaluation methods such as GREEN (Ostmeier et al., 2024). However, evaluating a completed report does not reveal how much an earlier error affects later content, or how much improvement could be achieved by correcting that error.

Supervised fine-tuning trains a model on reference prefixes, whereas inference relies on its own predictions. This mismatch motivates learning from model-generated states (Ross et al., 2011; Bengio et al., 2015). However, errors do not necessarily accumulate throughout generation, as models can also recover from incorrect prefixes (He et al., 2021). On-policy distillation addresses this mismatch by providing teacher feedback on student-generated trajectories (Agarwal et al., 2024), while on-policy self-distillation uses reference information to guide the same model's learning (Zhao et al., 2026). Recent studies further examine unreliable teacher feedback (Fu et al., 2026) and differences in the usefulness of teacher–student disagreement (Wang et al., 2026). For clinical reporting, these findings motivate identifying which errors affect subsequent claims and allocating supervision according to their estimated impact.

Our approach considers both where corrective supervision should begin and how much supervision each error should receive. First, generated and reference reports may describe the same clinical facts using different words and ordering, so token differences alone cannot reliably identify clinical errors. Correction should instead target a claim that explicitly contradicts the reference facts. Once an incorrect claim enters the prefix, next-token supervision cannot remove it, and subsequent predictions may

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.

be influenced by the incorrect context. This motivates applying corrective supervision immediately before the claim is generated. Second, local prediction errors or teacher-student disagreement do not directly measure the downstream benefit of a correction. Some errors may affect several later claims, whereas others may have little influence on subsequent content. Rollout-based cost-sensitive learning evaluates decisions through their subsequent outcomes (Chang et al., 2015). Applying this idea to clinical reporting requires comparing an incorrect claim with a reference-supported correction while separating their effects on later content from the direct benefit of editing.

We propose a novel method named as Clinical Error Frontier On-Policy Distillation (CEF-OPD). Using clinical entities and relation extraction (Jain et al., 2021; Delbrouck et al., 2024), CEF-OPD locates the earliest detected explicit contradiction with reference facts and constructs a minimal correction that preserves other clinical content. The erroneous and corrected claims are appended to the same student-generated prefix, and the model produces paired continuations from the two branches. Their difference in downstream clinical cost estimates the benefit of the correction, excluding the edited claim and retaining penalties for omitted reference facts. This comparison measures changes in subsequent clinical content rather than simply rewarding the removal of the initial error. The positive part of this estimate weights local correction and self-distillation, assigning stronger supervision to errors with greater estimated downstream impact. Reference facts provide privileged guidance from the pre-error state, directing learning toward the decision that introduces the contradiction. This objective complements standard report-generation training, and all auxiliary components are used only during training. Our contributions are threefold:

- We define a clinical-semantic intervention point and construct reference-supported minimal corrections at states visited by the student.
- We use paired continuations to estimate the downstream value of correcting a claim, isolating changes in downstream clinical cost from the directly edited error.
- We use this estimate to guide local correction and privileged self-distillation, supported by theoretical analysis and experiments on two report-generation benchmarks.

2 Related Work

Radiology report generation. Research on radiology report generation has focused on improving the connection between image evidence and report language. Memory-based methods capture report patterns and visual-textual correspondence (Chen et al., 2020; 2021). More recent approaches combine visual representations with pretrained language models or introduce diagnostic prompts and additional knowledge to guide generation (Wang et al., 2023; Jin et al., 2024; Hou et al., 2025). These approaches improve the information available to the generator and how that information is used. However, stronger visual-language alignment alone does not determine how to correct a clinical error once the model encounters it in its own generation. Our CEF-OPD complements these modeling approaches by studying where corrective supervision should begin and how its strength should depend on the estimated consequences of an error.

Clinical factuality in report generation. Clinical factuality requires more than matching the wording of a reference report. Previous work incorporates entity agreement, entailment consistency, and semantic rewards into training to improve the correctness and completeness of generated findings (Miura et al., 2021; Delbrouck et al., 2022). Clinical representations such as RadGraph support comparisons of entities and relations (Jain et al., 2021), while GREEN identifies clinically significant discrepancies between generated and reference reports (Ostmeier et al., 2024). These methods provide useful signals about which clinical statements are correct or incorrect. However, an error score alone does not establish how correcting one statement would affect subsequent content. Our CEF-OPD uses extracted reference facts to define a local correction and compares continuations from the erroneous and corrected prefixes.

On-policy knowledge distillation. On-policy distillation provides teacher feedback on student-generated sequences, reducing the mismatch between the states used for training and those encountered during inference (Agarwal et al., 2024). On-policy self-distillation further allows one model to serve as both teacher and student by giving the teacher additional reference information (Zhao et al., 2026). Recent work also studies how to select useful supervision by distinguishing learnable

teacher-student disagreement from less useful distributional differences (Wang et al., 2026). These approaches show that both the training states and the choice of teacher feedback matter. Our CEF-OPD builds on this foundation but selects corrective supervision according to clinical meaning and estimated downstream benefit.

3 METHODOLOGY

For sample i, let $\mathcal{I}_i$ denote the input radiographs and $R_i = (y_{i,1}, \ldots, y_{i,T_i})$ the reference report with T_i tokens. A visual encoder converts the images into visual tokens for the language model. The resulting generator, with trainable parameters θ, produces a report $\hat{R}_i$. We initialize the model through supervised fine-tuning (SFT) with $\mathcal{L}_{\mathrm{SFT}}^{(i)} = -\frac{1}{T_i}\sum_{t=1}^{T_i} \log p_\theta(y_{i,t} \mid \mathcal{I}_i, y_{i,<t})$. While SFT conditions on reference prefixes, CEF-OPD introduces corrective supervision at states reached through the model's own generation.

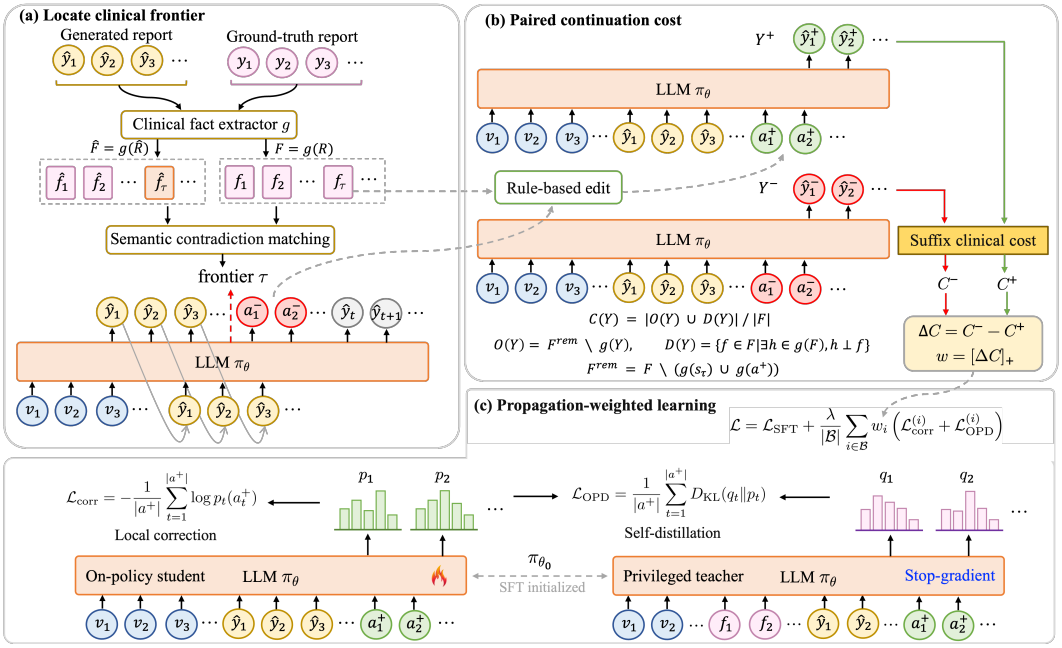


Figure 1: Overview of CEF-OPD. (a) Clinical fact matching identifies the first explicit contradiction and supports a minimal correction. (b) Paired continuations estimate the downstream cost difference, excluding the edited claim. (c) The resulting weight scales local correction and self-distillation, with reference facts guiding a teacher that shares the student's parameters.

