TARING: TRAINING-ANCHORED REWEIGHTING FOR MASKED RADIOLOGY REPORT GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Radiology report generators can assign high probability to corpus-common content even when it is only weakly supported by the current examination. This ambiguity is particularly important for iterative masked generators, where token predictions depend jointly on image evidence and the evolving partial report. We introduce Training-Anchored Reweighting for Iterative Non-autoregressive Generation (TARING), a training-free decoding procedure that contrasts current-study prediction with a deterministic training-mean evidence reference under a matched generative state. At each unresolved position, the same frozen model evaluates the current study and the reference while holding the partial report, refinement time, report support, queried coordinates, and valid actions fixed. TARING reweights the current-study distribution according to the predictive contrast between the two contexts. A KL-regularized objective yields this update in closed form, providing a principled trade-off between reference-relative contrast and fidelity to the original current-study prediction.

On MIMIC-CXR-JPG, TARING increases three-seed CheXbert micro-F1 from 0.512 ± 0.006 without reference guidance to 0.534 ± 0.002, primarily through an increase in recall from 0.489 ± 0.006 to 0.525 ± 0.002. Alternative empirical references yield closely matched F1 values, suggesting that the observed effect is not specific to one exact reference construction. Several higher-order lexical-overlap metrics decrease modestly, indicating a trade-off between extracted clinical-label agreement and reference-text similarity. These results characterize the behavior of training-anchored reweighting for one frozen generator and dataset and do not establish clinical factuality or cross-institution generalization.

1 INTRODUCTION

Radiology report generation must select findings supported by an examination while maintaining coherent language. Memory-driven decoding (Chen et al., 2020), visual–language alignment (Wang et al., 2023b), and grounding findings to image locations (Bannur et al., 2024) strengthen the available representations. Content selection nevertheless remains ambiguous: a high score can reflect image evidence, common reporting patterns, or surrounding text. An explicit evidence reference allows this decision to account for both current-study compatibility and reference-relative predictive support.

Inference-time guidance uses contextual or expert predictions to adjust token selection (Shi et al., 2024; Liu et al., 2021; Zhang et al., 2026). Visual Contrastive Decoding is a related approach that contrasts predictions from original and distorted visual inputs (Leng et al., 2024). For masked report generation, the central design question is how to construct an evidence reference and use it consistently as the report evolves.

Masked generators predict unresolved positions from visible text on both sides (Ghazvininejad et al., 2019; Nie et al., 2025). Because the partial report evolves, we isolate evidence-context effects by holding the partial report, refinement time, query position, and valid actions fixed across the two branches. The procedure must also specify the report support and the positions assigned together at each reveal round.

We propose Training-Anchored Reweighting for Iterative Non-autoregressive Generation (TARING), combining training-evidence reference construction, current-study support prediction, and matched-state masked generation. Averaging the trained generator's evidence representations at corresponding

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.

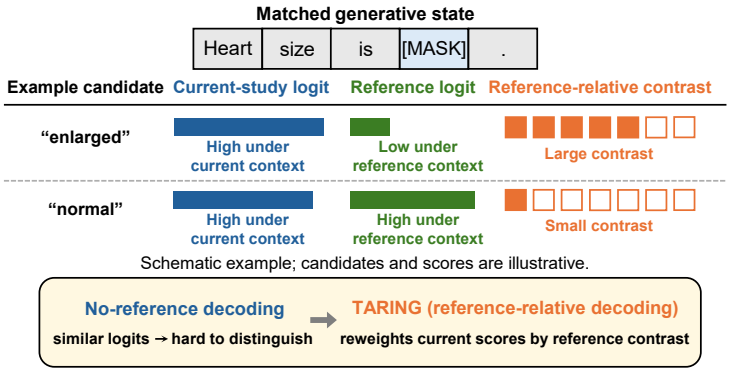


Figure 1: Schematic motivation for reference-relative decoding. Under the same matched generative state, candidate tokens can have similar current-study logits but different contrasts against a training-mean evidence reference. TARING combines the current-study score with this reference-relative contrast, distinguishing the candidates in the illustrated example. The candidates and scores are illustrative, not measured predictions.

query slots supplies a deterministic reference without retrieval or additional parameter learning. A current-study terminal prediction initializes the report support. The same frozen model then evaluates each queried position under the current context and training mean; their predictive ratio informs simultaneous token assignments. This continues until all content positions are filled. The reference prediction depends on visible text, so it is neither an unconditional prior nor clinically verified evidence.

We formulate content selection as expected predictive log-ratio maximization with a Kullback–Leibler penalty on departure from the current-study distribution. The resulting standard entropic mirror-ascent step (Beck & Teboulle, 2003) yields a closed-form score for every candidate. This derivation connects the reference construction to the implemented decision rule without attributing novelty to mirror ascent itself. Fixed-state analysis shows cancellation of shared report-state terms under an explicit decomposition, nondecreasing expected contrast and KL displacement with guidance strength, and an exact threshold at which relative advantage offsets a current-prediction margin. These results characterize the decoder's decisions, not their clinical correctness.

Our contributions are:

- **A reference-relative decoding formulation.** We specify training-mean evidence construction, current-only support prediction, and synchronous matched-state content generation within one procedure.
- **A principled objective and exact local analysis.** A KL-proximal derivation yields the content update and identifies its information budget and deterministic switching thresholds.
- **A controlled empirical evaluation.** On a frozen masked radiology generator, a three-seed paired comparison shows consistent gains in CheXbert recall and micro-F1. Reference-sensitivity, paired uncertainty, and post-hoc content analyses characterize the robustness and limitations of the decoding intervention.

2 Background

Masked report prediction. At refinement round j, a masked model with fixed parameters $\boldsymbol{\theta}$ predicts token a at unresolved coordinate i from the partial report $\mathbf{z}_j$, refinement time t_j, and evidence context $\mathbf{C}$. We write $h_{j,i} = (\mathbf{z}_j, t_j, i, \mathcal{A})$, where $\mathcal{A}$ is the finite set of valid content tokens and $p_{\boldsymbol{\theta}}(a \mid h_{j,i}, \mathbf{C})$ is normalized over $\mathcal{A}$. Because masked generators condition on visible tokens on both sides of an unresolved position (Ghazvininejad et al., 2019; Nie et al., 2025), the predictive state changes as the partial report evolves. A terminal boundary determines the content support, while a reveal schedule determines which unresolved coordinates are assigned at each round.

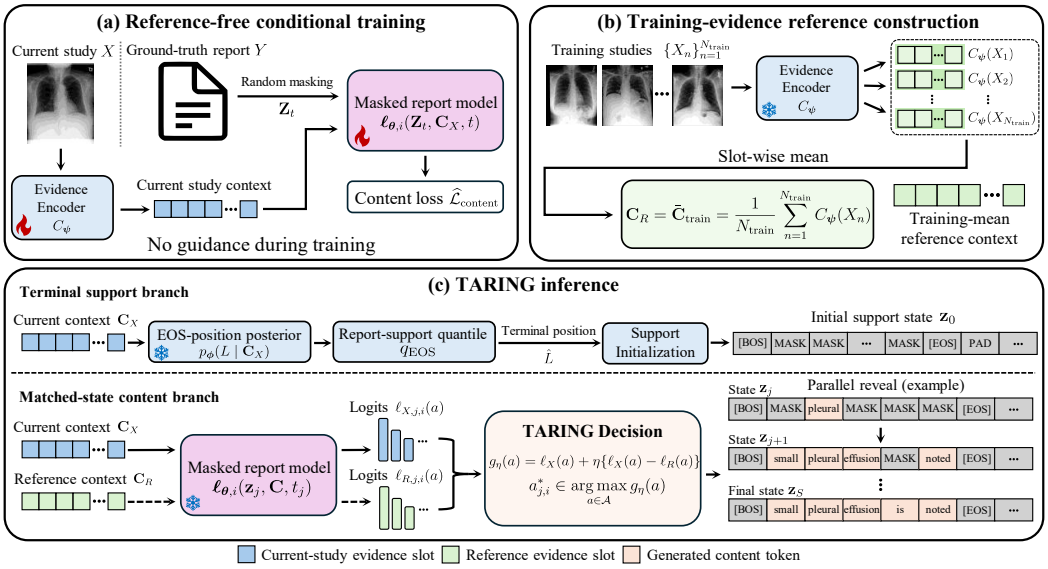


Figure 2: Overview of TARING. (a) Reference-free conditional training reconstructs masked report tokens from current-study evidence. (b) After training, the frozen evidence encoder produces training evidence contexts whose slot-wise mean defines the reference $\mathbf{C}_R$. (c) At inference, a current-study terminal head determines report support using a report-support quantile. For content generation, the same frozen model evaluates $\mathbf{C}_X$ and $\mathbf{C}_R$ at identical partial reports, refinement times, report supports, queried coordinates, and valid actions. The score $\mathbf{g}_\eta = \boldsymbol{\ell}_X + \eta(\boldsymbol{\ell}_X - \boldsymbol{\ell}_R)$, where $\eta = s - 1$, determines token selection by argmax. Queried positions are updated simultaneously, while terminal support remains unchanged.

Predictive contrast. At a fixed state $h_{j,i}$, let $\mathbf{p}_X$ and $\mathbf{p}_R$ denote the predictive distributions under the current-study and reference evidence contexts. Their tokenwise log ratio $r(a) = \log p_X(a) - \log p_R(a)$ measures relative predictive support across the two contexts. For finite branch logits $\boldsymbol{\ell}_X$ and $\boldsymbol{\ell}_R$, softmax normalization over the same action set $\mathcal{A}$ makes $r(a)$ equal to $\ell_X(a) - \ell_R(a)$ up to an action-independent constant. This relation underlies two-branch contrastive decoding (Shi et al., 2024; Li et al., 2023).

KL-regularized reweighting. For a candidate law $\boldsymbol{\pi} \in \Delta(\mathcal{A})$, where $\Delta(\mathcal{A})$ is the probability simplex, define $\mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X) = \sum_{a \in \mathcal{A}} \pi(a) \log[\pi(a)/p_X(a)]$, with terms satisfying $\pi(a) = 0$ defined to contribute zero. Maximizing $\mathbb{E}_{\boldsymbol{\pi}}[r] - \eta^{-1} \mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X)$ for step size $\eta > 0$ trades expected contrast against departure from $\mathbf{p}_X$, yielding the standard entropic mirror-ascent step (Beck & Teboulle, 2003). Deterministic decoding selects a mode of the resulting law.

Section 3 instantiates this construction with a training-mean evidence reference, current-only report support, and synchronous matched-state updates.

3 Problem and Method

TARING is a reference-relative procedure for generating a report from a radiographic study. It constructs a fixed training-evidence reference, predicts report support from the current study, and fills that support through synchronous masked updates. At each queried position, a KL-regularized objective favors current-to-reference predictive contrast while penalizing departure from the current-study distribution. Its closed-form solution supplies the content score used throughout decoding.

Setup and training-evidence reference. Let X be a radiographic study and $\mathbf{C}_X = C_{\boldsymbol{\psi}}(X) \in \mathbb{R}^{Q \times d_e}$ its evidence context, where Q is the number of learned query slots and d_e their dimension. The evidence encoder $C_{\boldsymbol{\psi}}$ and report model $p_{\boldsymbol{\theta}}$ have parameters $(\boldsymbol{\psi}, \boldsymbol{\theta})$, frozen after conditional

training. The shared ordered query bank aligns the slots across studies, so the N_{train} training contexts admit the slotwise mean

$$\mathbf{C}_R = \bar{\mathbf{C}}_{\text{train}} = \frac{1}{N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} C_{\boldsymbol{\psi}}(X_n). \tag{1}$$

The reference-construction step uses evidence tensors, not report text or clinical labels. The mean is their Euclidean barycenter: it minimizes summed squared Frobenius distance and avoids selecting or retrieving a donor. This representation-space property does not imply an optimal predictive reference or exact marginalization through the nonlinear report model.

Let H be the token horizon, S the number of reveal rounds, and $\mathbf{z}_j$ the length-H partial report before round $j \in \{0, \dots, S-1\}$, at time $t_j = j/S$. The nonempty finite set $\mathcal{A}$ contains the valid content tokens, excluding special and model-only symbols. At a queried, still-masked coordinate i, write $h_{j,i} = (\mathbf{z}_j, t_j, i, \mathcal{A})$ and define

$$p_X(a) = p_{\boldsymbol{\theta}}(a \mid h_{j,i}, \mathbf{C}_X), \qquad p_R(a) = p_{\boldsymbol{\theta}}(a \mid h_{j,i}, \mathbf{C}_R), \quad a \in \mathcal{A}. \tag{2}$$

Both laws are softmax distributions normalized over the same $\mathcal{A}$; we assume finite logits there, hence positive probabilities. Parameters, partial report, time, coordinate, report support, and valid actions are identical within this pair: only the evidence context changes. This matching applies to the two context evaluations within each decision; each subsequent decision uses the newly updated report. Because $\mathbf{p}_R$ receives the already revealed text in $\mathbf{z}_j$, it is a local context-replacement prediction, not an unconditional report prior.

Objective and closed-form content update. Suppress (j, i) for this fixed decision and define the predictive log ratio

$$r(a) = \log p_X(a) - \log p_R(a). \tag{3}$$

A large ratio alone can favor a token with low current-study probability. We therefore retain $\mathbf{p}_X$ as the base distribution and penalize displacement from it. For step size $\eta > 0$, let $\Delta(\mathcal{A}) = \{\boldsymbol{\pi} \geq 0 : \sum_{a \in \mathcal{A}} \pi(a) = 1\}$ and solve

$$\boldsymbol{\pi}_\eta = \arg\max_{\boldsymbol{\pi} \in \Delta(\mathcal{A})} \left\{ \mathbb{E}_{\boldsymbol{\pi}}[r] - \eta^{-1} \, \mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X) \right\}. \tag{4}$$

Here $\mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X) = \sum_{a \in \mathcal{A}} \pi(a) \log[\pi(a)/p_X(a)]$, with $0 \log 0 = 0$. This is one standard entropic mirror-ascent step for the fixed reward r (Beck & Teboulle, 2003; Kivinen & Warmuth, 1997). Strict concavity gives uniqueness. Lagrangian stationarity gives $\log \pi(a) = \log p_X(a) + \eta r(a) + \text{constant}$; normalization yields

$$\pi_\eta(a) = \frac{p_X(a) e^{\eta r(a)}}{\sum_b p_X(b) e^{\eta r(b)}} = \frac{p_X(a)^{1+\eta} p_R(a)^{-\eta}}{\sum_b p_X(b)^{1+\eta} p_R(b)^{-\eta}}. \tag{5}$$

All sums are over $\mathcal{A}$. The appendix establishes global optimality, including boundary distributions. Each new state supplies a new base $\mathbf{p}_X$ and reward r; the decoder does not iterate one fixed report-level objective.

Let $\ell_{X,j,i}(a) = \ell_{\boldsymbol{\theta},i}(\mathbf{z}_j, \mathbf{C}_X, t_j)(a)$ and $\ell_{R,j,i}(a) = \ell_{\boldsymbol{\theta},i}(\mathbf{z}_j, \mathbf{C}_R, t_j)(a)$ be the corresponding network logits. The log-ratio vector differs from $\boldsymbol{\ell}_X - \boldsymbol{\ell}_R$ only by an action-independent normalization constant. Substituting $\mathbf{p}_X \propto e^{\boldsymbol{\ell}_X}$ and $\mathbf{p}_R \propto e^{\boldsymbol{\ell}_R}$ into equation 5 therefore gives

$$g_\eta(a) = \ell_X(a) + \eta\{\ell_X(a) - \ell_R(a)\} = \ell_R(a) + s\{\ell_X(a) - \ell_R(a)\}, \quad s = 1 + \eta. \tag{6}$$

The implementation evaluates this affine score directly, without computing probability powers or numerically solving equation 4. Restoring (j, i), it selects

$$a^*_{j,i} \in \arg\max_{a \in \mathcal{A}} g_{\eta,j,i}(a) \tag{7}$$

with a fixed tie rule. The zero-step limit is $\boldsymbol{\pi}_0 = \mathbf{p}_X$; setting $s = 1$ ($\eta = 0$) leaves the current-study logits unchanged. The chosen action can change only when its relative advantage offsets its current-branch margin (Proposition 3); this is a predictive trade-off, not a correctness guarantee. The KL penalty regularizes the categorical law; it does not impose a tokenwise probability floor.

Decoding with current-only support. Let $p_{\boldsymbol{\phi}}(L \mid \mathbf{C}_X)$ be a separate frozen head over end-of-sequence positions $L \in \{1, \ldots, H-1\}$. BOS, MASK, EOS, and PAD denote sequence start, unresolved content, sequence end, and padding. For a report-support quantile $q_{\mathrm{EOS}} \in (0,1)$, the boundary and initial state are

$$\widehat{L} = \min\left\{ d : \sum_{l=1}^{d} p_{\boldsymbol{\phi}}(l \mid \mathbf{C}_X) \geq q_{\mathrm{EOS}} \right\}, \qquad \mathbf{z}_0 = (\texttt{BOS}, \texttt{MASK}^{\widehat{L}-1}, \texttt{EOS}, \texttt{PAD}^{H-\widehat{L}-1}). \quad (8)$$

Positions $1, \ldots, \widehat{L}-1$ alone are editable. The terminal head uses the current study to determine the support, while the reference enters content scoring only within that support. This factorization is an algorithmic choice, not a consequence of equation 4.

Let $\kappa : [0,1] \to [0,1]$ be a fixed nondecreasing reveal schedule with $\kappa(t_0) = 0$ and $\kappa(t_S) = 1$, and let $M_j = \{i : 1 \leq i < \widehat{L},\ z_{j,i} = \texttt{MASK}\}$. When $\kappa(t_j) < 1$, independent uniforms $U_{j,i} \sim \mathrm{Unif}(0,1)$ select

$$\rho_j = \frac{\kappa(t_{j+1}) - \kappa(t_j)}{1 - \kappa(t_j)}, \qquad J_j = \{i \in M_j : U_{j,i} < \rho_j\}.$$

If $\kappa(t_j) = 1$, we set $J_j = \varnothing$. All surviving positions are queried in the final round. Each nonempty J_j is scored from one pre-update $\mathbf{z}_j$ under both evidence contexts, and all selected tokens are assigned simultaneously. Empty rounds leave $\mathbf{z}_j$ unchanged, and revealed tokens are not remasked. The resulting procedure has one terminal boundary, no active MASK after the final round, and only PAD after EOS. Algorithm 1 summarizes the procedure.

Algorithm 1 TARING: reference-relative masked decoding

Require: study X; fixed reference $\mathbf{C}_R$; guidance scale s; report-support quantile q_{EOS}; reveal schedule; valid actions $\mathcal{A}$

1: $\mathbf{C}_X \leftarrow C_{\boldsymbol{\psi}}(X)$; choose $\widehat{L}$ from $p_{\boldsymbol{\phi}}(L \mid \mathbf{C}_X)$ using q_{EOS}; initialize $\mathbf{z}_0$
2: **for** $j = 0, \ldots, S-1$ **do**
3: form queried set J_j from still-masked coordinates; query all survivors at $j = S-1$
4: $\mathbf{z}_{j+1} \leftarrow \mathbf{z}_j$
5: **if** $J_j \neq \varnothing$ **then**
6: evaluate $(\boldsymbol{\ell}_X, \boldsymbol{\ell}_R)$ in one batch on the same $(\mathbf{z}_j, t_j, J_j)$
7: **for all** $i \in J_j$ **in parallel do**
8: $z_{j+1,i} \leftarrow \arg\max_{a \in \mathcal{A}} \left[\ell_{R,j,i}(a) + s\{\ell_{X,j,i}(a) - \ell_{R,j,i}(a)\}\right]$
9: **return** $(\mathbf{z}_S, \widehat{L})$

Implementation and complexity. The evidence contexts $\mathbf{C}_X$ and $\mathbf{C}_R$ are reused throughout decoding, while both current and reference logits are recomputed as the partial report changes. Current and reference requests are concatenated in one batch using shared weights. This avoids a second parameter copy, but not the computation of the second logical branch. Model dimensions, decoding hyperparameters, and operating-point selection are described in Section 4 and the appendix.

Given encoded training contexts, forming the mean requires $O(N_{\mathrm{train}} Q d_e)$ additions and $O(Q d_e)$ retained reference storage. At inference, each nonempty round uses two logical branch evaluations, with at most S batched content-model calls per study. Since every content coordinate is assigned once, score combination and mode selection require $O((\widehat{L}-1)|\mathcal{A}|)$ arithmetic in total. Peak activation memory and latency depend on batching.

4 Experimental Protocol

Data and model. We use MIMIC-CXR and MIMIC-CXR-JPG (Johnson et al., 2019b;a). The processed protocol contains 172,608 training studies, 1,391 validation studies, and 2,357 test studies. The frozen report generator combines RAD-DINO image features (Pérez-García et al., 2025), a two-layer 128-query evidence resampler, and an 8B LLaDA backbone (Nie et al., 2025) with rank-16 LoRA adapters. TARING does not retrain this checkpoint. The reference context is the slotwise mean of the frozen evidence representations computed over the training split.

Implementation details. The canonical configuration uses $q_{\mathrm{EOS}} = 0.60$, guidance scale $s = 1.75$, $S = 80$ reveal rounds, and the cumulative reveal schedule $\kappa(t) = \sin^2(\pi t/2)$, with no plausibility threshold. Current- and reference-context predictions are evaluated with the same frozen model and are concatenated into a shared batch when queried at the same refinement state. Token selection uses deterministic argmax with a fixed tie rule, while stochasticity enters only through the reveal process. The test evaluation uses reveal seeds $\{0, 1, 2\}$.

Operating-point selection. The report-support quantile q_{EOS} and guidance scale s are selected exclusively on the validation split using CheXbert micro-F1 (Smit et al., 2020). The initial sweep evaluates $q_{\mathrm{EOS}} \in \{0.55, 0.58, 0.60, 0.90\}$ and $s \in \{1.00, 1.25, 1.50, 1.75, 2.00\}$ under reveal seed 0. The two strongest guided configurations at $q_{\mathrm{EOS}} = 0.60$, corresponding to $s = 1.50$ and $s = 1.75$, are then evaluated under reveal seeds 1 and 2. Their three-seed validation micro-F1 values are 0.5469 ± 0.0082 and 0.5474 ± 0.0040, respectively. We therefore fix $(q_{\mathrm{EOS}}, s) = (0.60, 1.75)$ before canonical test evaluation. This final selection uses the validation sweep described above; earlier test inspection during project development is disclosed in Section 7. The complete seed-0 validation grid is reported in Appendix D.1 (Table 4).

Test evaluation. The frozen canonical configuration is evaluated on all 2,357 test studies under reveal seeds $\{0, 1, 2\}$. We report BLEU-1 through BLEU-4 (Papineni et al., 2002), METEOR (Banerjee & Lavie, 2005), and ROUGE-L (Lin, 2004) for lexical agreement. Clinical-label agreement is measured using CheXbert micro-precision, micro-recall, micro-F1, and macro-F1 (Smit et al., 2020). Unless otherwise stated, results are reported as the mean and sample standard deviation across the three reveal seeds. These seeds quantify variation in the reveal process for one frozen checkpoint and do not constitute independent training runs.

Controlled comparison and reference sensitivity. The primary controlled comparison removes reference guidance by setting $s = 1$, for which equation 6 reduces to the current-study logits. The no-reference and TARING conditions otherwise use the same frozen checkpoint, report-support mechanism, reveal schedule, token-action set, and evaluation pipeline. To examine sensitivity to the reference construction, we additionally evaluate a validation-set mean and the validation exemplar nearest to that mean while holding q_{EOS}, s, and the reveal seeds fixed. These validation-derived references are post-hoc sensitivity controls only and are not used by the canonical TARING configuration.

5 RESULTS

Comparison with prior methods. Table 1 places TARING in the context of representative autoregressive and non-autoregressive radiology report generators. Among the listed Non-AR methods, TARING reports the highest source-reported BLEU-1, BLEU-2, ROUGE-L, and CheXbert precision, recall, and micro-F1 values. Its BLEU-3 and BLEU-4 scores remain below ControlDiff (Tian et al., 2024), while its METEOR score remains below both ControlDiff and MeDiM (Mao et al., 2025). These source-reported results provide context across published systems rather than a controlled head-to-head ranking, since the methods differ in architecture, training procedure, input information, and evaluation conditions.

Effect of training-anchored reweighting. Table 2 isolates the effect of TARING within the same frozen report generator. TARING increases CheXbert micro-F1 from 0.512 ± 0.006 to 0.534 ± 0.002 and macro-F1 from 0.309 ± 0.006 to 0.323 ± 0.001. The change is primarily recall-driven: recall increases from 0.489 ± 0.006 to 0.525 ± 0.002, while precision also increases from 0.538 ± 0.005 to 0.544 ± 0.003.

The gain in clinical-label agreement does not coincide with uniformly higher lexical overlap. BLEU-1 remains nearly unchanged (0.430 ± 0.001 versus 0.431 ± 0.001), whereas BLEU-4 decreases from 0.122 ± 0.001 to 0.117 ± 0.001 and ROUGE-L from 0.314 ± 0.002 to 0.308 ± 0.001. Thus, the selected TARING configuration improves extracted clinical-label agreement while modestly reducing higher-order reference-text overlap. The CheXbert gain does not extend to RadGraph report–reference agreement: all three variants decrease under the canonical configuration (Table 9).