3.1 CLINICAL ERROR FRONTIER AND MINIMAL CORRECTION

We periodically collect training episodes using a fixed generator snapshot to generate reports $\hat{R}_i \sim \pi_\theta(\cdot \mid \mathcal{I}_i)$. To compare the reports at the clinical level, a fixed extractor g that combines a frozen pretrained RadGraph-XL model (Delbrouck et al., 2024) with deterministic normalization rules then maps the reference and generated reports to the fact sets $F_i = g(R_i)$ and $\hat{F}_i = g(\hat{R}_i)$, respectively. Each fact is represented by $f = (o, a, l, r)$, containing an observation o, assertion a, explicit anatomical location l, and observation-anatomy relation r. The extractor combines clinical entity and relation extraction (Jain et al., 2021) with fixed normalization rules. Source spans are stored separately for locating claims and edits. All set operations below compare normalized semantic fields rather than source spans. We retain only facts covered by the predefined schema and exclude ambiguous extraction results and unresolved reference conflicts.

Let $\hat{f} \perp f$ denote a reference-supported contradiction between two comparable clinical facts. The primary case is a present-absent disagreement for the same finding within a compatible anatomical

scope. Let $b_{i,j}$ denote the starting token position of the smallest self-contained claim containing the generated fact $\hat{f}_{i,j}$. For reports with at least one detected contradiction, we define the clinical error frontier and its preceding prefix as

$$\tau_i = \min\{b_{i,j} \mid \exists f \in F_i,\ \hat{f}_{i,j} \perp f\}, \qquad s_{\tau_i} = \hat{y}_{i,<\tau_i}. \tag{1}$$

The frontier τ_i marks the start of the earliest detected contradictory claim. Denoting this claim by a_i^-, we decompose the generated report as $\hat{R}_i = s_{\tau_i} \oplus a_i^- \oplus \hat{Y}_i$, where $\oplus$ denotes concatenation and $\hat{Y}_i$ is the original suffix.

To correct a_i^-, let $f_i^- \in g(a_i^-)$ denote the selected erroneous fact and $f_i^+ \in F_i$ its uniquely matched reference counterpart, with $f_i^- \perp f_i^+$. We construct the corrected claim a_i^+ using a deterministic rule-based editor. The editor matches predefined phrase patterns to locate the conflicting expression and determines its replacement from f_i^+. For assertion contradictions, it replaces the matched polarity expression while preserving the finding and anatomical expressions. For anatomical contradictions, it replaces the conflicting location phrase with the reference-supported location. Editing requires a unique pattern match and an isolated target span. Claims with ambiguous negation scope or inseparable multiple errors are excluded. We then apply the extractor g to both claims, obtaining $G_i^- = g(a_i^-)$ and $G_i^+ = g(a_i^+)$, which contain all clinical facts extracted from the erroneous and corrected claims, respectively. Unlike the individual target facts f_i^- and f_i^+, these sets also include any non-target facts expressed in the claims. The correction is accepted only when

$$G_i^+ = (G_i^- \setminus \{f_i^-\}) \cup \{f_i^+\}, \qquad f_i^- \notin G_i^+. \tag{2}$$

These conditions require replacing the erroneous fact with its reference-supported counterpart while preserving all other extracted facts. Each accepted episode contains $(\mathcal{I}_i, F_i, s_{\tau_i}, a_i^-, a_i^+)$.

3.2 Paired Continuations and Clinical Cost

To assess the downstream effect of correcting the frontier claim, we generate paired continuations from the erroneous and corrected prefixes using the same generator snapshot

$$Y_i^\sigma \sim \pi_\theta(\cdot \mid \mathcal{I}_i, s_{\tau_i}, a_i^\sigma), \qquad \sigma \in \{-, +\}, \tag{3}$$

where σ identifies the erroneous or corrected input branch. Each Y_i^σ contains only newly generated suffix tokens, excluding the shared prefix and the appended claim. Both branches use the same generation limit and share the remaining reference fact set $F_i^{\text{rem}} = F_i \setminus \left(g(s_{\tau_i}) \cup G_i^+\right)$ for omission scoring. This set excludes reference facts already expressed by the shared prefix or the corrected claim, including the target fact and preserved non-target facts. For a suffix Y, we define the omitted reference fact set $O_i(Y)$ and the contradicted reference fact set $D_i(Y)$ as

$$O_i(Y) = F_i^{\text{rem}} \setminus g(Y), \qquad D_i(Y) = \{f \in F_i \mid \exists h \in g(Y),\ h \perp f\}. \tag{4}$$

Contradictions are checked against the full reference set F_i, allowing errors that reappear in the suffix to be counted. Since only newly generated suffix tokens are evaluated and the target fact is excluded from omission scoring, the direct correction itself does not contribute to the measured improvement. The clinical cost is the fraction of reference facts affected by either error

$$C_i(Y) = \frac{|O_i(Y) \cup D_i(Y)|}{|F_i|}. \tag{5}$$

An accepted episode has $|F_i| \geq 1$, and both error sets are subsets of F_i, so $C_i(Y) \in [0, 1]$. Each reference fact contributes at most once to the cost, even if it is associated with multiple detected errors. The forced claim is outside Y and therefore is not scored. If its erroneous fact appears again in the generated suffix, it is counted through $D_i(Y)$. Contradictions of other previously expressed facts are counted in the same way. The omission term penalizes empty or incomplete continuations when reference targets remain. Findings unmentioned in the reference are not automatically counted as errors. Thus the score measures reference-relative fidelity within the extractor's coverage. We compute the clinical cost of each continuation and use their difference to obtain a nonnegative propagation weight

$$C_i^\sigma = C_i(Y_i^\sigma), \quad \Delta C_i = C_i^- - C_i^+, \quad w_i = [\Delta C_i]_+. \tag{6}$$

Here $[x]_+ = \max(0, x)$, giving $w_i \in [0, 1]$. It gives greater weight to corrections with larger estimated reductions in downstream clinical cost. Corrections that do not reduce this cost receive zero weight.

3.3 PROPAGATION-WEIGHTED SELF-DISTILLATION

To prioritize corrections with larger downstream cost reductions, we perform propagation-weighted self-distillation. The student and privileged teacher use the same SFT-initialized report generator and share the current parameters θ. The student receives the images $\mathcal{I}_i$, the pre-error prefix s_{τ_i}, and the corrected-claim history $a^+_{i,<t}$. The teacher additionally receives the reference facts F_i. For token position t within the corrected claim a^+_i, their output distributions are

$$p_{i,t} = p_\theta(\cdot \mid \mathcal{I}_i, s_{\tau_i}, a^+_{i,<t}), \quad q_{i,t} = \mathrm{sg}\big[p_\theta(\cdot \mid \mathcal{I}_i, F_i, s_{\tau_i}, a^+_{i,<t})\big], \tag{7}$$

where sg denotes the stop-gradient operator, which prevents gradients from flowing through the teacher outputs. The teacher parameters remain shared with the student and evolve as the shared model is updated. The reference facts F_i are serialized as additional teacher input. The prefix s_{τ_i} comes from the student rollout, while the corrected-claim history $a^+_{i,<t}$ is teacher-forced in both branches. The local correction loss supervises the reference-supported correction, and the self-distillation loss aligns the student distribution with the privileged teacher distribution

$$\mathcal{L}^{(i)}_{\mathrm{corr}} = -\frac{1}{|a^+_i|}\sum_{t=1}^{|a^+_i|} \log p_{i,t}(a^+_{i,t}), \quad \mathcal{L}^{(i)}_{\mathrm{OPD}} = \frac{1}{|a^+_i|}\sum_{t=1}^{|a^+_i|} D_{\mathrm{KL}}(q_{i,t} \| p_{i,t}). \tag{8}$$

Here $|a^+_i|$ denotes the number of tokens in the corrected claim. For an episode batch $\mathcal{B}$, we weight both losses by the corresponding propagation weight w_i and combine them with the SFT objective

$$\mathcal{L} = \mathcal{L}_{\mathrm{SFT}} + \frac{\lambda}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} w_i \left(\mathcal{L}^{(i)}_{\mathrm{corr}} + \mathcal{L}^{(i)}_{\mathrm{OPD}}\right), \tag{9}$$

where λ controls the overall contribution of frontier supervision. Training episodes and their weights are periodically refreshed using the updated generator.