Method	NLG Metrics ↑						CE Metrics ↑		
	B-1	B-2	B-3	B-4	MTR	R-L	P	R	F1
Autoregressive (AR)									
R2Gen Chen et al. (2020)	0.353	0.218	0.145	0.103	0.142	0.277	0.333	0.273	0.276
R2GenCMN Chen et al. (2021)	0.353	0.218	0.148	0.106	0.142	0.278	0.334	0.275	0.278
KiUT Huang et al. (2023)	0.393	0.243	0.159	0.113	0.160	0.285	0.371	0.318	0.321
METransformer Wang et al. (2023a)	0.386	0.250	0.169	0.124	0.152	0.291	0.364	0.309	0.311
RGRG Tanida et al. (2023)	0.373	0.249	0.175	0.126	0.168	0.264	0.461	0.475	0.447
R2GenGPT Wang et al. (2023b)	0.411	0.267	0.186	0.134	0.160	0.297	0.392	0.387	0.389
MAN Shen et al. (2024)	0.396	0.244	0.162	0.115	0.151	0.274	0.411	0.398	0.389
DCG Liang et al. (2024)	0.397	0.258	0.166	0.126	0.162	0.295	0.441	0.414	0.404
EKAGen Bu et al. (2024)	0.419	0.258	0.170	0.119	0.157	0.287	0.517	0.483	0.499
PromptMRG Jin et al. (2024a)	0.398	–	–	0.112	0.157	0.268	0.501	0.509	0.476
R2-LLM Liu et al. (2024)	0.402	0.262	0.180	0.128	0.175	0.291	0.465	0.482	0.473
DART Park et al. (2025)	**0.437**	0.279	0.191	0.137	0.175	0.310	0.520	0.546	0.533
MPO Xiao et al. (2025)	0.416	0.269	0.191	0.139	0.162	0.309	0.436	0.376	0.353
MLRG‡ Liu et al. (2025)	0.411	0.277	0.204	0.158	0.176	0.320	0.549	0.468	0.505
LLM-RG4 Wang et al. (2025b)	0.377	–	–	0.144	–	0.318	**0.583**	**0.593**	**0.588**
PriorRG‡ Liu et al. (2026)	0.412	**0.290**	**0.220**	**0.175**	**0.189**	**0.324**	0.541	0.485	0.511
Non-autoregressive (Non-AR)									
D^2-Net Jin et al. (2024b)	0.365	0.230	0.153	0.107	0.136	0.278	–	–	–
DMR^2G Ouyang et al. (2025)	0.373	0.217	0.136	0.099	–	0.267	–	–	–
ControlDiff Tian et al. (2024)	0.411	0.265	**0.183**	**0.132**	0.186	0.299	0.477	0.484	0.480
MeDiM Mao et al. (2025)	0.328	0.185	0.109	–	**0.265**	0.297	–	–	–
TARING (Ours)	**0.431**	**0.274**	0.178	0.117	0.170	**0.308**	**0.544**	**0.525**	**0.534**

Table 1: Comparison with representative radiology report generation methods on MIMIC-CXR. AR and Non-AR denote autoregressive and non-autoregressive generation, respectively. B-1 through B-4 denote BLEU-1 through BLEU-4, MTR denotes METEOR, and R-L denotes ROUGE-L. CE denotes CheXbert-based clinical efficacy; P, R, and F1 denote micro-precision, micro-recall, and micro-F1, respectively. Boldface and underlining indicate the best and second-best source-reported values within each generation family. A dash indicates that the cited paper does not report the corresponding metric. ‡ denotes a method that uses patient-specific longitudinal information and/or additional clinical context.

Method	P	R	Micro-F1	Macro-F1
No reference	0.538±.005	0.489±.006	0.512±.006	0.309±.006
TARING	**0.544±.003**	**0.525±.002**	**0.534±.002**	**0.323±.001**
Δ	+.006	+.036	+.022	+.014

Table 2: Controlled comparison between no-reference decoding and TARING on the MIMIC-CXR-JPG test set. TARING uses the training-mean reference with $q_{\mathrm{EOS}} = 0.60$ and $s = 1.75$. Values are mean ± sample standard deviation over reveal seeds $\{0, 1, 2\}$.

Reference sensitivity. Table 3 examines whether the observed effect depends strongly on the exact reference construction. The training mean, validation mean, and mean-nearest validation exemplar yield closely matched CheXbert micro-F1 values of 0.5343, 0.5334, and 0.5331, respectively. All three guided conditions increase recall relative to no-reference decoding, although their precision–recall profiles differ slightly.

We use the training mean as the canonical reference because it is constructed exclusively from training evidence. The similarity among the three guided conditions nevertheless indicates that the observed effect is not specific to one exact empirical reference, nor do these results establish the training mean as uniquely optimal.

Reference	P	R	Micro-F1	Macro-F1
None	0.538±.005	0.489±.006	0.512±.006	0.309±.006
Training mean	0.544±.003	**0.525±.002**	**0.534±.002**	**0.323±.001**
Validation mean	0.545±.003	0.523±.003	0.533±.003	0.322±.004
Nearest val. exemplar	**0.546±.004**	0.520±.004	0.533±.003	0.313±.006

Table 3: Sensitivity to reference construction on the test set. All guided conditions use $q_{\mathrm{EOS}} = 0.60$ and $s = 1.75$. Values are mean ± sample standard deviation over reveal seeds $\{0, 1, 2\}$. Validation-derived references are sensitivity controls only and are not used by the canonical TARING configuration.

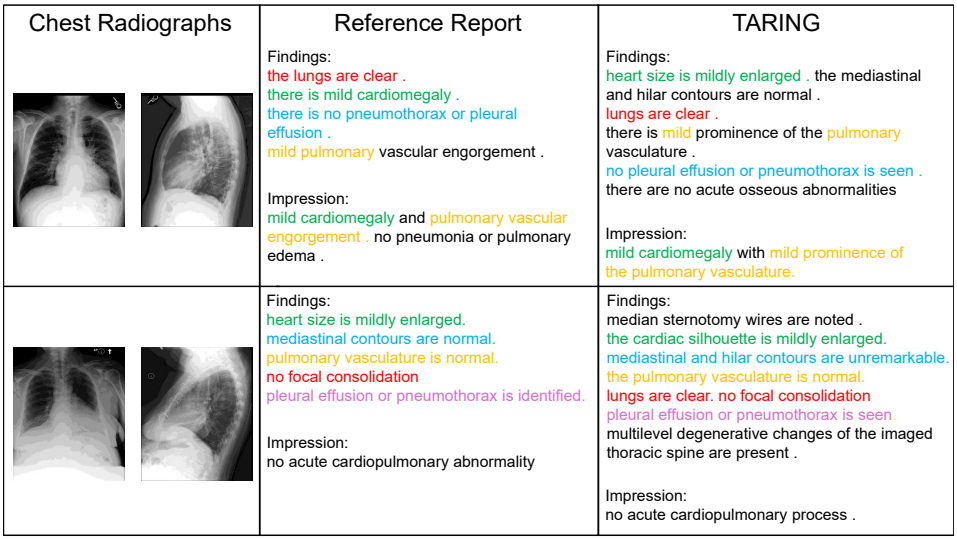


Figure 3: Qualitative examples of TARING-generated reports. Each row shows chest radiographs (left), the reference report (middle), and the TARING output (right). Matching colors within each example highlight corresponding report content, including paraphrases. The examples illustrate textual correspondence and differences; they do not establish the image-grounded correctness of every generated statement.

6 Ablation Study

Reference-guidance ablation and pairing. The no-reference condition sets $s = 1$, for which the TARING score reduces to the current-study logits and the reference branch has no effect on token selection. The canonical TARING condition instead uses the training-mean reference with $s = 1.75$. Both conditions share the frozen checkpoint, study inputs, report-support mechanism, reveal schedule, queried-coordinate construction, valid action set, token-selection rule, numerical precision, batching policy, evaluation pipeline and reveal seeds. Once their token choices diverge, however, the resulting partial reports also diverge; matched-state evaluation applies within each current/reference decision pair, not across the subsequent trajectories of the two decoding conditions.

Contribution of reference guidance. The paired micro-F1 gain is 0.0223 (95% subject-cluster bootstrap interval $[0.0165, 0.0281]$; Table 5), and all three reveal seeds show positive effects. The gain is primarily recall-driven; the smaller precision change has an interval that includes zero. These improvements coexist with lower higher-order lexical overlap and RadGraph agreement.

Reference construction. The reference ablation in Table 3 shows that replacing the training mean with the validation mean or the mean-nearest validation exemplar changes CheXbert micro-F1 by less than 0.002 in absolute value. This weak sensitivity suggests that the observed improvement is not tied to one exact empirical reference. We nevertheless use the training mean in the canonical procedure because it is deterministic and constructed exclusively from training evidence.

Guidance strength and report support. Validation performance varies non-monotonically with guidance strength and report support; the complete selection grid is reported in Appendix D.1.

7 Discussion and Limitations

TARING's recall-driven CheXbert gains are accompanied by lower higher-order lexical overlap and RadGraph agreement. A post-hoc seed-0 GREEN analysis shows fewer missing-finding errors but more false-finding and severity errors, alongside a lower GREEN score (Table 11). This pattern is consistent with an omission–commission trade-off rather than uniform report-quality improvement. Automated report agreement does not establish image-grounded clinical factuality, and the fixed-state theory characterizes predictive behavior rather than token-level correctness or complete-report factuality.

The training mean is a deterministic, training-only reference requiring neither retrieval nor an additional model, but similar results with alternative empirical references do not establish its unique optimality. The reported comparisons do not include random, null, or corrupted reference contexts, so they do not establish superiority over these alternatives. It may encode dataset prevalence, acquisition conditions, projection, scanner characteristics, institutional practice, and reporting regularities; it should not be treated as a pathology-free or clinically neutral reference.

Evaluation is limited to one frozen checkpoint and one institution-derived dataset; the three reveal seeds measure decoding variability, not independent training or external replication. The canonical operating point was selected using validation data only, but earlier project development inspected test behavior under other configurations. The test set is therefore not an untouched project-level holdout. External-institution evaluation, independent checkpoint replication, and direct radiologist assessment are needed before claiming clinical utility or broad generalization.

8 Conclusion

We presented TARING, a training-free decoding procedure for iterative masked radiology report generation that contrasts current-study prediction with a fixed training-derived evidence reference under a matched generative state. A KL-regularized objective yields the corresponding closed-form reweighting rule without modifying or retraining the frozen report generator.

On MIMIC-CXR-JPG, TARING raises three-seed CheXbert micro-F1 from 0.512 to 0.534, primarily through recall, while higher-order lexical overlap and RadGraph agreement decrease. These findings are limited to one frozen generator and dataset; clinical validation, external generalization, and independent replication remain open.

Reproducibility Statement

Section 3 and Algorithm 1 specify the decoding procedure, with derivations and fixed-state proofs provided in the appendix. Section 4 and Appendix C describe the experimental protocol, model configuration, and training and reveal processes. Appendix D reports complete metrics, seed-specific results, paired subject-cluster bootstrap intervals, and matched-state and GREEN analyses. The qualitative comparison appears in Appendix E, and limits of end-to-end reconstruction are discussed in Appendix F.

Ethics Statement

All authors have read and commit to adhering to the ICLR Code of Ethics. This study uses the de-identified MIMIC-CXR dataset and involves no new patient recruitment or clinical intervention. Generated reports are research outputs and are not intended for clinical deployment. Automated agreement with reference reports does not establish clinical correctness, and generated statements may contain unsupported findings or reflect reporting biases.

AI-USE STATEMENT

In this work, we used generative AI tools, including OpenAI ChatGPT and Codex, to assist with mathematical exposition, drafting code and experiment specifications, organizing results, and automated audits. Additionally, we used these tools for manuscript planning, drafting, editing, and polishing.

The authors retained control of experimental decisions and reviewed the AI-assisted work. In particular, the authors independently reviewed source papers and research artifacts, checked mathematical steps and generated code, and reviewed and revised AI-assisted text. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

REFERENCES

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*, 2005.

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

Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

Shenshen Bu, Taiji Li, Yuedong Yang, and Zhiming Dai. Instance-level expert knowledge and aggregate discriminative attention for radiology report generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14194–14204, 2024.

Lifeng Chen, Tianqi You, Hao Liu, Zhimin Bao, Jile Jiao, Xiao Han, Zhicai Ou, Tao Sun, Xiaofeng Mou, Xiaojie Jin, and Yi Xu. ECHO: Efficient chest X-ray report generation with one-step block diffusion. *arXiv preprint arXiv:2604.09450*, 2026.

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

Zhihong Chen, Yaling Shen, Yan Song, and Xiang Wan. Cross-modal memory networks for radiology report generation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2021.

Alessandro Favero, Luca Zancato, Matthew Trager, Siddharth Choudhary, Pramuditha Perera, Alessandro Achille, Ashwin Swaminathan, and Stefano Soatto. Multi-modal hallucination control by visual information grounding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Marjan Ghazvininejad, Omer Levy, Yinhan Liu, and Luke Zettlemoyer. Mask-Predict: Parallel decoding of conditional masked language models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2019.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Zhongzhen Huang, Xiaofan Zhang, and Shaoting Zhang. KiUT: Knowledge-injected U-transformer for radiology report generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19809–19818, 2023.

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, 2024a.

Yuda Jin, Weidong Chen, Yuanhe Tian, Yan Song, Chenggang Yan, and Zhendong Mao. Improving radiology report generation with D^2-Net: When diffusion meets discriminator. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024b.

Alistair Johnson, Matt Lungren, Yifan Peng, Zhiyong Lu, Roger Mark, Seth Berkowitz, and Steven Horng. MIMIC-CXR-JPG - chest radiographs with structured labels. *PhysioNet*, 2019a. Version 2.0.0.

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

Jyrki Kivinen and Manfred K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. Mitigating object hallucinations in large vision-language models through visual contrastive decoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Pengxiang Li, Shilin Yan, Jiayin Cai, Renrui Zhang, Ruichuan An, Ziyu Guo, and Xiaowei Gao. Adaptive classifier-free guidance via dynamic low-confidence masking. In *Advances in Neural Information Processing Systems*, 2025.

Xiang Lisa Li, Ari Holtzman, Daniel Fried, Percy Liang, Jason Eisner, Tatsunori Hashimoto, Luke Zettlemoyer, and Mike Lewis. Contrastive decoding: Open-ended text generation as optimization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2023.

Xiao Liang, Yanlei Zhang, Di Wang, Haodi Zhong, Ronghan Li, and Quan Wang. Divide and conquer: Isolating normal-abnormal attributes in knowledge graph-enhanced radiology report generation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4967–4975, 2024.

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

Alisa Liu, Maarten Sap, Ximing Lu, Swabha Swayamdipta, Chandra Bhagavatula, Noah A. Smith, and Yejin Choi. DExperts: Decoding-time controlled text generation with experts and anti-experts. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2021.

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, 2024.

Kang Liu, Zhuoqi Ma, Xiaolu Kang, Yunan Li, Kun Xie, Zhicheng Jiao, and Qiguang Miao. Enhanced contrastive learning with multi-view longitudinal data for chest X-ray report generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10348–10359, 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 7206–7214, 2026.

Jiawei Mao, Yuhan Wang, Lifeng Chen, Can Zhao, Yucheng Tang, Dong Yang, Liangqiong Qu, Daguang Xu, and Yuyin Zhou. Discrete diffusion models with MLLMs for unified medical multimodal generation, 2025.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large language diffusion models. In *Advances in Neural Information Processing Systems*, 2025.

Sophie Ostmeier, Justin Xu, Zhihong Chen, Maya Varma, Louis Blankemeier, Christian Bluethgen, Arne Edward Michalson, Michael Moseley, Curtis Langlotz, Akshay S. Chaudhari, and Jean-Benoit Delbrouck. GREEN: Generative radiology report evaluation and error notation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 374–390. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-emnlp.21. URL https://aclanthology.org/2024.findings-emnlp.21/.

Huan Ouyang, Zheng Chang, Binghao Tang, and Si Li. DMR2G: Diffusion model for radiology report generation. *Multimedia Tools and Applications*, 84(23):27029–27043, 2025. doi: 10.1007/s11042-024-20206-x.

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

Jonggwon Park, Byungmu Yoon, Soobum Kim, and Kyoyun Choi. RA-RRG: Multimodal retrieval-augmented radiology report generation with key phrase extraction. In *Findings of the Association for Computational Linguistics: ACL*, 2026.

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

Fernando Pérez-García, Harshita Sharma, Sam Bond-Taylor, Kenza Bouzid, Valentina Salvatelli, Maximilian Ilse, Shruthi Bannur, Daniel C. Castro, Anton Schwaighofer, Matthew P. Lungren, Maria Teodora Wetscherek, Noel Codella, Stephanie L. Hyland, Javier Alvarez-Valle, and Ozan Oktay. Exploring scalable medical image encoders beyond text supervision. *Nature Machine Intelligence*, 7:119–130, 2025.

Kevin Rojas, Ye He, Chieh-Hsin Lai, Yuhta Takida, Yuki Mitsufuji, and Molei Tao. Improving classifier-free guidance in masked diffusion: Low-dim theoretical insights with high-dim impact. In *International Conference on Learning Representations*, 2026.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems*, 2024.

Yair Schiff, Subham Sekhar Sahoo, Hao Phung, Guanghan Wang, Sam Boshar, Hugo Dalla-Torre, Bernardo P. de Almeida, Alexander Rush, Thomas Pierrot, and Volodymyr Kuleshov. Simple guidance mechanisms for discrete diffusion models. In *International Conference on Learning Representations*, 2025.

Hongyu Shen, Mingtao Pei, Juncai Liu, and Zhaoxing Tian. Automatic radiology reports generation via memory alignment network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4776–4783, 2024.

Weijia Shi, Xiaochuang Han, Mike Lewis, Yulia Tsvetkov, Luke Zettlemoyer, and Wen-tau Yih. Trusting your evidence: Hallucinate less with context-aware decoding. In *Proceedings of the NAACL Conference on Human Language Technologies*, 2024.

Akshay Smit, Saahil Jain, Pranav Rajpurkar, Anuj Pareek, Andrew Ng, and Matthew Lungren. Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2020.

Shantam Srivastava, Mahesh Bhosale, David Doermann, and Mingchen Gao. CWCD: Category-wise contrastive decoding for structured medical report generation. In *Proceedings of the International Conference on Medical Imaging with Deep Learning*, 2026.

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

Yuanhe Tian, Fei Xia, and Yan Song. Diffusion networks with task-specific noise control for radiology report generation. In *Proceedings of the ACM International Conference on Multimedia*, 2024.

Han Wang, Archiki Prasad, Elias Stengel-Eskin, and Mohit Bansal. AdaCAD: Adaptively decoding to balance conflicts between contextual and parametric knowledge. In *Proceedings of the NAACL Conference on Human Language Technologies*, 2025a.

Zhanyu Wang, Lingqiao Liu, Lei Wang, and Luping Zhou. METransformer: Radiology report generation by transformer with multiple learnable expert tokens. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11558–11567, 2023a.

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

Zhuhao Wang, Yihua Sun, Zihan Li, Xuan Yang, Fang Chen, and Hongen Liao. LLM-RG4: Flexible and factual radiology report generation across diverse input contexts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 8250–8258, 2025b.

Ting Xiao, Lei Shi, Peng Liu, Zhe Wang, and Chenjia Bai. Radiology report generation via multi-objective preference optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 8664–8672, 2025.

Shiying Yu, Jielei Wang, and Guoming Lu. AnchorDiff: Topology-aware masked diffusion with confidence-based rewriting for radiology report generation. *arXiv preprint arXiv:2605.17071*, 2026.

Xi Zhang, Zaiqiao Meng, Jake Lever, and Edmond S. L. Ho. CCD: Mitigating hallucinations in radiology MLLMs via clinical contrastive decoding. In *Findings of the Association for Computational Linguistics: ACL*, 2026.

A Complete Derivations and Proofs

A.1 Fixed-state setup and support conventions

Bold symbols denote complete vectors, token arrays, and evidence tensors; scalar components and scalar-valued functions remain in ordinary type. Vector logarithms, exponentials, powers, and scalar offsets act componentwise.

Fix $h_{j,i} = (\mathbf{z}_j, t_j, i)$, an active coordinate $i \in J_j$, the terminal boundary, the masks, the finite nonempty action set $\mathcal{A}$, and frozen parameters $(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\phi})$. The distributions $\mathbf{p}_X, \mathbf{p}_R$ in Eq. (2) vary only with $\mathbf{C}_X$ versus $\mathbf{C}_R$. All logarithms are natural. We assume both distributions are strictly positive on $\mathcal{A}$, as follows from finite real logits on this common support, and define

$$\Delta(\mathcal{A}) = \left\{ \boldsymbol{\pi} \in \mathbb{R}^{|\mathcal{A}|} : \pi(a) \geq 0,\ \sum_{a \in \mathcal{A}} \pi(a) = 1 \right\}. \tag{9}$$

For $q(a) > 0$, the convention $0 \log(0/q(a)) = 0$ is used throughout. The reward $r(a) = \log p_X(a) - \log p_R(a)$ is finite and remains fixed as $\boldsymbol{\pi}$ varies. This is a local optimization over one categorical law, not optimization over reports, contexts, network weights, or reference tensors.

A.2 Existence, strict concavity, and boundary exclusion

For finite $\eta > 0$, write

$$\mathcal{J}_\eta(\boldsymbol{\pi}) = \sum_a \pi(a) r(a) - \eta^{-1} \sum_a \pi(a) \log \pi(a) + \eta^{-1} \sum_a \pi(a) \log p_X(a). \tag{10}$$

The extended function $x \log x$ is continuous and strictly convex on $[0, \infty)$. Consequently $\mathcal{J}_\eta$ is continuous on the compact simplex and has a maximizer. The two outer terms in Eq. (10) are linear, while the middle term is strictly concave, so that maximizer is unique. On the relative interior, the same conclusion follows from the tangent-space Hessian: for any nonzero $\mathbf{v}$ satisfying $\sum_a v(a) = 0$,

$$\mathbf{v}^\top \nabla^2 \mathcal{J}_\eta(\boldsymbol{\pi}) \mathbf{v} = -\frac{1}{\eta} \sum_{a \in \mathcal{A}} \frac{v(a)^2}{\pi(a)} < 0. \tag{11}$$

If $|\mathcal{A}| = 1$, the simplex has a single point and uniqueness is immediate.

For completeness, a boundary point cannot maximize the objective when $|\mathcal{A}| > 1$. Suppose $\pi(a) = 0$ and $\pi(b) > 0$, and transfer mass $\epsilon \in (0, \pi(b))$ from b to a. The reward and the already positive b coordinate change by $O(\epsilon)$, whereas the new entropy contribution is $-\eta^{-1} \epsilon \log \epsilon$. Hence

$$\mathcal{J}_\eta(\boldsymbol{\pi} + \epsilon \mathbf{e}_a - \epsilon \mathbf{e}_b) - \mathcal{J}_\eta(\boldsymbol{\pi}) = -\frac{\epsilon \log \epsilon}{\eta} + O(\epsilon) > 0 \tag{12}$$

for sufficiently small ϵ. Interior stationarity therefore applies. Here $\mathbf{e}_a$ and $\mathbf{e}_b$ are coordinate unit vectors.

A.3 Lagrangian solution and normalization

The equality-constraint Lagrangian is

$$\mathcal{L}(\boldsymbol{\pi}, \lambda) = \mathcal{J}_\eta(\boldsymbol{\pi}) + \lambda \left(\sum_a \pi(a) - 1 \right). \tag{13}$$

For each a, differentiate with respect to $\pi(a)$:

$$0 = r(a) - \eta^{-1} \left(\log \frac{\pi(a)}{p_X(a)} + 1 \right) + \lambda, \tag{14}$$

$$\log \frac{\pi(a)}{p_X(a)} = \eta r(a) + \eta \lambda - 1, \tag{15}$$

$$\pi(a) = p_X(a) \exp\{\eta r(a)\} \exp\{\eta \lambda - 1\}. \tag{16}$$

The last factor is shared by every action. Let

$$Z_\eta = \sum_a p_X(a) \exp\{\eta r(a)\} \in (0, \infty). \tag{17}$$

Normalization then requires $\exp\{\eta\lambda - 1\} = Z_\eta^{-1}$, or $\lambda = (1 - \log Z_\eta)/\eta$. Thus

$$\pi_\eta(a) = p_X(a) \exp\{\eta r(a) - \log Z_\eta\}. \tag{18}$$

The resulting distribution has positive entries summing to one. Substituting $r(a) = \log p_X(a) - \log p_R(a)$ yields

$$p_X(a)e^{\eta r(a)} = p_X(a)e^{\eta \log p_X(a) - \eta \log p_R(a)} \tag{19}$$

$$= p_X(a)^{1+\eta} p_R(a)^{-\eta}. \tag{20}$$

This proves Eq. (5).