3.4 THEORETICAL PROPERTIES OF PROPAGATION-WEIGHTED LEARNING

To clarify how propagation-weighted supervision guides learning, we first interpret the propagation weight and then characterize the training target induced by local correction and privileged self-distillation. For a fixed episode and reference scorer, define the policy-specific cost-to-go and paired intervention value as

$$Q_i(a) = \mathbb{E}_{Y\sim\pi_\theta(\cdot|\mathcal{I}_i, s_{\tau_i}, a)}[C_i(Y)], \qquad \Delta C_i = Q_i(a^-_i) - Q_i(a^+_i). \tag{10}$$

Here Y contains only the newly generated suffix following $s_{\tau_i} \oplus a$, and the appended claim itself is excluded from scoring. Hence, ΔC_i measures the change in downstream clinical cost associated with replacing the erroneous claim by its correction under a fixed continuation policy. Under the deterministic decoding used in our implementation, this quantity reduces to the realized difference $C^-_i - C^+_i$ in Eq. 6.

Theorem 1 (Propagation-weighted local risk control) *Fix the contexts, correction tokens, detached teacher distributions, nonnegative weights w_i, and continuation costs $\mathcal{Q}_{i,t}(v) \in [0,1]$, where t indexes a position in the corrected claim and v denotes a candidate token at that position. Let $\delta_{a^+_{i,t}}$ denote the point-mass distribution on the correction token $a^+_{i,t}$ and define $m_{i,t} = (\delta_{a^+_{i,t}} + q_{i,t})/2$. Then*

$$-\log p_{i,t}(a^+_{i,t}) + D_{\mathrm{KL}}(q_{i,t}\|p_{i,t}) = 2D_{\mathrm{KL}}(m_{i,t}\|p_{i,t}) + 2\,\mathrm{JSD}(\delta_{a^+_{i,t}}, q_{i,t}), \tag{11}$$

where JSD *denotes the Jensen–Shannon divergence. Let μ select an episode uniformly and then a token position uniformly within its corrected claim. For any candidate token distribution $r_{i,t}$, define*

$$J_w(r) = \mathbb{E}_{(i,t)\sim\mu}\left[w_i \sum_{v\in\mathcal{V}} r_{i,t}(v)\mathcal{Q}_{i,t}(v)\right], \quad D_{\mathrm{mix},w} = \mathbb{E}_{(i,t)\sim\mu}\left[w_i D_{\mathrm{KL}}(m_{i,t}\|p_{i,t})\right]. \tag{12}$$

Let $W = \mathbb{E}_\mu[w_i]$. The deviation of the student's propagation-weighted local clinical risk from the average risk of the hard correction target and the privileged teacher is bounded by

$$\left| J_w(p) - \frac{J_w(\delta) + J_w(q)}{2} \right| \le \sqrt{\frac{W D_{\mathrm{mix},w}}{2}}. \tag{13}$$

Here, $J_w(p)$, $J_w(q)$, and $J_w(\delta)$ denote $J_w(r)$ evaluated at $r_{i,t} = p_{i,t}$, $q_{i,t}$, and $\delta_{a_{i,t}^+}$, respectively.

The mixed-target identity shows that local correction and self-distillation train the student toward a mixture of the hard correction target and the privileged teacher distribution, while the JSD term contributes no student gradient. The risk bound further shows that reducing the propagation-weighted divergence to this combined target controls the deviation of the student's weighted local clinical risk from the corresponding combined-target risk. Together, these results connect the implemented propagation-weighted objective to local clinical risk under the stated assumptions. Complete proofs and sensitivity analyses are provided in Appendix A.

4 EXPERIMENTS

Datasets. We evaluate CEF-OPD on two widely used medical report generation benchmarks MIMIC-CXR (Johnson et al., 2019) and IU-Xray (Demner-Fushman et al., 2016). MIMIC-CXR is a large-scale radiology report generation dataset containing 377,110 chest X-ray images and 227,835 reports collected at Beth Israel Deaconess Medical Center. Following the preprocessing and data split protocol of R2Gen (Chen et al., 2020), we use 270,790, 2,130, and 3,858 samples for training, validation, and testing, respectively. IU-Xray comprises 3,955 reports and 7,470 chest X-ray images with frontal and lateral views. Following the split protocol of previous work (Chen et al., 2021; Liu et al., 2021), we use 2,069, 296, and 590 samples for training, validation, and testing, respectively.

Evaluation metrics. We evaluate generated reports using natural language generation (NLG) and clinical efficacy (CE) metrics (Jin et al., 2024; Liu et al., 2025; 2026). For NLG evaluation, we report BLEU-1 through BLEU-4 (Papineni et al., 2002), METEOR (Banerjee & Lavie, 2005), ROUGE-L (Lin, 2004), and CIDEr (Vedantam et al., 2015) to assess textual similarity between generated and reference reports. For CE evaluation, CheXbert (Smit et al., 2020) extracts 14 observation labels, from which we compute sample-averaged precision, recall, and F1 following the PromptMRG (Jin et al., 2024) implementation. We additionally report GREEN (Ostmeier et al., 2024) and RadGraph Error Rate (RG-ER) (Jain et al., 2021) for clinical evaluation. GREEN assesses clinically significant discrepancies between generated and reference reports, whereas RG-ER measures entity- and relation-level clinical errors.

Implementation details. We adopt BiomedCLIP (Zhang et al., 2023) as the visual encoder and LLaMA-2-7B-Chat (Touvron et al., 2023) as the language decoder. A trainable projection followed by LayerNorm maps the visual features from 1,024 dimensions to the 4,096-dimensional LLaMA embedding space. We first train the SFT baseline for five epochs and subsequently initialize CEF-OPD post-training from the resulting checkpoint. Training uses AdamW with a learning rate of 2×10^{-5} and weight decay of 0.05 on two NVIDIA RTX A6000 48GB GPUs. During CEF-OPD post-training, the clinical error extractor selects the earliest reference-supported explicit commission error with a confidence threshold of 0.9, and the frontier-supervision weight is set to $\lambda = 1.0$ for all experiments.

4.1 MAIN RESULTS

Table 1 compares CEF-OPD with representative radiology report generation methods, including memory-based approaches (Chen et al., 2020; 2021), knowledge- and contrastive-learning methods (Li et al., 2023; Huang et al., 2023; Liu et al., 2024a), diagnosis-guided generation (Jin et al., 2024), and LLM-based methods (Liu et al., 2024b). Most existing approaches mainly improve visual-textual representations, inject clinical knowledge, or strengthen decoding guidance. In contrast, CEF-OPD focuses on errors arising under the model's own autoregressive context and explicitly weights their correction according to downstream clinical consequences.

On MIMIC-CXR, CEF-OPD achieves stronger overall clinical efficacy while maintaining competitive NLG performance. Compared with PromptMRG, CEF-OPD yields higher precision and F1 score, suggesting that propagation-aware correction helps suppress unsupported clinical statements rather than simply increasing predicted findings. Although its BLEU-4 score is not the highest, this is consistent with our objective, which prioritizes clinically consequential corrections rather than direct n-gram matching. On IU-Xray, CEF-OPD achieves the best results across all reported NLG and CE metrics. Overall, these results show that using the subsequent impact of clinical errors to guide corrective supervision improves the reliability of radiology report generation.

Table 1: Overall performance comparison on the MIMIC-CXR and IU-Xray datasets. NLG denotes natural language generation metrics, where B-1–B-4, MTR, and R-L denote BLEU-1–BLEU-4, METEOR, and ROUGE-L, respectively. CE denotes clinical efficacy, where P, R, and F1 represent clinical precision, recall, and F1 score, respectively. The best and second-best results are shown in bold and underlined, respectively.