A.4 An exact global optimality certificate

An independent identity certifies the candidate against every feasible law, including those on the boundary. From $\log \boldsymbol{\pi}_\eta = \log \mathbf{p}_X + \eta\mathbf{r} - \log Z_\eta$,

$$\mathrm{KL}(\boldsymbol{\pi} \| \boldsymbol{\pi}_\eta) = \mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X) - \eta \sum_a \pi(a) r(a) + \log Z_\eta, \tag{21}$$

$$\mathcal{J}_\eta(\boldsymbol{\pi}) = \frac{\log Z_\eta}{\eta} - \frac{1}{\eta} \mathrm{KL}(\boldsymbol{\pi} \| \boldsymbol{\pi}_\eta). \tag{22}$$

KL is nonnegative, with equality precisely when its arguments agree. Consequently

$$\boxed{\mathcal{J}_\eta(\boldsymbol{\pi}_\eta) - \mathcal{J}_\eta(\boldsymbol{\pi}) = \eta^{-1} \mathrm{KL}(\boldsymbol{\pi} \| \boldsymbol{\pi}_\eta) \geq 0,} \tag{23}$$

with equality only for $\boldsymbol{\pi} = \boldsymbol{\pi}_\eta$. This is a standard exponential-tilt optimality certificate for a fixed linear reward.

A.5 Why this is one entropic mirror-ascent step

Let $\Omega(\boldsymbol{\pi}) = \sum_a \pi(a) \log \pi(a)$ be negative entropy. For a strictly positive base law $\mathbf{p}_X$, its Bregman divergence is

$$D_\Omega(\boldsymbol{\pi}, \mathbf{p}_X) = \Omega(\boldsymbol{\pi}) - \Omega(\mathbf{p}_X) - \langle \nabla\Omega(\mathbf{p}_X), \boldsymbol{\pi} - \mathbf{p}_X \rangle \tag{24}$$

$$= \sum_a \pi(a) \log \frac{\pi(a)}{p_X(a)} = \mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X), \tag{25}$$

because $\nabla\Omega(\mathbf{p}_X)_a = 1 + \log p_X(a)$ and both laws have unit mass. For the fixed linear reward $f(\boldsymbol{\pi}) = \langle \boldsymbol{\pi}, \mathbf{r} \rangle$, one mirror-ascent step from $\mathbf{p}_X$ solves

$$\arg\max_{\boldsymbol{\pi} \in \Delta(\mathcal{A})} \left\{ \eta \langle \nabla f(\mathbf{p}_X), \boldsymbol{\pi} - \mathbf{p}_X \rangle - D_\Omega(\boldsymbol{\pi}, \mathbf{p}_X) \right\}. \tag{26}$$

Since $\nabla f(\mathbf{p}_X) = \mathbf{r}$, dropping the constant $-\eta\langle \mathbf{r}, \mathbf{p}_X \rangle$ recovers the same maximizer as Eq. (4) after division by $\eta > 0$. This is the familiar entropic mirror/exponentiated-gradient construction (Beck & Teboulle, 2003; Kivinen & Warmuth, 1997). At the next report state, a new $\mathbf{p}_X, \mathbf{p}_R, r$ is evaluated. The procedure is not repeatedly optimizing one fixed report-level objective, and it does not carry the previous coordinate's $\boldsymbol{\pi}_\eta$ forward as the next mirror base.

A.6 Restricted logits, additive constants, and the mode

Let $\boldsymbol{\ell}_X$ and $\boldsymbol{\ell}_R$ denote the network logits restricted to $\mathcal{A}$. Then

$$p_X(a) = \frac{e^{\ell_X(a)}}{Z_X}, \quad p_R(a) = \frac{e^{\ell_R(a)}}{Z_R}, \qquad Z_X = \sum_{b \in \mathcal{A}} e^{\ell_X(b)}, \quad Z_R = \sum_{b \in \mathcal{A}} e^{\ell_R(b)}. \tag{27}$$

The predictive log ratio and raw logit contrast differ by a shared scalar:

$$r(a) = \ell_X(a) - \ell_R(a) - \log Z_X + \log Z_R. \tag{28}$$

Thus raw logit contrast is not numerically the log-probability ratio, but the difference is action-independent. In particular, the ratio has an absolute probability meaning that an arbitrarily shifted raw logit contrast does not. Taking logs of Eq. (18) gives

$$\log \pi_\eta(a) = (1+\eta)\ell_X(a) - \eta\ell_R(a) - (1+\eta)\log Z_X + \eta \log Z_R - \log Z_\eta \tag{29}$$

$$= g_\eta(a) - \log \sum_{b \in \mathcal{A}} e^{g_\eta(b)}, \qquad \mathbf{g}_\eta = (1+\eta)\boldsymbol{\ell}_X - \eta\boldsymbol{\ell}_R. \tag{30}$$

The equality follows since $\log \sum_b e^{g_\eta(b)} = \log Z_\eta + (1+\eta)\log Z_X - \eta \log Z_R$. With $s = 1+\eta$,

$$\mathbf{g}_\eta = \boldsymbol{\ell}_X + \eta(\boldsymbol{\ell}_X - \boldsymbol{\ell}_R) = \boldsymbol{\ell}_R + s(\boldsymbol{\ell}_X - \boldsymbol{\ell}_R) = \mathbf{g}^{(s)}. \tag{31}$$

Softmax is order-preserving, so

$$\arg\max_{a \in \mathcal{A}} \pi_\eta(a) = \arg\max_{a \in \mathcal{A}} g_\eta(a). \tag{32}$$

Applying the same fixed deterministic selector τ to either argmax set produces the same action. The optimizing distribution may have several modes despite being uniquely determined. The derivation does not require a particular smallest-ID or vocabulary-order convention; the exact frozen tie ordering is not recovered by this algebra. At $\eta = 0$, the expression with η^{-1} is not evaluated. Instead, finite support gives $Z_\eta \to 1$ and $\boldsymbol{\pi}_\eta \to \mathbf{p}_X$ as $\eta \downarrow 0$; the score directly reduces to $\mathbf{g}_0 = \boldsymbol{\ell}_X = \mathbf{g}^{(1)}$. The equivalent scaled objective

$$\widetilde{\mathcal{J}}_\eta(\boldsymbol{\pi}) = \eta \sum_a \pi(a) r(a) - \mathrm{KL}(\boldsymbol{\pi} \| \mathbf{p}_X) \tag{33}$$

is well-defined at zero and then uniquely maximized by $\mathbf{p}_X$. If $\mathbf{p}_R = \mathbf{p}_X$, the reward is identically zero and $\boldsymbol{\pi}_\eta = \mathbf{p}_X$ for every finite $\eta \geq 0$.

No numerical optimizer, explicit probability powers, or entropy term needs to be evaluated by the deployed mode decoder. For mathematical checks one can use log-softmax and log-sum-exp. For deployment, the equivalent affine score on finite allowed logits suffices. Unsupported actions must be excluded before expressions such as $(-\infty) - (-\infty)$ are formed. Exact-arithmetic equivalence does not guarantee bitwise equality of reassociated floating-point operations near a tie.

A.7 The training mean is an ambient Euclidean barycenter

Assume $N_{\text{train}} \geq 1$ finite frozen tensors $\mathbf{C}_n$ occupy the same $Q \times d_e$ coordinate system. Shared ordered query slots make the slotwise Frobenius objective well defined. Define

$$B(\mathbf{C}) = \frac{1}{N_{\text{train}}} \sum_n \|\mathbf{C} - \mathbf{C}_n\|_F^2, \qquad \bar{\mathbf{C}}_{\text{train}} = \frac{1}{N_{\text{train}}} \sum_n \mathbf{C}_n. \tag{34}$$

Expand $\mathbf{C} - \mathbf{C}_n = (\mathbf{C} - \bar{\mathbf{C}}_{\text{train}}) + (\bar{\mathbf{C}}_{\text{train}} - \mathbf{C}_n)$. The average cross term vanishes because $\sum_n (\bar{\mathbf{C}}_{\text{train}} - \mathbf{C}_n) = 0$, giving

$$B(\mathbf{C}) = B(\bar{\mathbf{C}}_{\text{train}}) + \|\mathbf{C} - \bar{\mathbf{C}}_{\text{train}}\|_F^2. \tag{35}$$

Therefore $\bar{\mathbf{C}}_{\text{train}}$ is the unique minimizer over the full ambient space. Equivalently, $\nabla B(\mathbf{C}) = 2(\mathbf{C} - \bar{\mathbf{C}}_{\text{train}})$ and the Hessian is $2\mathbf{I}$ after vectorization. No independent-sampling, population-consistency, or calibrated-feature assumption is needed for this finite-sample identity.

This objective concerns distances between evidence tensors, not report prediction. The result does not establish that the mean is a realizable single-study context, a preferable decoding reference, or an optimal report distribution. In general,

$$p_{\boldsymbol{\theta}}(a \mid h, \bar{\mathbf{C}}_{\text{train}}; \mathcal{A}) \not\equiv \frac{1}{N_{\text{train}}} \sum_n p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_n; \mathcal{A}). \tag{36}$$

Here $\not\equiv$ indicates that the equality is not a general identity, although it may hold in special cases. TARING uses the mean-context reference on the left, whereas the right side averages predictions over individual training contexts. The mean is computed from frozen evidence tensors produced by an encoder learned during conditional training.

A.8 State matching is a definition of the intervention

For $h = (\mathbf{z}_j, t_j, i)$, the defined reward is $r_h(a) = \log p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_X; \mathcal{A}) - \log p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_R; \mathcal{A})$. Replacing the reference state by $\widetilde{h}$ instead gives

$$\log \frac{p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_X; \mathcal{A})}{p_{\boldsymbol{\theta}}(a \mid \widetilde{h}, \mathbf{C}_R; \mathcal{A})} = r_h(a) + \log \frac{p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_R; \mathcal{A})}{p_{\boldsymbol{\theta}}(a \mid \widetilde{h}, \mathbf{C}_R; \mathcal{A})}. \tag{37}$$

The additional term is a report-state contrast that can depend on the action. Matching the masked state and time removes this state-mismatch term, but does not by itself imply universal removal of language bias or separation of clinical evidence from all language–image interactions. Such claims require additional structure and are not used here.

A fixed $\mathbf{C}_R$ does not make $\mathbf{p}_R$ fixed: changing $\mathbf{z}_j$, t_j, or i changes its conditioning. In particular, the reference branch receives any study-specific words already revealed in $\mathbf{z}_j$. It is therefore a local context-replacement prediction, not an unconditional report prior. Within one TARING decision both branches share $\mathbf{z}_j$. Across complete no-reference and TARING runs, selected content may differ and later report states generally differ; sharing the support and reveal variables does not prevent that endogenous divergence.

A.9 Terminal support is outside the optimized simplex

Using the current-study terminal boundary $\widehat{L}$ and initial state $\mathbf{z}_0$ defined in Eq. (8), the simplex $\Delta(\mathcal{A})$ contains content actions, not terminal positions or special tokens. At each round the shared reveal variables select J_j; for every $i \in J_j$, Eq. (7) is computed from the same pre-update $\mathbf{z}_j$ and assigned simultaneously. Empty queried sets require no content update. The terminal head, report-support quantile, and reveal schedule are fixed outside the variational problem. Accordingly, the intervention analyzed here is confined to content-score construction.

A.10 Limits of the derivation

The derivation assumes finite real logits on a common valid support and strictly positive probabilities for both branches. Structural zeros within $\mathcal{A}$ violate these assumptions; in particular, $p_R(a) = 0$ with $p_X(a) > 0$ makes the reward nonfinite. Zeros outside $\mathcal{A}$ do not affect the derivation. Clipping probabilities or adding an epsilon changes the objective.

The KL penalty regularizes the categorical distribution but does not impose a tokenwise plausibility constraint or provide a clinically meaningful risk bound. The implemented method uses no such threshold.

The optimality result applies to a single categorical distribution at a fixed state. It does not imply joint report optimality or a calibrated posterior for simultaneous coordinatewise mode decoding. A larger predictive ratio alone does not establish a causal interpretation, factual accuracy, hallucination reduction, safety, or clinical utility.

B Fixed-State Theory: Proofs and Counterexamples

B.1 Standing assumptions and relation to the decoder

We retain the positive finite laws and common finite nonempty action support from Appendix A.1. Parameters, $h = (\mathbf{z}_j, t_j, i)$, masks, terminal boundary, $\mathbf{C}_X$ and $\mathbf{C}_R$ are fixed. All logarithms are natural. Scale comparisons hold these objects fixed; they are not comparisons of states reached by separately generated reports. There is no probability floor, changing action threshold, temperature change, or optimization over network parameters. The production method takes the mode of the represented law, not a draw from it.

We use $\boldsymbol{\pi}_\eta$ with $\boldsymbol{\pi}_0 = \mathbf{p}_X$ and $\eta = s - 1 \geq 0$. The normalized predictive ratio and raw logit contrast satisfy

$$r(a) = d(a) - \log Z_X + \log Z_R, \qquad d(a) = \ell_X(a) - \ell_R(a). \tag{38}$$

Their pairwise differences agree, but their absolute values need not.