Method	NLG Metrics						CE Metrics		
	B-1	B-2	B-3	B-4	MTR	R-L	P	R	F1
MIMIC-CXR									
R2Gen (Chen et al., 2020)	0.353	0.218	0.145	0.103	0.142	0.270	0.333	0.273	0.276
R2GenCMN (Chen et al., 2021)	0.353	0.218	0.148	0.106	0.142	0.278	0.344	0.275	0.278
DCL (Li et al., 2023)	–	–	–	0.109	0.150	0.284	0.471	0.352	0.373
KiUT (Huang et al., 2023)	0.393	0.243	0.159	0.113	0.160	0.285	0.371	0.318	0.321
Bootstrapping (Liu et al., 2024b)	0.402	0.262	0.180	<u>0.128</u>	**0.175**	0.291	0.465	0.482	0.473
PromptMRG (Jin et al., 2024)	0.398	–	–	0.112	0.157	0.268	<u>0.501</u>	**0.509**	<u>0.476</u>
Multi-Grained (Liu et al., 2024a)	<u>0.406</u>	<u>0.267</u>	<u>0.190</u>	**0.141**	0.163	<u>0.309</u>	0.457	0.337	0.330
DDaTR (Song et al., 2025)	0.401	–	–	0.113	0.162	0.272	0.438	0.465	0.441
CEF-OPD (ours)	**0.425**	**0.278**	**0.194**	0.127	<u>0.171</u>	**0.314**	**0.529**	<u>0.491</u>	**0.485**
IU-Xray									
DCL (Li et al., 2023)	0.354	–	–	0.074	0.152	0.267	0.257	0.202	0.177
CvT2DistilGPT2 (Nicolson et al., 2023)	<u>0.473</u>	<u>0.304</u>	<u>0.224</u>	<u>0.175</u>	<u>0.200</u>	<u>0.376</u>	<u>0.325</u>	0.166	0.155
PromptMRG (Jin et al., 2024)	0.401	–	–	0.098	0.160	0.281	0.213	0.229	0.211
DDaTR (Song et al., 2025)	0.421	–	–	0.103	0.160	0.307	0.304	<u>0.291</u>	<u>0.262</u>
ARDGen (Ahsan et al., 2025)	0.431	–	–	0.113	0.170	0.311	0.229	0.245	0.236
CEF-OPD (ours)	**0.492**	**0.331**	**0.242**	**0.194**	**0.208**	**0.387**	**0.337**	**0.303**	**0.285**

4.2 Ablation study

Contribution of CEF-OPD components. We ablate the main training components of CEF-OPD on MIMIC-CXR. The ablation considers three components: reference-supported frontier correction, privileged on-policy distillation (OPD), and propagation-aware weighting based on the estimated downstream benefit ΔC. Continued-SFT yields only marginal improvements over SFT, indicating that the gains of CEF-OPD cannot be explained by additional post-training alone. Removing either the reference-supported correction or privileged on-policy distillation consistently degrades clinical performance, showing that explicit correction and soft privileged supervision provide complementary signals. Notably, replacing the propagation-aware weights with uniform weights also leads to a performance drop. Additional analyses of supervision efficiency and sensitivity to the frontier supervision weight λ are provided in Appendix B. This comparison indicates that treating all detected frontier errors equally is suboptimal and supports our use of the estimated downstream consequence to allocate stronger supervision to more consequential errors. The full model achieves the best overall clinical performance, with improved precision and recall, suggesting that the gains arise from more reliable clinical predictions.

Table 2: Ablation of the correction, OPD, and propagation-aware weighting components of CEF-OPD on MIMIC-CXR.

	Components			Metrics				
Setting	Correction	OPD	ΔC weight	GREEN↑	RG-ER↓	P↑	R↑	F1↑
SFT	×	×	×	0.602	0.402	0.486	0.469	0.458
Continued-SFT	×	×	×	0.606	0.397	0.490	0.471	0.461
Weighted Correction	×	✓	✓	0.623	0.370	0.501	0.476	0.469
Weighted OPD	✓	×	✓	0.629	0.361	0.509	0.480	0.473
Uniform Weight	✓	✓	×	<u>0.638</u>	<u>0.347</u>	<u>0.518</u>	<u>0.484</u>	<u>0.478</u>
CEF-OPD	✓	✓	✓	**0.651**	**0.326**	**0.529**	**0.491**	**0.485**

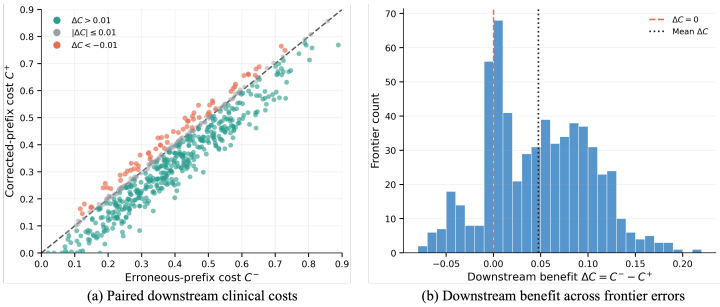
Comparison of corrective training strategies. We further compare CEF-OPD with alternative strategies for mitigating autoregressive exposure mismatch and providing privileged supervision. *Scheduled sampling*, which exposes the model to its own predictions during training, does not improve over SFT, indicating that such exposure alone is insufficient when supervision is not targeted to clinically meaningful errors. *Frozen-teacher OPD* performs on-policy distillation using a fixed privileged teacher and yields moderate improvements, suggesting that privileged supervision is beneficial but that a fixed teacher provides limited adaptation to evolving student-generated states. *Fact-OPD* introduces sample-specific reference facts as privileged guidance and further improves clinical performance, showing the benefit of structured clinical supervision. *Frontier-uniform* applies correction and OPD at clinically identified error frontiers while assigning equal weight to all accepted episodes, achieving stronger results but remaining below CEF-OPD. In contrast, *CEF-OPD* additionally weights frontier supervision according to the estimated downstream benefit ΔC, indicating that the estimated propagation effect provides useful information beyond locating the intervention point alone. CEF-OPD achieves the best GREEN, RadGraph error rate (RG-ER), and Clinical Efficacy F1 score (CE-F1), supporting the joint use of clinical-semantic frontier selection, privileged guidance, and propagation-aware weighting.

Table 3: Comparison of corrective training strategies on IU-Xray across both natural language generation and clinical efficacy metrics.

Method	GREEN	RG-ER	CE-F1	BLEU-4	METEOR	CIDEr
SFT	0.676	0.349	0.248	0.096	0.193	0.394
Scheduled sampling	0.653	0.342	0.239	0.091	0.187	0.376
Frozen-teacher OPD	0.684	0.334	0.257	0.100	0.192	0.401
Fact-OPD	0.691	0.319	0.264	0.105	0.188	**0.513**
Frontier-uniform	<u>0.699</u>	<u>0.306</u>	<u>0.274</u>	<u>0.181</u>	<u>0.202</u>	0.456
CEF-OPD	**0.707**	**0.293**	**0.285**	**0.194**	**0.208**	<u>0.475</u>

4.3 Model analysis

Downstream effect of frontier correction. We examine whether correcting a frontier claim affects subsequent clinical generation rather than only removing the detected error itself. For each accepted frontier, we generate paired continuations from the erroneous and corrected branches while keeping the image, pre-error prefix, generator snapshot, and scoring targets fixed. The corrected claim itself is excluded from downstream scoring, so the comparison isolates changes in the subsequently generated clinical content.



(a) Paired downstream clinical costs (b) Downstream benefit across frontier errors

Figure 2: Paired-continuation analysis of frontier correction.

Table 4: Paired-continuation summary on MIMIC-CXR.

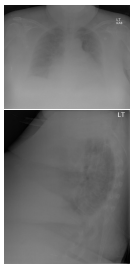
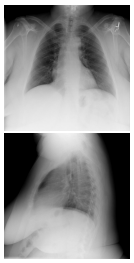
Statistic	Value
Mean C^-	0.394
Mean C^+	0.347
Mean ΔC	0.048
Median ΔC	0.047
$\Delta C > 0.01$	68.8%
$\lvert\Delta C\rvert \leq 0.01$	20.7%
$\Delta C < -0.01$	10.5%

Figure 2 visualizes the paired downstream costs and the distribution of the resulting benefit $\Delta C = C^- - C^+$, with the corresponding statistics summarized in Table 4. The mean downstream cost decreases from 0.394 for the erroneous branch to 0.347 for the corrected branch, yielding a mean ΔC of 0.048. At the individual-frontier level, 68.8% of frontiers have a clear positive benefit ($\Delta C > 0.01$), 20.7% show only a small change ($|\Delta C| \leq 0.01$), and 10.5% have a negative difference ($\Delta C < -0.01$). The broad distribution of ΔC shows that frontier errors differ substantially in their downstream consequences: correcting some frontiers changes later clinical predictions considerably, whereas others have little measurable effect. This heterogeneity supports frontier-specific propagation

weighting, showing that different errors have different effects on subsequent clinical generation and should not receive identical corrective supervision. A qualitative paired-continuation example illustrating how a frontier correction changes subsequent clinical content is provided in Appendix C.

Entity-level qualitative analysis. We further examine how CEF-OPD affects a specific clinical decision. We select two IU-Xray cases involving the same focal clinical entity, heart-size/cardiomegaly status, but with opposite reference assertions. In Case 1, the reference indicates stable cardiomegaly, whereas both SFT and Fact-OPD incorrectly describe the heart size as normal. CEF-OPD corrects this focal contradiction and predicts an enlarged heart, consistent with the reference finding. The corresponding positive downstream benefit ($\Delta C = +0.11$) further indicates that correcting this frontier claim is associated with a lower clinical cost in the subsequent continuation. Case 2 provides a matched failure mode for the same entity. Here, the reference specifies a cardiomediastinal silhouette within normal limits, and both SFT and Fact-OPD preserve the correct normal heart-size assertion. In contrast, CEF-OPD incorrectly predicts an enlarged heart. The negative downstream benefit ($\Delta C = -0.04$) is consistent with the absence of a beneficial propagation effect for this intervention. Together, these cases show that CEF-OPD can alter the local clinical decision at the error frontier.

Table 5: Qualitative analysis of focal correction and matched failure cases. Both cases examine heart-size/cardiomegaly status, with opposite reference assertions. We compare the reference report with outputs from SFT, Fact-OPD, and CEF-OPD. Blue denotes heart-size predictions consistent with the reference, whereas orange denotes predictions that contradict the reference. C^- and C^+ denote the downstream costs following the erroneous and corrected frontier claims, respectively, and $\Delta C = C^- - C^+$ denotes the estimated downstream benefit of the correction.

Case 1: CXR2285_IM-0870	$\mathbf{C^- = 0.36,\ C^+ = 0.25,\ \Delta C = +0.11}$
	**Ground Truth:** Sternotomy XXXX and mediastinal postsurgical changes. Stable cardiomegaly. Crowded bronchovascular and interstitial markings XXXX related to low lung volumes and technique. Grossly stable appearance of the lungs compared to prior exam without overt edema or gross airspace consolidation.
	SFT: The cardiomediastinal silhouette is within normal limits. There is no pneumothorax or large pleural effusion. There is no focal airspace disease. There are degenerative changes of the thoracic spine.
	Fact-OPD: The heart is normal in size. The mediastinum is unremarkable. The lungs are clear. There is no pneumothorax or pleural effusion.
	CEF-OPD: The heart size is enlarged. The mediastinal contours are within normal limits. The lungs are clear. There is no pleural effusion or pneumothorax.
Case 2: CXR3817_IM-1925	$\mathbf{C^- = 0.27,\ C^+ = 0.31,\ \Delta C = -0.04}$
	**Ground Truth:** Lungs are clear bilaterally with no focal infiltrate, pleural effusion, or pneumothoraces. Cardiomediastinal silhouette is within normal limits. XXXX and soft tissues are unremarkable.
	SFT: The cardiomediastinal silhouette and pulmonary vasculature are within normal limits in size. The lungs are clear of focal airspace disease, pneumothorax, or pleural effusion. There are no acute bony findings.
	Fact-OPD: The cardiomediastinal silhouette and pulmonary vasculature are within normal limits in size. The lungs are clear of focal airspace disease, pneumothorax, or pleural effusion. There are no acute bony findings.
	CEF-OPD: The heart size is enlarged. The mediastinal contours are within normal limits. The lungs are clear bilaterally. No pleural effusion or pneumothorax is seen.

5 Conclusion

We proposed CEF-OPD to address clinically consequential errors arising during autoregressive radiology report generation. The method identifies the earliest reference-supported contradiction, constructs a minimal correction, and uses paired continuations to estimate its downstream clinical impact. The estimated benefit is then used to weight local correction and privileged self-distillation from the pre-error state. Experiments on MIMIC-CXR and IU-Xray demonstrate improved clinical report quality, while ablation and paired-continuation analyses verify the contributions of localized correction and propagation-aware weighting. Qualitative results further illustrate both successful entity-level corrections and remaining failure cases. Overall, CEF-OPD shows that modeling downstream error consequences can improve the clinical reliability of radiology report generation.

AI USE STATEMENT

Generative AI tools were used solely for language polishing, including improving grammar, clarity, readability, and academic writing style. All AI-assisted revisions were carefully reviewed and verified by the authors, who take full responsibility for the final content of the paper.

ETHICS STATEMENT

This work uses publicly available and de-identified chest X-ray datasets for research purposes. The proposed method is intended to support radiology report generation research and is not designed for direct clinical decision-making.

REPRODUCIBILITY STATEMENT

We provide the model architecture, training procedure, evaluation protocol, and implementation details in the main paper and appendix. Additional experimental details are included in the supplementary material.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

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

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

Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. *Advances in Neural Information Processing Systems*, 28, 2015.

Kai-Wei Chang, Akshay Krishnamurthy, Alekh Agarwal, Hal Daumé III, and John Langford. Learning to search better than your teacher. In *International Conference on Machine Learning*, pp. 2058–2066. PMLR, 2015.

Zhihong Chen, Yan Song, Tsung-Hui Chang, and Xiang Wan. Generating radiology reports via memory-driven transformer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 1439–1449, 2020.

Zhihong Chen, Yaling Shen, Yan Song, and Xiang Wan. Cross-modal memory networks for radiology report generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International joint Conference on Natural Language Processing (volume 1: long papers)*, pp. 5904–5914, 2021.

Jean-Benoit Delbrouck, Pierre Chambon, Christian Bluethgen, Emily Tsai, Omar Almusa, and Curtis Langlotz. Improving the factual correctness of radiology report generation with semantic rewards. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 4348–4360, 2022.

Jean-Benoit Delbrouck, Pierre Chambon, Zhihong Chen, Maya Varma, Andrew Johnston, Louis Blankemeier, Dave Van Veen, Tan Bui, Steven Truong, and Curtis Langlotz. RadGraph-XL: A large-scale expert-annotated dataset for entity and relation extraction from radiology reports. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 12902–12915, 2024.

Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2016.

Yuqian Fu, Haohuan Huang, Kaiwen Jiang, Jiacai Liu, Zhuo Jiang, Yuanheng Zhu, and Dongbin Zhao. Revisiting on-policy distillation: Empirical failure modes and simple fixes. *arXiv preprint arXiv:2603.25562*, 2026.

Tianxing He, Jingzhao Zhang, Zhiming Zhou, and James Glass. Exposure bias versus self-recovery: Are distortions really incremental for autoregressive text generation? In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 5087–5102, 2021.

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

Zhongzhen Huang, Xiaofan Zhang, and Shaoting Zhang. KiUT: Knowledge-injected U-transformer for radiology report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19809–19818. IEEE, 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 2607–2615, 2024.

Alistair EW Johnson, Tom J Pollard, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Yifan Peng, Zhiyong Lu, Roger G Mark, Seth J Berkowitz, and Steven Horng. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*, 2019.

Mingjie Li, Bingqian Lin, Zicong Chen, Haokun Lin, Xiaodan Liang, and Xiaojun Chang. Dynamic graph enhanced contrastive learning for chest X-ray report generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3334–3343. IEEE, 2023.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, 2004.

Aohan Liu, Yuchen Guo, Jun-Hai Yong, and Feng Xu. Multi-grained radiology report generation with sentence-level image-language contrastive learning. *IEEE Transactions on Medical Imaging*, 43(7):2657–2669, 2024a.

Chang Liu, Yuanhe Tian, Weidong Chen, Yan Song, and Yongdong Zhang. Bootstrapping large language models for radiology report generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 18635–18643, 2024b.

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

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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 552–562. Springer, 2025.

Yunyi Liu, Yingshu Li, Tong Chen, Lingqiao Liu, Lei Wang, and Luping Zhou. SAT-RRG: LLM-guided self-adaptive training for radiology report generation with token-level push-pull optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 35363–35372, 2026.