The monotone score conclusions below apply to any maximizers, including ties. When a claim concerns the *identity* of the selected action under a shrinking tie set, we explicitly assume a fixed total priority order on $\mathcal{A}$. This is a mathematical tie convention, not a claim to have recovered the archived token ordering.

B.2 Matched-state cancellation and mismatch residuals

Proposition 1 (Matched-state cancellation)**.** *Suppose*

$$\ell(a; h, \mathbf{C}) = b(a; h) + e(a; h, \mathbf{C}) + k(h, \mathbf{C}), \tag{39}$$

where b is independent of evidence context and k is action-independent. For $\Delta_{ab} f = f(a) - f(b)$, matched-state ratio contrasts satisfy

$$r_h(a) - r_h(b) = \Delta_{ab}\mathbf{e}(h, \mathbf{C}_X) - \Delta_{ab}\mathbf{e}(h, \mathbf{C}_R). \tag{40}$$

Querying the reference at $\tilde{h}$ adds the action-dependent residual $\Delta_{ab}\{\mathbf{b}(h) - \mathbf{b}(\tilde{h})\} + \Delta_{ab}\{\mathbf{e}(h, \mathbf{C}_R) - \mathbf{e}(\tilde{h}, \mathbf{C}_R)\}$.

Proof of Proposition 1. Let $\ell^h_X(a) = \ell(a; h, \mathbf{C}_X)$ and $\ell^h_R(a) = \ell(a; h, \mathbf{C}_R)$. Softmax normalizers are action-independent, so

$$r_h(a) - r_h(b) = \Delta_{ab}(\boldsymbol{\ell}^h_X - \boldsymbol{\ell}^h_R). \tag{41}$$

Substitute Eq. (39). The terms $b(\cdot; h)$ cancel between contexts, and $k(h, \mathbf{C}_X) - k(h, \mathbf{C}_R)$ cancels between actions. This proves Eq. (40). For the mismatched reward

$$\widetilde{r}(a) = \log p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_X; \mathcal{A}) - \log p_{\boldsymbol{\theta}}(a \mid \widetilde{h}, \mathbf{C}_R; \mathcal{A}), \tag{42}$$

add and subtract $\boldsymbol{\ell}^h_R$ in its pairwise contrast to obtain

$$\begin{aligned} \widetilde{r}(a) - \widetilde{r}(b) &= r_h(a) - r_h(b) + \Delta_{ab}(\boldsymbol{\ell}^h_R - \boldsymbol{\ell}^{\widetilde{h}}_R) \\ &= r_h(a) - r_h(b) + \Delta_{ab}\{\mathbf{b}(h) - \mathbf{b}(\widetilde{h})\} \\ &\quad + \Delta_{ab}\{\mathbf{e}(h, \mathbf{C}_R) - \mathbf{e}(\widetilde{h}, \mathbf{C}_R)\}. \end{aligned} \tag{43}$$

This is the complete contamination term: the reference interaction can change along with the shared state term. A changed refinement time is included by changing the time component of h. □

What does not cancel. Under the same decomposition the guided logits themselves are

$$g_\eta(a) = b(a; h) + (1 + \eta)e(a; h, \mathbf{C}_X) - \eta e(a; h, \mathbf{C}_R) + (1 + \eta)k(h, \mathbf{C}_X) - \eta k(h, \mathbf{C}_R). \tag{44}$$

Thus b is retained once, not eliminated from the decoder. Moreover, for any action/state function v, replacing b by $b + v$ and e by $e - v$ leaves all logits unchanged. Without independent structural information, the decomposition is not identifiable and can even be chosen trivially with $b = 0$. Proposition 1 therefore proves cancellation of a stipulated shared term, not universal language bias removal, causal separation, or an empirically validated mechanism. Mismatch can have zero residual by coincidence; the proposition does not say that matching is necessary in every possible model.

A decision-level mismatch bound. A useful statement does not require the decomposition. Define

$$\delta(a) = \log p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_R; \mathcal{A}) - \log p_{\boldsymbol{\theta}}(a \mid \widetilde{h}, \mathbf{C}_R; \mathcal{A}). \tag{45}$$

The mismatched distributional score differs from the matched score by $\eta\delta(a)$, up to an irrelevant action-independent constant. If the matched score has unique winner a_η and margins $M_b = g_\eta(a_\eta) - g_\eta(b) > 0$, that winner is preserved whenever

$$M_b > \eta\{\delta(b) - \delta(a_\eta)\} \quad (b \neq a_\eta). \tag{46}$$

Indeed, subtract the two perturbed scores. A sufficient, shift-invariant condition is $\min_{b \neq a_\eta} M_b > \eta \operatorname{osc}(\delta)$, where $\operatorname{osc}(\delta) = \max_a \delta(a) - \min_a \delta(a)$. For a singleton action set the claim is vacuous. This bound describes local decision sensitivity rather than clinical correctness. Appendix D.6 evaluates the corresponding one-step reference-time perturbation empirically in a post-hoc seed-0 diagnostic of the frozen decoder.

B.3 Induced information budget and trust-region interpretation

Proposition 2 (Induced information budget). *Let* $A(\eta) = \log \sum_a p_X(a) e^{\eta r(a)}$ *and* $\varepsilon_\eta = \mathrm{KL}(\boldsymbol{\pi}_\eta \| \mathbf{p}_X)$. *Then*

$$A'(\eta) = \mathbb{E}_{\boldsymbol{\pi}_\eta} r, \quad A''(\eta) = \mathrm{Var}_{\boldsymbol{\pi}_\eta}(r), \quad \varepsilon'_\eta = \eta \, \mathrm{Var}_{\boldsymbol{\pi}_\eta}(r) \geq 0. \tag{47}$$

For $\eta > 0$, $\boldsymbol{\pi}_\eta$ *also uniquely maximizes* $\mathbb{E}_{\boldsymbol{\pi}} r$ *over the KL ball with radius* ε_η.

Mirror step and unique tilt. Appendix A.5 derives the entropic mirror-ascent interpretation for the fixed reward $\mathbf{r}$. The update is unique for $\eta > 0$ and reduces to $\mathbf{p}_X$ at $\eta = 0$.

Trust-region proof, including boundary candidates. Write $\mu_\eta = \langle \boldsymbol{\pi}_\eta, \mathbf{r} \rangle$ and $\varepsilon_\eta = \mathrm{KL}(\boldsymbol{\pi}_\eta \| \mathbf{p}_X)$. Recall that $A(\eta) = \log Z_\eta = \log \sum_{a \in \mathcal{A}} p_X(a) e^{\eta r(a)}$. For any simplex law $\mathbf{q}$, including laws with zero masses,

$$\mathrm{KL}(\mathbf{q} \| \boldsymbol{\pi}_\eta) = \mathrm{KL}(\mathbf{q} \| \mathbf{p}_X) - \eta \langle \mathbf{q}, \mathbf{r} \rangle + A(\eta). \tag{48}$$

Use $A(\eta) = \eta \mu_\eta - \varepsilon_\eta$ to rearrange this as

$$\eta \{ \mu_\eta - \langle \mathbf{q}, \mathbf{r} \rangle \} = \varepsilon_\eta - \mathrm{KL}(\mathbf{q} \| \mathbf{p}_X) + \mathrm{KL}(\mathbf{q} \| \boldsymbol{\pi}_\eta). \tag{49}$$

If $\mathbf{q}$ is feasible for the induced trust-region problem in Proposition 2, both terms on the right are nonnegative. Equality requires $\mathbf{q} = \boldsymbol{\pi}_\eta$. This proves the trust-region part of Proposition 2 for every finite $\eta > 0$. For $\eta = 0$, the induced budget is zero, whose sole feasible law is $\mathbf{p}_X$.

B.4 Exact derivatives and the meaning of the budget

Proof of the derivative part of Proposition 2. Every term of $Z_\eta = \sum_a p_X(a) e^{\eta r(a)}$ is positive and smooth. Because the sum is finite,

$$A'(\eta) = \frac{\sum_a p_X(a) r(a) e^{\eta r(a)}}{Z_\eta} = \sum_a \pi_\eta(a) r(a), \tag{50}$$

$$\frac{d \pi_\eta(a)}{d \eta} = \pi_\eta(a) \{ r(a) - \mu_\eta \}, \tag{51}$$

$$A''(\eta) = \sum_a \pi_\eta(a) r(a)^2 - \mu_\eta^2 = \mathrm{Var}_{\boldsymbol{\pi}_\eta}(r). \tag{52}$$

The exact log ratio to the base is $\log(\boldsymbol{\pi}_\eta / \mathbf{p}_X) = \eta \mathbf{r} - A(\eta)$; averaging it gives

$$\varepsilon_\eta = \eta A'(\eta) - A(\eta), \qquad \varepsilon'_\eta = A'(\eta) + \eta A''(\eta) - A'(\eta) = \eta \, \mathrm{Var}_{\boldsymbol{\pi}_\eta}(r). \tag{53}$$

For nonconstant r, the variance is positive at every finite η since $\boldsymbol{\pi}_\eta$ has full support. Hence μ_η strictly increases, and ε_η strictly increases on $[0, \infty)$ even though its right derivative at zero is zero. Constant $r = k$ implies $\mathbf{p}_X = e^k \mathbf{p}_R$; summing yields $k = 0$ and $\mathbf{p}_X = \mathbf{p}_R$, proving the degenerate clause. □

What monotonicity does and does not cover. A fixed set of states, averaged with fixed nonnegative weights, inherits the monotonicities by summation. Separately generated trajectories need not visit that same set or use scale-independent weights. Neither this proof nor the mirror-step derivation implies monotone report metrics, per-token truth probability, or external factuality. Even distributional entropy need not fall: an initially favored action can lose mass to a higher-ratio competitor before that competitor dominates. The derivative of KL is nonnegative because $\eta \geq 0$; on the interpretive negative-step extension its sign is reversed. Finally, $\boldsymbol{\pi}_\eta$ is not a Dirac mass at the implementation's argmax, so its KL is not a KL bound on the actual deterministic output law.

B.5 Deterministic modes, breakpoints, and tie conventions

Proposition 3 (Deterministic evidence–compatibility frontier). *Let* $a_0 \in \arg\max_a \ell_X(a)$. *For competitor* b, *define the current-branch margin* $m_b = \ell_X(a_0) - \ell_X(b) \geq 0$ *and relative advantage* $\Delta r_b = r(b) - r(a_0)$. *Then* b *strictly outranks* a_0 *under* TARING *exactly when*

$$\Delta r_b > 0 \quad \textit{and} \quad \eta > \frac{m_b}{\Delta r_b}. \tag{54}$$

Across increasing η, *the selected action's relative contrast cannot decrease and its current-branch logit cannot increase; equality in (54) is a tie.*

Proof of Proposition 3. Write $c(a) = \ell_X(a)$ and $d(a) = \ell_X(a) - \ell_R(a)$. For $0 \le \eta_1 < \eta_2$, choose $a_k = a_{\eta_k} \in \arg\max_{a\in\mathcal{A}}\{c(a) + \eta_k d(a)\}$ and set $c_k = c(a_k)$ and $d_k = d(a_k)$ for $k = 1, 2$. Optimality at each scale gives

$$c_1 + \eta_1 d_1 \ge c_2 + \eta_1 d_2, \qquad c_2 + \eta_2 d_2 \ge c_1 + \eta_2 d_1. \tag{55}$$

Consequently

$$\eta_1(d_2 - d_1) \le c_1 - c_2 \le \eta_2(d_2 - d_1). \tag{56}$$

The interval is nonempty only if $d_2 - d_1 \ge 0$ because $\eta_2 > \eta_1$. Its left endpoint is nonnegative because $\eta_1 \ge 0$, proving $c_1 \ge c_2$. The same ordering holds for r by Eq. (38), and for selected p_X because its normalizer is fixed. No uniqueness or differentiation of argmax is required.

For a current-prediction maximizer a_0 and competitor b, subtract their guided scores:

$$g_\eta(b) - g_\eta(a_0) = -m_b + \eta\Delta r_b. \tag{57}$$

If $\Delta r_b > 0$, the gap is negative, zero, or positive according as η is below, equal to, or above $m_b/\Delta r_b$. If $\Delta r_b \le 0$, it is never positive for $\eta \ge 0$. This proves the strict-overtake criterion. □

Exact first departure from the current-prediction mode. With a_0 selected by a fixed total priority order, set

$$\eta_\star = \min_{b:\Delta r_b > 0} \frac{m_b}{\Delta r_b}, \qquad \min \varnothing = +\infty. \tag{58}$$

For $0 \le \eta < \eta_\star$, no competitor strictly beats a_0. Any persistent tied competitor was already tied at $\eta = 0$ and has the same slope; priority therefore keeps a_0 selected. If $\eta_\star < \infty$, a_0 remains a maximizer at $\eta = \eta_\star$, with selection determined by priority, and is not a maximizer at any $\eta > \eta_\star$. When $m_b = 0$ and $\Delta r_b > 0$, the critical scale is zero and b strictly outranks a_0 for every positive step. If $\Delta r_b = 0$, equality with a_0 either persists at all scales ($m_b = 0$) or never occurs ($m_b > 0$). Initial ties with $\Delta r_b < 0$ disappear immediately for positive steps.

A competitor's crossing of a_0 is a *pairwise* event. The actual mode is determined by the upper envelope of all lines $c(a) + \eta d(a)$. This envelope is convex and piecewise affine: between finitely many intersections of unequal-slope lines its set of maximizers is fixed. Its one-sided slopes are ordered, explaining the monotone score ordering in Proposition 3. Under fixed priority, a fixed-state indicator $\mathbf{1}\{a_\eta \ne a_0\}$ cannot return from one to zero after its first departure. This does not imply a monotone change rate between separately generated reports with different endogenous states.

Why a fixed selector is not automatically fixed priority. For $\mathbf{c} = (0, 0, 0)$ and $\mathbf{d} = (1, 1, 0)$, all actions tie at zero and only the first two tie for positive η. A deterministic set-selector can pick the first action from the three-way set but the second from the two-way set. The first action maximizes both $\mathbf{c}$ and $\mathbf{d}$ yet its selected identity changes. These scores are realizable by positive laws: $\mathbf{p}_X = \mathrm{softmax}(\mathbf{c})$ and $\mathbf{p}_R = \mathrm{softmax}(\mathbf{c} - \mathbf{d})$; normalizing only shifts all slopes by a common constant. Thus statements about unchanging identity require fixed priority or uniqueness. The score-ordering inequalities and strict pairwise crossings do not require that extra convention.

B.6 Special cases and counterexamples to stronger claims

All constructions in this subsection are finite synthetic examples, not radiology measurements.

A large ratio does not impose a current-probability floor. For $0 < v < 1/2$ and $k > 1$, take

$$\mathbf{p}_X = (1 - v, v), \qquad \mathbf{p}_R = (1 - v^k, v^k). \tag{59}$$

The second action has current probability v and predictive ratio v^{1-k}, which diverges as $v \downarrow 0$. Its guided odds are

$$\frac{\pi_\eta(2)}{\pi_\eta(1)} = \frac{v^{1+\eta-k\eta}(1 - v^k)^\eta}{(1 - v)^{1+\eta}}. \tag{60}$$

For *any fixed* $\eta > 0$, choose $k > (1 + \eta)/\eta$; then the second action becomes the unique mode for sufficiently small v, even though $p_X(2)$ is arbitrarily small. At the frozen step $\eta = 0.75$, $k = 3$ suffices. At small enough step the same high-ratio action can still lose because its conditional margin has not been paid. Hence neither the proximal penalty nor the induced budget creates a tokenwise threshold.

Predictive gain can reduce correctness. Let action 1 be correct and action 2 incorrect in a binary, synthetic single-coordinate task, with

$$\mathbf{p}_X = (4/5, 1/5), \qquad \mathbf{p}_R = (99/100, 1/100). \tag{61}$$

The ratio advantage of action 2 is $\log(99/4)$; the current log margin is $\log 4$. The unique current-prediction mode is correct, but every

$$\eta > \frac{\log 4}{\log(99/4)} \tag{62}$$

selects the incorrect action, including $\eta = 0.75$. Meanwhile μ_η strictly increases and $d\pi_\eta(1)/d\eta = \pi_\eta(1)\{r(1) - \mu_\eta\} < 0$. Thus both deterministic accuracy and probability assigned to truth can fall while expected predictive contrast rises. Treating the binary token as the only assertion in a report also disproves any universal report factuality implication. This is a logical counterexample, not a claim about a measured clinical case. Entropy initially increases here: writing $u = \pi_\eta(1)$ gives $H' = -u' \log[u/(1-u)] > 0$ while $u > 1/2$ and $u' < 0$.

A local tilt budget is not an argmax-output budget. Take $\mathbf{p}_X = \mathbf{p}_R = (1/2, 1/2)$. Then $\varepsilon_\eta = 0$ for every step. A deterministic tie-selected output has law $\boldsymbol{\delta}_{a_\star}$, with $\mathrm{KL}(\boldsymbol{\delta}_{a_\star} \| \mathbf{p}_X) = \log 2 > 0$. This simple example rules out transferring the local tilt budget to the actual deterministic output distribution. No report-level KL bound is asserted for the implemented coordinatewise argmax trajectory.

C Conditional Training and Terminal/Reveal Process

C.1 Conditional masked generator

For a study $X = \{(\mathbf{x}_m, v_m)\}_{m=1}^M$, a frozen image encoder supplies visual tokens. After normalization and view/modality embeddings, a learned-query set encoder maps the valid evidence rows to $\mathbf{C}_X = C_{\boldsymbol{\psi}}(X) \in \mathbb{R}^{Q \times d_e}$. Let L be the terminal position, $I_L = \{1, \dots, L-1\}$ the content coordinates, and $n = |I_L|$. At a sampled mask probability $u \in (0, 1)$, independently draw $\xi_i \sim \mathrm{Bernoulli}(u)$ and mask each $i \in I_L$ when $\xi_i = 1$. If $\mathbf{Z}_t^{(i)}$ denotes the state with coordinate i forced masked, the inverse-mask-weighted estimator is

$$\widehat{\mathcal{L}}_{\mathrm{content}} = \frac{1}{\max(n, 1)} \sum_{i \in I_L} \frac{\xi_i}{u} \mathrm{CE}_{\mathcal{A}}(\boldsymbol{\ell}_{\boldsymbol{\theta}, i}(\mathbf{Z}_t, \mathbf{C}_X, t), y_i) \tag{63}$$

Conditioning on all masks except ξ_i gives $\mathbb{E}[\xi_i/u \mid \boldsymbol{\xi}_{-i}] = 1$, so (63) is unbiased for the mean forced-mask reconstruction risk at the sampled mask rate. A separate categorical head predicts the terminal position. The empirical-reference branch is absent during training.

Model configuration. The evidence interface has dimension 768, and the two-layer 128-query resampler uses 12 attention heads. The backbone is `GSAI-ML/LLaDA-8B-Instruct`, with a 128-token horizon and a categorical terminal head over the non-BOS positions. The LoRA configuration and target modules are recorded in the experiment metadata.

C.2 Reveal survival law

The canonical decoder uses the cumulative reveal schedule

$$\kappa(t) = \sin^2\left(\frac{\pi t}{2}\right), \qquad t \in [0, 1],$$

which satisfies $\kappa(0) = 0$ and $\kappa(1) = 1$. Let M_j be the positions still masked at grid time t_j. For $\kappa(t_j) < 1$, conditional on being masked, a coordinate is revealed with

$$\rho_j = \frac{\kappa(t_{j+1}) - \kappa(t_j)}{1 - \kappa(t_j)}. \tag{64}$$

If $\Pr(i \in M_j) = 1 - \kappa(t_j)$, then

$$\Pr(i \in M_{j+1}) = (1 - \kappa(t_j))(1 - \rho_j) = 1 - \kappa(t_{j+1}), \tag{65}$$

q_{EOS}	s=1.00	1.25	1.50	1.75	2.00
0.55	0.520	0.534	0.541	0.537	0.543
0.58	0.512	0.532	0.537	0.542	0.546
0.60	0.523	0.543	0.556	0.551	0.548
0.90	0.501	0.516	0.519	0.529	0.531

Table 4: Validation CheXbert micro-F1 under reveal seed 0. The two strongest guided configurations at $q_{\text{EOS}} = 0.60$ are subsequently evaluated over seeds $\{0, 1, 2\}$, yielding 0.5469 ± 0.0082 for $s = 1.50$ and 0.5474 ± 0.0040 for $s = 1.75$.

Metric	**No ref.**	**TARING**	**Δ**	**95% CI**
CheX P	0.538	0.5438	+0.0063	[-.0001, .0130]
CheX R	0.489	0.5251	+0.0363	[.0295, .0429]
Micro-F1	0.5120	0.5343	+0.0223	[.0165, .0281]
Macro-F1	0.3087	0.3227	+0.0140	[.0050, .0230]
BLEU-1	0.4303	0.4305	+0.0003	[-.0050, .0053]
BLEU-2	0.2796	0.2737	-0.0059	[-.0098, -.0020]
BLEU-3	0.1841	0.1779	-0.0062	[-.0092, -.0032]
BLEU-4	0.1222	0.1171	-0.0050	[-.0076, -.0025]
METEOR	0.1710	0.1701	-0.0008	[-.0019, .0002]
ROUGE-L	0.3140	0.3077	-0.0064	[-.0081, -.0045]

Table 5: Canonical controlled effects on the test set. Differences are TARING minus no-reference. Confidence intervals are paired 10,000-draw subject-cluster percentile bootstrap intervals. CheX denotes CheXbert-derived metrics.

so induction from the all-masked initialization establishes the grid-point survival law while $\kappa(t_j) < 1$.

If $\kappa(t_j) = 1$, monotonicity gives $\kappa(t_{j+1}) = 1$ and $M_j = \varnothing$ almost surely. When $\kappa(t_{S-1}) < 1$, the last transition has $\rho_{S-1} = 1$.

All positions selected at round j are scored from the same pre-update state and updated simultaneously. In the final round, every remaining masked position is included in the queried set.

D Extended Ablation Studies

D.1 Validation operating-point selection

Table 4 reports the complete seed-0 grid used for validation-only operating-point selection. The two strongest guided configurations at $q_{\text{EOS}} = 0.60$ were subsequently evaluated over reveal seeds $\{0, 1, 2\}$, yielding 0.5469 ± 0.0082 for $s = 1.50$ and 0.5474 ± 0.0040 for $s = 1.75$. The latter was fixed before canonical test evaluation.

Unless otherwise stated, this section uses the canonical TARING configuration defined in Section 4 and its no-reference ablation ($s = 1$) defined in Section 6. The saved evaluation files record the no-reference condition as `conditional`.

D.2 Full controlled effects and paired uncertainty

Table 5 reports the complete controlled comparison between no-reference decoding and TARING. Values for each condition are the means over reveal seeds $\{0, 1, 2\}$. Differences are TARING minus no-reference. Confidence intervals are paired 95% percentile intervals from 10,000 subject-cluster bootstrap resamples that retain the paired decoding conditions and associated reveal seeds.

D.3 Reveal-seed-specific results

Table 6 reports CheXbert micro-F1 for each reveal seed before aggregation. TARING improves micro-F1 for all three seeds, with gains ranging from $+0.0192$ to $+0.0259$. This directional consistency

Reveal seed	No reference	TARING	Δ
0	0.5178	0.5369	+0.0192
1	0.5064	0.5323	+0.0259
2	0.5118	0.5337	+0.0219

Table 6: Reveal-seed-specific CheXbert micro-F1 on the canonical test set. Differences are TARING minus no-reference. These seed-specific corpus effects are descriptive and are distinct from the subject-cluster bootstrap intervals reported in Table 5.

Condition	Pos./report	FP/report	FN/report	Exact match	Words	Sentences	Negations
No reference	2.5921	1.1988	1.4574	0.0653	62.01	8.43	2.31
TARING	2.7525	1.2557	1.3538	0.0714	61.06	7.61	1.77
Δ	+0.1604	+0.0569	-0.1036	+0.0061	-0.95	-0.82	-0.54

Table 7: Report-level coverage and disagreement summaries over the canonical test outputs. FP and FN denote CheXbert-derived disagreements with reference-report labels, not clinician-adjudicated errors. Exact match denotes agreement of the complete 14-label CheXbert vector.

indicates that the aggregate improvement is not driven by a single reveal realization. The seeds quantify reveal-order variability for one frozen checkpoint and are not independent training runs; inferential uncertainty is instead reported using the paired subject-cluster bootstrap in Table 5.

D.4 CheXbert coverage and disagreement patterns

Table 7 summarizes report-level coverage and disagreement patterns. Relative to no-reference decoding, TARING produces 0.1604 more positive CheXbert labels per report, reduces false-negative disagreements by 0.1036 per report, and increases false-positive disagreements by 0.0569 per report. The complete 14-label exact-match rate increases by 0.0061. Reports are slightly shorter under TARING (62.01 to 61.06 words on average), so the increased positive-label coverage is not explained by longer outputs. These quantities are derived from reference-report CheXbert labels and are not clinician-adjudicated errors.

The aggregate pattern is heterogeneous across labels. Cardiomegaly, edema, pleural effusion, pneumothorax, and support devices show substantial reductions in false-negative disagreements together with increases in false-positive disagreements, whereas atelectasis shows the opposite pattern. Table 8 reports the full label-specific counts.