Yasuhide Miura, Yuhao Zhang, Emily Tsai, Curtis Langlotz, and Dan Jurafsky. Improving factual completeness and consistency of image-to-text radiology report generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5288–5304, 2021.

Aaron Nicolson, Jason Dowling, and Bevan Koopman. Improving chest X-ray report generation by leveraging warm starting. *Artificial Intelligence in Medicine*, 144:102633, 2023.

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

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

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 627–635. JMLR Workshop and Conference Proceedings, 2011.

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 *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 1500–1519, 2020.

Shanshan Song, Hui Tang, Honglong Yang, and Xiaomeng Li. DDaTR: Dynamic difference-aware temporal residual network for longitudinal radiology report generation. *IEEE Transactions on Medical Imaging*, 44(12):5345–5357, 2025.

Ryutaro Tanno, David GT Barrett, Andrew Sellergren, Sumedh Ghaisas, Sumanth Dathathri, Abigail See, Johannes Welbl, Charles Lau, Tao Tu, Shekoofeh Azizi, et al. Collaboration between clinicians and vision–language models in radiology report generation. *Nature Medicine*, 31(2):599–608, 2025.

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.

Ramakrishna Vedantam, C Lawrence Zitnick, and Devi Parikh. CIDEr: Consensus-based image description evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4566–4575, 2015.

Yuanyi Wang, Su Lu, Yanggan Gu, Pengkai Wang, Yifan Yang, Zhaoyi Yan, Congkai Xie, Jianmin Wu, and Hongxia Yang. Not all disagreement is learnable: Token teachability in on-policy distillation. *arXiv preprint arXiv:2605.26844*, 2026.

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

Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. *arXiv preprint arXiv:2303.00915*, 2023.

Siyan Zhao, Zhihui Xie, Mengchen Liu, Jing Huang, Guan Pang, Feiyu Chen, and Aditya Grover. Self-distilled reasoner: On-policy self-distillation for large language models. *arXiv preprint arXiv:2601.18734*, 2026.

A Theoretical Analysis and Complete Proofs

This appendix provides the derivations supporting the paired intervention value, mixed supervision target, and weighted local risk bound in Section 3.4. It also examines how perturbations in clinical scores affect the propagation weights and the auxiliary objective. We use the notation of the main text throughout. All logarithms are natural, and $\lambda \geq 0$. During episode collection, θ denotes a fixed generator snapshot. During an optimization step, the collected episodes and weights remain fixed, while the current shared generator parameters are updated through the student and SFT branches.

A.1 Bounded Clinical Cost and Scoring Targets

Fix an accepted episode i, including $\mathcal{I}_i$, F_i, s_{τ_i}, a_i^-, and a_i^+. Both branches use the same remaining reference set $F_i^{\mathrm{rem}} = F_i \setminus \big(g(s_{\tau_i}) \cup G_i^+\big)$. From Eq. 4,

$$O_i(Y) \subseteq F_i^{\mathrm{rem}} \subseteq F_i, \qquad D_i(Y) \subseteq F_i. \tag{14}$$

Consequently, $O_i(Y) \cup D_i(Y) \subseteq F_i$. Since an accepted episode satisfies $|F_i| \geq 1$, the clinical cost in Eq. 5 obeys

$$0 \leq C_i(Y) = \frac{|O_i(Y) \cup D_i(Y)|}{|F_i|} \leq 1. \tag{15}$$

The union counts each reference fact at most once, including when a fact is both omitted and contradicted. It follows that the realized branch difference and its positive part satisfy

$$-1 \leq C_i(Y_i^-) - C_i(Y_i^+) \leq 1, \qquad 0 \leq w_i \leq 1. \tag{16}$$

The appended claim is excluded from Y_i^- and Y_i^+. In addition, facts covered by the corrected claim are removed from omission scoring in both branches. Thus the direct replacement of a_i^- by a_i^+ does not itself reduce the scored cost. If a conflicting fact is generated again in the suffix, it is counted through $D_i(Y)$ because contradictions are checked against the full reference set. For an empty suffix, $g(Y) = \varnothing$ and $C_i(Y) = |F_i^{\mathrm{rem}}|/|F_i|$. Empty generation therefore incurs an omission cost whenever reference targets remain. For fixed F_i, F_i^{rem}, and contradiction rules, two suffixes with the same normalized fact set have the same cost. Indeed, $g(Y) = g(Y')$ implies $O_i(Y) = O_i(Y')$ and $D_i(Y) = D_i(Y')$. This is a property of the scorer. It does not require the generator to assign the same continuation distribution to semantically equivalent prefix strings.

A.2 Interpretation of the Paired Intervention Value

The policy π_θ denotes the actual continuation rule, including its decoding settings and generation limit. In particular, greedy decoding defines a deterministic policy, even though the underlying model probabilities p_θ need not be point masses. For a fixed episode and policy, let $Y_i(a, U)$ denote the newly generated suffix after appending claim a to s_{τ_i}, where U contains any randomness used by the decoder. Then Eq. 10 can be written as

$$Q_i(a) = \mathbb{E}_U\big[C_i(Y_i(a, U))\big]. \tag{17}$$

Therefore, $\Delta C_i = Q_i(a_i^-) - Q_i(a_i^+) = \mathbb{E}_U\big[C_i(Y_i(a_i^-, U)) - C_i(Y_i(a_i^+, U))\big]$. The last expression represents a coupling of the two continuation branches. Its expectation depends only on their marginal distributions, so shared decoding randomness is not required for the identity. The images, pre-error prefix, reference sets, scorer, and continuation policy are held fixed. The quantity ΔC_i consequently measures the expected downstream cost change associated with this particular claim replacement.

For deterministic decoding, let $Y_i(a)$ denote the unique suffix produced under the fixed decoding rule. Each branch distribution is concentrated on that suffix, giving

$$Q_i(a_i^\sigma) = C_i(Y_i^\sigma), \qquad \Delta C_i = C_i^- - C_i^+, \qquad \sigma \in \{-, +\}. \tag{18}$$

This establishes agreement between the expected-value definition in Eq. 10 and the implemented difference in Eq. 6. A positive difference identifies a reduction under this fixed intervention and continuation policy. A zero difference can occur even when the frontier claim is successfully corrected, because its direct correction is excluded from the cost.

A.3 Mixed Supervision Target and Shared-StopGrad Gradient

Fix an episode i and a position t within its corrected claim. Let $\mathcal{V}$ denote the model vocabulary. The student distribution $p_{i,t}$ and detached teacher distribution $q_{i,t}$ are evaluated on the contexts in Eq. 7. Define

$$m_{i,t} = \frac{\delta_{a^+_{i,t}} + q_{i,t}}{2}, \quad \ell_{i,t} = -\log p_{i,t}(a^+_{i,t}) + D_{\mathrm{KL}}(q_{i,t} \| p_{i,t}). \tag{19}$$

For finite losses, assume $p_{i,t}(v) > 0$ whenever $m_{i,t}(v) > 0$. This holds for a full softmax with finite logits. We use the usual convention that terms with zero probability contribute zero.

Proof of the mixed-target identity. Because $\delta_{a^+_{i,t}}$ is a point mass, its entropy is zero and

$$-\log p_{i,t}(a^+_{i,t}) = D_{\mathrm{KL}}(\delta_{a^+_{i,t}} \| p_{i,t}). \tag{20}$$

Splitting the logarithm in each KL divergence through $m_{i,t}$ gives

$$\begin{aligned} \ell_{i,t} &= D_{\mathrm{KL}}(\delta_{a^+_{i,t}} \| m_{i,t}) + D_{\mathrm{KL}}(q_{i,t} \| m_{i,t}) + \sum_{v \in \mathcal{V}} \left[\delta_{a^+_{i,t}}(v) + q_{i,t}(v)\right] \log \frac{m_{i,t}(v)}{p_{i,t}(v)} \\ &= 2\,\mathrm{JSD}(\delta_{a^+_{i,t}}, q_{i,t}) + 2 D_{\mathrm{KL}}(m_{i,t} \| p_{i,t}), \end{aligned} \tag{21}$$

where $\mathrm{JSD}(\delta_{a^+_{i,t}}, q_{i,t}) = \frac{1}{2} D_{\mathrm{KL}}(\delta_{a^+_{i,t}} \| m_{i,t}) + \frac{1}{2} D_{\mathrm{KL}}(q_{i,t} \| m_{i,t})$. This proves Eq. 11. With the correction token and teacher target fixed, the JSD term is independent of the student prediction. The unconstrained minimizer over student distributions is therefore $p_{i,t} = m_{i,t}$. At this minimum, the correction token receives probability $m_{i,t}(a^+_{i,t}) = (1 + q_{i,t}(a^+_{i,t}))/2 \geq 1/2$. This describes the local distributional target rather than assuming that a shared-parameter model attains every such minimum simultaneously.

Gradient through the shared model. During differentiation, the correction token and complete teacher distribution are held fixed. The resulting gradient is

$$\nabla_\theta \ell_{i,t} = -\nabla_\theta \log p_{i,t}(a^+_{i,t}) - \sum_{v \in \mathcal{V}} q_{i,t}(v) \nabla_\theta \log p_{i,t}(v) = -2 \sum_{v \in \mathcal{V}} m_{i,t}(v) \nabla_\theta \log p_{i,t}(v). \tag{22}$$

Equivalently, for student logits $z_{i,t}$ with $p_{i,t} = \mathrm{softmax}(z_{i,t})$,

$$\nabla_{z_{i,t}} \ell_{i,t} = 2p_{i,t} - \delta_{a^+_{i,t}} - q_{i,t} = 2(p_{i,t} - m_{i,t}). \tag{23}$$

Stop-gradient removes differentiation through the computation of $q_{i,t}$. The teacher is nevertheless recomputed from the shared parameters and can change after an update. Thus the mixed-target equivalence holds at each update without requiring a permanently frozen teacher checkpoint.

A.4 Weighted Local Risk Bound and Objective Correspondence

For a nonempty episode batch $\mathcal{B}$, define the distribution used in the local analysis by

$$\mu(i,t) = \frac{1}{|\mathcal{B}|\,|a^+_i|}, \qquad i \in \mathcal{B}, \quad t = 1, \ldots, |a^+_i|. \tag{24}$$

It samples an episode uniformly and then a position uniformly within that episode's corrected claim. This matches the per-episode length normalization in Eq. 8. The same construction applies to a fixed episode cache by replacing $\mathcal{B}$ with the cache. Uniform sampling over all concatenated correction tokens would define a different distribution when claim lengths differ.

Fix a common continuation rule and costs $\mathcal{Q}_{i,t}(v) \in [0,1]$. The costs are held fixed when comparing candidate next-token distributions. The definitions in Eq. 12 give

$$J_w(m) = \mathbb{E}_{(i,t)\sim\mu}\left[w_i \sum_{v \in \mathcal{V}} \frac{\delta_{a^+_{i,t}}(v) + q_{i,t}(v)}{2} \mathcal{Q}_{i,t}(v)\right] = \frac{J_w(\delta) + J_w(q)}{2}, \tag{25}$$

where $J_w(\delta) = \mathbb{E}_{(i,t)\sim\mu}[w_i \mathcal{Q}_{i,t}(a^+_{i,t})]$.

Proof of the local risk bound. For any two vocabulary distributions u and r, let

$$\mathrm{TV}(u, r) = \frac{1}{2} \sum_{v \in \mathcal{V}} |u(v) - r(v)|. \tag{26}$$

The positive and negative parts of $u - r$ have equal mass, each equal to $\mathrm{TV}(u, r)$. Since every token cost lies in $[0, 1]$,

$$\left| \sum_{v \in \mathcal{V}} [u(v) - r(v)] \mathcal{Q}_{i,t}(v) \right| \leq \mathrm{TV}(u, r). \tag{27}$$

Pinsker's inequality with natural logarithms yields

$$\mathrm{TV}(p_{i,t}, m_{i,t}) \leq \sqrt{\frac{D_{\mathrm{KL}}(m_{i,t} \| p_{i,t})}{2}}. \tag{28}$$

For completeness, take $A = \{v \in \mathcal{V} : m_{i,t}(v) > p_{i,t}(v)\}$, $u = m_{i,t}(A)$, and $z = p_{i,t}(A)$. The log-sum inequality gives $D_{\mathrm{KL}}(m_{i,t} \| p_{i,t}) \geq \mathrm{kl}(u \| z)$, where kl denotes Bernoulli KL divergence. For fixed $z \in (0, 1)$, this divergence has value and first derivative zero at $u = z$, and its second derivative with respect to u is $1/[u(1-u)] \geq 4$. Hence $\mathrm{kl}(u \| z) \geq 2(u - z)^2$. Since $u - z = \mathrm{TV}(m_{i,t}, p_{i,t})$, Eq. 28 follows. Boundary cases follow by continuity or have infinite KL divergence.

Combining Eqs. 27 and 28, multiplying by $w_i \geq 0$, and averaging gives

$$\begin{aligned} |J_w(p) - J_w(m)| &\leq \mathbb{E}_\mu \left[w_i \sqrt{\frac{D_{\mathrm{KL}}(m_{i,t} \| p_{i,t})}{2}} \right] = \frac{1}{\sqrt{2}} \mathbb{E}_\mu \left[\sqrt{w_i} \sqrt{w_i D_{\mathrm{KL}}(m_{i,t} \| p_{i,t})} \right] \\ &\leq \sqrt{\frac{\mathbb{E}_\mu[w_i] \mathbb{E}_\mu[w_i D_{\mathrm{KL}}(m_{i,t} \| p_{i,t})]}{2}} = \sqrt{\frac{W D_{\mathrm{mix},w}}{2}}. \end{aligned} \tag{29}$$

The penultimate step is Cauchy–Schwarz. Substituting Eq. 25 proves Eq. 13. If $W = 0$, nonnegative weights imply that $w_i = 0$ almost surely under μ, so both weighted risks and $D_{\mathrm{mix},w}$ vanish. No division by W is needed.

Correspondence with the implemented objective. Let $\mathcal{L}_{\mathrm{aux}}$ denote the auxiliary term in Eq. 9. By Eq. 24 and the mixed-target identity,

$$\mathcal{L}_{\mathrm{aux}} = \frac{\lambda}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} w_i \left(\mathcal{L}_{\mathrm{corr}}^{(i)} + \mathcal{L}_{\mathrm{OPD}}^{(i)} \right) = \lambda \mathbb{E}_\mu[w_i \ell_{i,t}] = 2\lambda D_{\mathrm{mix},w} + 2\lambda \mathbb{E}_\mu \left[w_i \, \mathrm{JSD}(\delta_{a_{i,t}^+}, q_{i,t}) \right]. \tag{30}$$

The second term contributes no gradient during the current update because the weights, correction tokens, and teacher outputs are detached. This is an exact identity for the batch distribution in Eq. 24. It concerns the auxiliary term, while the SFT term remains in the full training objective. The risk bound consequently applies to the corrected histories used by the implemented loss and to the combined supervision target. The contexts, weights, teacher targets, and continuation costs are fixed within this comparison.

A.5 Sensitivity to Clinical Scoring Perturbations

Fix the accepted episodes, generated suffixes, correction tokens, teacher targets, and current student parameters. Let $\widetilde{C}_i$ be a perturbed scorer with branchwise errors satisfying

$$\left| \widetilde{C}_i(Y_i^\sigma) - C_i(Y_i^\sigma) \right| \leq \epsilon_i^\sigma, \qquad \sigma \in \{-, +\}. \tag{31}$$

Define $\widetilde{\Delta C}_i = \widetilde{C}_i(Y_i^-) - \widetilde{C}_i(Y_i^+)$ and $\widetilde{w}_i = [\widetilde{\Delta C}_i]_+$. The triangle inequality gives

$$\left| \widetilde{\Delta C}_i - \Delta C_i \right| \leq \epsilon_i^- + \epsilon_i^+. \tag{32}$$

The positive-part function is nonexpansive, since $|[u]_+ - [v]_+| \leq |u - v|$ for all real u, v. Therefore,

$$|\widetilde{w}_i - w_i| \leq |\widetilde{\Delta C}_i - \Delta C_i| \leq \epsilon_i^- + \epsilon_i^+. \tag{33}$$

In particular, branchwise score errors bounded by ϵ produce weight errors bounded by 2ϵ. A positive intervention value remains positive when $\Delta C_i > \epsilon_i^- + \epsilon_i^+$. Values closer to zero can change sign under the allowed scoring perturbation.