D.5 Graph-based report agreement

We additionally evaluate report/reference graph agreement using RadGraph. Scores are averaged across the three reveal seeds within each study before paired subject-cluster resampling. All three graph-agreement variants decrease under TARING, showing that the improvement in CheXbert-based label agreement does not extend uniformly to graph-based report agreement.

These graph scores measure agreement between generated and reference reports and do not establish image-grounded factuality or radiologist-assessed clinical correctness.

D.6 Matched-state mismatch diagnostic

We perform a post-hoc fixed-trajectory diagnostic on all $2{,}357$ test studies under reveal seed 0. The canonical TARING trajectory and current branch are held fixed, while only the reference-side refinement time is shifted by one valid step. The mismatched decision is never fed back into generation. Canonical replay reproduces all $182{,}409$ queried token decisions. Ninety-four decisions are excluded because no previous refinement time is available.

For each eligible queried decision, the current-study logits and canonical trajectory are fixed, and only the reference branch is evaluated at the preceding valid refinement time. The guided flip rate is the

Label	NR FP	TARING FP	ΔFP	NR FN	TARING FN	ΔFN
Enlarged cardiomediastinum	65	53	-12	231	237	+6
Cardiomegaly	520	602	+82	276	184	-92
Lung opacity	150	134	-16	711	705	-6
Lung lesion	13	31	+18	124	121	-3
Edema	448	525	+77	218	159	-59
Consolidation	80	64	-16	189	194	+5
Pneumonia	57	57	0	423	414	-9
Atelectasis	498	327	-171	465	520	+55
Pneumothorax	61	126	+65	47	32	-15
Pleural effusion	449	487	+38	194	173	-21
Pleural other	13	25	+12	78	76	-2
Fracture	48	36	-12	110	111	+1
Support devices	248	327	+79	225	142	-83
No finding	150	139	-11	103	109	+6

Table 8: CheXbert label-specific disagreement counts for the fixed seed-0 decode on the 2,357-study test set. NR denotes no-reference decoding, and differences are TARING minus no-reference. This analysis is descriptive and is not used for efficacy inference. FP and FN are defined relative to reference-report-derived CheXbert labels and are not radiologist-adjudicated clinical errors.

RadGraph	No ref.	TARING	Δ	95% CI
Complete	0.1775	0.1724	-0.0052	[-0.0078, -0.0026]
Partial	0.2456	0.2408	-0.0048	[-0.0075, -0.0020]
Simple	0.2702	0.2667	-0.0035	[-0.0063, -0.0008]

Table 9: Canonical RadGraph report/reference agreement. Differences are TARING minus no-reference; higher is better. Confidence intervals use 10,000 paired subject-cluster bootstrap draws after averaging reveal-seed scores within study.

fraction of decisions whose TARING argmax changes under this perturbation, whereas the reference top-1 flip rate measures the corresponding change in the reference branch alone. JS divergence is computed between the matched and mismatched reference distributions over the common valid action set.

Let

$$\delta(a) = \log p_{\boldsymbol{\theta}}(a \mid h, \mathbf{C}_R; \mathcal{A}) - \log p_{\boldsymbol{\theta}}(a \mid \widetilde{h}, \mathbf{C}_R; \mathcal{A}),$$

and

$$\mathrm{osc}(\delta) = \max_{a \in \mathcal{A}} \delta(a) - \min_{a \in \mathcal{A}} \delta(a).$$

For the matched winner a_η, let $M_b = g_\eta(a_\eta) - g_\eta(b)$. The exact-bound rate is the fraction of eligible decisions for which Eq. (46) holds for every competitor $b \neq a_\eta$. The strong-bound rate is the fraction satisfying the stronger sufficient condition

$$\min_{b \neq a_\eta} M_b > \eta \ \mathrm{osc}(\delta).$$

Table 10 summarizes the overall and stage-specific mismatch diagnostics. The overall guided flip rate is 0.00484 (95% CI $[0.00453, 0.00516]$), and the reference top-1 flip rate is 0.00622 (95% CI $[0.00583, 0.00661]$). Sensitivity is concentrated early in refinement: guided flip rates are 1.08%, 0.38%, and 0.14% in the early, middle, and late portions, respectively. Thus, changing only the reference-side refinement time measurably perturbs the local reference-relative signal, while the large majority of selected tokens remain unchanged. This diagnostic characterizes state sensitivity and does not establish clinical superiority of matched-state decoding.

D.7 GREEN ERROR DECOMPOSITION

We evaluate the saved canonical seed-0 no-reference and TARING reports against their corresponding reference reports using GREEN (Ostmeier et al., 2024). We use the `StanfordAIMI/GREEN-radllama2-7b` checkpoint with the authors' released implementa-

Metric	Overall	Early	Middle	Late
Guided flip rate	0.0048	0.0108	0.0038	0.0014
Reference top-1 flip rate	0.0062	0.0126	0.0054	0.0018
Mean JS divergence	0.000053	0.000090	0.000050	0.000024
Mean $\mathrm{osc}(\delta)$	0.2284	0.2102	0.2348	0.2325
Exact bound satisfied	0.9931	0.9851	0.9946	0.9975
Strong bound satisfied	0.9403	0.8608	0.9563	0.9825

Table 10: Post-hoc reference-state mismatch diagnostic. The canonical trajectory and current-study branch are fixed, while only the reference-side refinement time is shifted by one valid step.

Endpoint	No ref.	TARING	Δ	95% CI
GREEN	0.2866	0.2792	-0.0073	[-0.0147, -0.0003]
Significant errors	5.2490	5.3407	+0.0916	[0.0121, 0.1673]
False findings	2.1256	2.3195	+0.1939	[0.1360, 0.2530]
Missing findings	2.3237	2.1824	-0.1413	[-0.1907, -0.0919]
Location errors	0.3237	0.3076	-.0161	[-0.0387, 0.0076]
Severity errors	0.3971	0.4434	+0.0462	[0.0180, 0.0746]
Matched findings	1.9559	1.9194	-0.0365	[-0.0802, 0.0043]

Table 11: Post-hoc GREEN evaluation of the canonical reports. Differences are TARING minus no-reference. Confidence intervals are paired subject-cluster bootstrap intervals. GREEN categories are automated report/reference outputs rather than radiologist-adjudicated or image-grounded clinical errors.

tion.[1] This post-hoc automated report/reference analysis does not alter the decoding protocol. Differences are TARING minus no-reference.

Table 11 shows fewer missing-finding errors but more false-finding and severity errors under TARING, alongside a lower overall GREEN score. The confidence intervals for location errors and matched findings include zero. These seed-0 automated results indicate an omission–commission trade-off rather than uniform improvement across report-quality measures.

E Qualitative Reference-Guidance Comparison

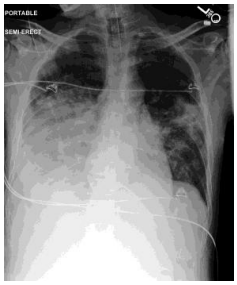
Figure 4 complements the aggregate results with a report-level comparison of no-reference decoding and TARING. The example illustrates differences in generated content rather than an estimate of the average effect.

F Reproducibility and Provenance

F.1 Limits of end-to-end reconstruction

The inference comparison is reproducible from its frozen inputs, but complete raw-data-to-checkpoint reconstruction is not. The archival record does not recover the original row-level filtering ledger, all report-normalization details, the exact tokenizer/action snapshot used during training, evidence-cache precision history, or the immutable launch configuration for the source checkpoint. These gaps do not alter the paired inference rows reported here, but they prevent a claim that an independent party can recreate the checkpoint from the raw MIMIC-CXR release using only the archived package.

[1] https://github.com/Stanford-AIMI/GREEN



Baseline

finding :
the et tube tip is 45 cm above the carina.
the right ij central line tip is at the level of mid svc.
the bilateral pulmonary opacities are present.
there is no pulmonary edema .
there are small bilateral pleural effusions .
there is no pneumothorax .
the heart and the hilar contours are normal .
impression : the et tube is in appropriate position . the bilateral pulmonary opacities are present .

Target report

findings :
diffuse bilateral pulmonary opacities with prominent right and probably small left effusion are present .
the endotracheal tube is in position .

Taring

impression :
severe multifocal opacities are present on the right greater than the left .
severe consolidation is present in the right lower lobes .
moderate right pleural effusion is present.
there is a small left-sided pleural effusion .
the heart is normal in size .
endotracheal tube is in standard placement .
esophageal drainage tube passes into the esophagus and out of view in the stomach . there is no pneumothorax .

Figure 4: Qualitative comparison of no-reference decoding and TARING. The example shows the chest radiograph, reference report, no-reference output, and TARING output. Colors highlight corresponding clinical content across reports. The example illustrates how reference-relative decoding can alter report content; it does not establish the image-grounded correctness of individual generated statements.

G Extended Related Work

G.1 Radiology report generation

Radiology report generation has progressed through several complementary ways of strengthening image–text representations and the evidence available to the decoder. Memory-based and cross-modal architectures explicitly retain or align report-relevant information (Chen et al., 2020; 2021), while later systems adapt large pretrained language models to radiographic inputs (Wang et al., 2023b). Grounded multimodal models such as MAIRA-2 associate report content with image regions (Bannur et al., 2024). Other approaches introduce retrieval or explicit correction mechanisms: DART retrieves disease-aligned reports and subsequently re-aligns generated content with the image (Park et al., 2025), whereas RA-RRG retrieves clinically relevant key phrases that are supplied to the generator (Park et al., 2026).

These methods primarily change the learned representation, training objective, auxiliary evidence, or retrieval pathway available to report generation. TARING is complementary to that line of work: it does not retrain the frozen generator or retrieve a patient-specific report at inference time. Instead, it changes content selection by evaluating the same evolving masked-report state under the current-study evidence and a deterministic training-derived evidence reference.

G.2 Masked and non-autoregressive generation

Non-autoregressive text generation replaces strictly left-to-right decoding with parallel or iterative refinement. Mask-Predict established iterative conditional masked generation in which multiple unresolved positions are updated across refinement rounds (Ghazvininejad et al., 2019). Masked diffusion language models and LLaDA further develop denoising-style language generation with bidirectional context and iterative masked-state transitions (Sahoo et al., 2024; Nie et al., 2025). In radiology report generation, D^2-Net and ControlDiff explore diffusion-based generation (Jin et al.,

2024b; Tian et al., 2024); more recent work includes one-step block diffusion for efficient chest X-ray report generation in ECHO (Chen et al., 2026) and clinically anchored masked diffusion with confidence-based rewriting in AnchorDiff (Yu et al., 2026).

G.3 Inference-time guidance and contrastive decoding

TARING also relates to a broad family of decoding-time methods that combine predictions from multiple conditions or models. Classifier-free guidance uses conditional and unconditional denoiser predictions in an affine guidance rule (Ho & Salimans, 2022). DExperts combines base, expert, and anti-expert language-model scores for controllable generation (Liu et al., 2021), while Contrastive Decoding contrasts an expert and amateur model subject to a plausibility constraint (Li et al., 2023). Context-aware decoding instead compares one language model with and without external context, amplifying context-supported predictions without additional training (Shi et al., 2024). AdaCAD adapts the contrastive weight using the Jensen–Shannon divergence between context-conditioned and context-free predictions (Wang et al., 2025a). In multimodal generation, Visual Contrastive Decoding contrasts predictions induced by original and distorted visual inputs (Leng et al., 2024). M3ID contrasts image-conditioned and image-free predictions to strengthen visual conditioning during autoregressive generation (Favero et al., 2024).

Guidance has also been adapted to discrete and masked diffusion models (Schiff et al., 2025; Rojas et al., 2026). A-CFG constructs a dynamic alternative input by temporarily re-masking low-confidence tokens (Li et al., 2025). TARING instead preserves the partial report within each current/reference comparison and changes only the evidence context. Radiology-specific contrastive decoders provide still closer task-level precedents. CCD combines a radiology MLLM with structured signals from task-specific clinical expert models (Zhang et al., 2026), whereas CWCD performs category-wise report generation by contrasting normal and masked radiographs with category-specific visual prompts (Srivastava et al., 2026).

Positioning of TARING. Logit differencing, affine guidance, and probability-ratio reweighting have the precedents discussed above. The KL-regularized update uses the established entropic mirror-ascent or exponentiated-gradient construction (Beck & Teboulle, 2003; Kivinen & Warmuth, 1997). The paper-specific construction is their use with a deterministic training-mean evidence-space reference inside iterative masked radiology report generation. At each content decision, the current and reference branches share the partial report, refinement time, queried coordinate, report support, and valid action set; only the evidence context changes. TARING additionally keeps report-support prediction current-study-only and restricts reference-relative intervention to content actions.