Let $\ell_i = \mathcal{L}_{\text{corr}}^{(i)} + \mathcal{L}_{\text{OPD}}^{(i)}$ be the fixed, finite local loss. Since $\ell_i \geq 0$, replacing only the weights gives

$$\left|\widetilde{\mathcal{L}}_{\text{aux}} - \mathcal{L}_{\text{aux}}\right| = \left|\frac{\lambda}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}(\widetilde{w}_i - w_i)\ell_i\right| \leq \frac{\lambda}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}|\widetilde{w}_i - w_i|\ell_i \leq \frac{\lambda}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}(\epsilon_i^- + \epsilon_i^+)\ell_i. \tag{34}$$

If $\epsilon_i^\sigma \leq \epsilon$ and $\ell_i \leq L_{\max}$ throughout this batch, then

$$\left|\widetilde{\mathcal{L}}_{\text{aux}} - \mathcal{L}_{\text{aux}}\right| \leq 2\lambda\epsilon L_{\max}. \tag{35}$$

Equation 34 requires no uniform loss bound across model parameters. Equation 35 is conditional on the stated batchwise bound. The denominator is the fixed batch size, as in the implemented objective.

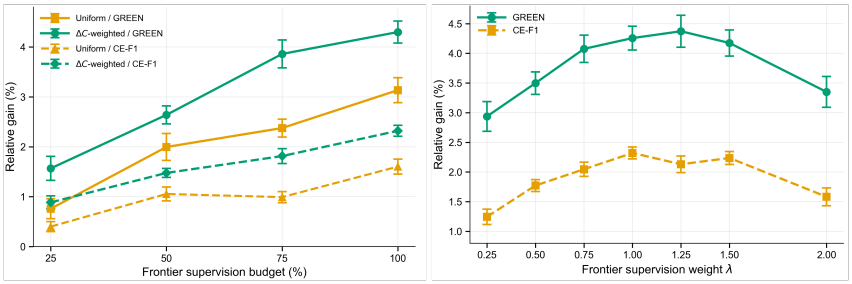
For the scorer in Eq. 5, define $E_i(Y) = O_i(Y) \cup D_i(Y)$. If a perturbed error set $\widetilde{E}_i(Y)$ is evaluated against the same reference set F_i, then

$$|\widetilde{C}_i(Y) - C_i(Y)| \leq \frac{|\widetilde{E}_i(Y) \triangle E_i(Y)|}{|F_i|}, \tag{36}$$

where $\triangle$ denotes symmetric set difference. Thus changes in the error status of at most h reference facts change the normalized score by at most $h/|F_i|$. These results concern score perturbations on fixed episodes. Changes to the extractor that also alter frontier selection, accepted corrections, teacher inputs, or generated suffixes require accounting for those additional changes separately.

B SUPERVISION EFFICIENCY AND SENSITIVITY ANALYSIS.

We further examine whether propagation-aware weighting improves the efficiency of frontier supervision and whether CEF-OPD is sensitive to the overall frontier-supervision weight λ. For the budget analysis, Uniform and ΔC-weighted variants use the same frontier episodes and optimization budget, while the available frontier supervision is restricted to 25%, 50%, 75%, and 100%. Figure 3(a) shows that increasing the supervision budget improves both GREEN and CE-F1, while ΔC-weighted supervision consistently yields larger gains than uniform weighting. This advantage across different budgets shows that propagation-aware weighting provides useful information for allocating corrective supervision.



(a) Supervision-budget efficiency (b) Sensitivity to frontier supervision weight

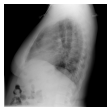
Figure 3: Supervision efficiency and sensitivity analysis on MIMIC-CXR. (a) Relative GREEN and CE-F1 gains under different frontier-supervision budgets. Uniform and ΔC-weighted variants use matched frontier episodes and optimization budgets. (b) Sensitivity to the overall frontier-supervision weight λ. Error bars denote standard deviations across repeated runs.

Figure 3(b) evaluates sensitivity to λ. Both GREEN and CE-F1 remain relatively stable over a broad intermediate range of λ, with performance decreasing mainly when the frontier objective is assigned either too little or excessively large weight. We select $\lambda = 1.0$ for all experiments, as it provides the best CE-F1 while achieving near-optimal GREEN within the stable region. This choice balances the two clinical metrics without relying on the peak of a single metric. Overall, the results show that CEF-OPD is robust to moderate variations in the frontier-supervision strength.

C QUALITATIVE ANALYSIS OF ERROR CORRECTION AND DOWNSTREAM PROPAGATION.

We conduct a qualitative case study to examine whether correcting a clinically important frontier error can influence subsequent report generation, rather than only fixing the local erroneous claim. We compare reports generated by SFT, Fact-OPD, Frontier-uniform, and CEF-OPD against the reference report, and further construct paired continuations from the same pre-error state using an erroneous frontier claim *Heart size is normal* and its corrected counterpart *Heart size is enlarged*. SFT incorrectly predicts a normal cardiac silhouette and captures only limited reference-consistent content. Fact-OPD recovers the increased interstitial markings but retains the same focal contradiction, illustrating that fact-level supervision alone does not necessarily correct the critical decision point. ROUGE-L further shows that CEF-OPD achieves the strongest report-level agreement with the reference, after frontier-uniform corrects the focal cardiac finding and CEF-OPD recovers additional reference-supported pulmonary abnormalities. The paired-continuation comparison provides a more direct view of the downstream effect: conditioning on the erroneous frontier claim leads to contradictions with the reference pulmonary findings, whereas the corrected claim yields a continuation that recovers the interstitial abnormality while preserving the absence of pleural effusion. Correspondingly, the downstream clinical cost decreases from $C^- = 0.48$ to $C^+ = 0.25$, giving $\Delta C = +0.23$. This example suggests that CEF-OPD can prioritize frontier corrections that are associated with meaningful improvements in subsequent clinical content.

Table 6: Qualitative analysis of entity-level correction and its downstream effect on IU-Xray. Blue and orange denote correct and contradictory predictions of the focal entity, respectively. green and purple denote clinical findings that are consistent with or contradictory to the reference. The paired-continuation analysis compares suffixes generated from erroneous and corrected frontier claims under the same pre-error state.

CXR699_IM-2263		**Focal entity: enlarged cardiac silhouette**
Frontal	**Lateral**	**Ground Truth:** Stable **enlargement of the cardiac silhouette**, stable mediastinal contours. Increased interstitial markings in the central lungs and bases, right greater than left. XXXX opacity on the lateral view over the heart also present on the previous exam suggesting chronic subsegmental atelectasis or scarring. No definite pleural effusion seen.
		

SFT (ROUGE-L=0.190): The cardiomediastinal silhouette is **normal in size and contour**. No focal consolidation, pneumothorax or large pleural effusion. Negative for acute bone abnormality.

Fact-OPD (ROUGE-L=0.173): The cardiomediastinal silhouette is **within normal limits for size and contour**. The lungs show increased interstitial markings without evidence of focal airspace disease, pleural effusion, or pneumothorax. No acute bone abnormality.

Frontier-uniform (ROUGE-L=0.211): Heart size is mildly **enlarged**, stable mediastinal and hilar contours. No focal alveolar consolidation, no definite pleural effusion seen. No typical findings of pulmonary edema.

CEF-OPD (ROUGE-L=0.247): Heart size moderately **enlarged**, stable mediastinal and hilar contours. No focal alveolar consolidation, no definite pleural effusion is seen. Increased central and basilar interstitial markings, greater on the right. No pneumothorax.

	Erroneous branch	**Corrected branch**
Frontier claim	**Heart size is normal.**	**Heart size is enlarged.**
Downstream continuation	No interstitial markings or focal pulmonary opacity is identified. No definite pleural effusion is seen. No pneumothorax.	Increased basilar interstitial markings, greater on the right. No focal alveolar consolidation. No pleural effusion is seen.
Paired continuation from the same pre-error state		$\mathbf{C^- = 0.48}$, $\mathbf{C^+ = 0.25}$, $\mathbf{\Delta C = +0.23}$