ProtoMedAgent: Supplementary Material

Alvaro Lopez Pellicer[1,†], Plamen Angelov[1], Marwan Bukhari[2], Yi Li[1], Eduardo Soares[3], Jemma Kerns[2]
[1]School of Computing and Communications, Lancaster University
[2]Lancaster Medical School, Lancaster University [3]PUC-Rio, Puc-Behring Institute for AI

1. Evaluation Metrics

This section provides a detailed definition of the evaluation metrics used in the results of our report.

Comparison Set Faithfulness (CSF). CSF measures whether the model's case–prototype comparison claims are consistent with evidence-derived differences [1]. For each query case and retrieved prototype, we compute deterministic reference partitions over evidence items Δ, where each item is either a visual assertion ID or a tabular factor-field ID. From the generated report, we extract the predicted partitions $\widehat{\Delta}$ (items claimed as shared, query-only, or prototype-only) using the report schema fields. We then compute:

$$\text{CSF-Precision} = \frac{|\widehat{\Delta}_{\text{diff}} \cap \Delta_{\text{diff}}|}{|\widehat{\Delta}_{\text{diff}}|} \tag{1}$$

$$\text{CSF-Recall} = \frac{|\widehat{\Delta}_{\text{diff}} \cap \Delta_{\text{diff}}|}{|\Delta_{\text{diff}}|} \tag{2}$$

$$\text{CSF-F1} = \frac{2\,\text{CSF-Precision}\,\text{CSF-Recall}}{\text{CSF-Precision} + \text{CSF-Recall}}, \tag{3}$$

where $\Delta_{\text{diff}} = \Delta_{\text{q-only}} \cup \Delta_{\text{p-only}}$ and $\widehat{\Delta}_{\text{diff}} = \widehat{\Delta}_{\text{q-only}} \cup \widehat{\Delta}_{\text{p-only}}$.

To summarize three-way assignment quality, we compute a weighted accuracy over the partition labels at the item level:

$$\text{CSF-WA} = \frac{\sum_c w_c \cdot \text{Acc}_c}{\sum_c w_c} \tag{4}$$

$$\text{Acc}_c = \frac{\left|\left\{i : \ell(i) = c \wedge \hat{\ell}(i) = c\right\}\right|}{|\{i : \ell(i) = c\}|}, \tag{5}$$

where $\ell(i)$ and $\hat{\ell}(i)$ denote the reference and predicted partition labels for item i, and w_c are class weights. Precision penalizes unsupported difference claims, recall measures coverage of evidence-backed differences, and CSF-WA summarizes overall three-way partition fidelity.

Privacy evaluation. We evaluate privacy at two complementary levels. First, we analyze the *release surface* induced by the export gate by sweeping semantic k-anonymity and ℓ-diversity over the candidate ProtoCard pool. Second, the main paper reports *artifact-level* attacks on the final exported evidence under a shared protocol across methods. In this supplement we therefore focus on the release-surface quantities that support the frontier in the main paper: weighted evidence utility, visible-card rate, redaction rate, and linkage exposure on the released ProtoCard surface.

[†]Corresponding author. Email: a.lopezpellicer@lancaster.ac.uk

Table 1. Detailed release-surface behavior at the selected operating point $(k, \ell) = (5, 2)$.

Measure	**Value**	**Scope**
Cited evidence utility over held-out reports	0.369	832 reports
Similarity-weighted utility on prototype release surface	0.673	54 candidate cards
Visible-card rate	0.593	54 candidate cards
Redaction rate	0.407	54 candidate cards
Top-1 linkage on released ProtoCards	0.167	54 candidate cards

Table 2. Compressed privacy–utility sweep over semantic k-anonymity and ℓ-diversity. Repeated regimes are merged to emphasize the active control variable.

k	ℓ	Utility	Visible	Linkage	Interpretation
3	1	0.882	0.852	0.312	loosest release
5	1	0.854	0.796	0.308	smoother trade-off
7	1	0.820	0.759	0.250	more restrictive
9	1	0.804	0.722	0.182	strongest k-only point
3,5,7,9	2	0.673	0.593	0.167	identical regime across all tested k
3,5,7,9	3	0.000	0.000	0.000	no shareable surface

Membership-Inference Accuracy (MIA). Let o_i denote a released artifact associated with source case i, and let $m_i \in \{0, 1\}$ indicate whether that source case belongs to the private training/prototype corpus ($m_i = 1$) or not ($m_i = 0$). Given an attack model a_{mia}, we report

$$\text{MIA} = \frac{1}{N_{\text{mia}}} \sum_{i=1}^{N_{\text{mia}}} \mathbf{1}[a_{\text{mia}}(o_i) = m_i] \,.$$

MIA measures whether the released artifact leaks the membership status of an individual case. It is relevant here because ProtoMedAgent explicitly seeks to reduce artifact-level membership leakage from the exported evidence surface and the final generated outputs.

Attribute-Inference Accuracy (AIA). Let $s_i \in \mathcal{S}$ denote a sensitive attribute of source case i, and let z_i denote any allowed auxiliary side information available to the attacker. If the attack model predicts $\hat{s}_i = a_{\text{aia}}(o_i, z_i)$, then we define

$$\text{AIA} = \frac{1}{N_{\text{aia}}} \sum_{i=1}^{N_{\text{aia}}} \mathbf{1}[\hat{s}_i = s_i] \,.$$

AIA measures how accurately an adversary can recover sensitive attributes from the released artifact, possibly combined with auxiliary information. It is especially relevant here because the semantic privacy gate is designed to limit attribute disclosure, which is precisely the threat probed by AIA.

Top-1 Linkage Success Rate (Link.). Let o_i be a released artifact, let $\mathcal{C}_i$ be the adversary's candidate set of identified records, and let $g(o_i, c)$ be a matching score between artifact o_i and candidate record c. The attacker returns the single best match

$$\hat{c}_i = \arg\max_{c \in \mathcal{C}_i} g(o_i, c).$$

Writing $c_i^\star$ for the true source record of o_i, the top-1 linkage success rate is

$$\text{Link.} \;\equiv\; \text{Link@1} = \frac{1}{N_{\text{link}}} \sum_{i=1}^{N_{\text{link}}} \mathbf{1}[\hat{c}_i = c_i^\star] \,.$$

This metric measures exact re-identification risk: whether the attacker's highest-scoring candidate is the correct source person/record. It is directly relevant here because the paper's k-anonymity gate is intended to make exact record linkage ambiguous and therefore suppress successful top-1 matching.

1.1. Privacy analysis and operating-point detail

The privacy gate is fit on a training population of 2,662 quantized records. The selected operating point $(k, \ell) = (5, 2)$ is applied to a prototype-support release surface containing 54 candidate support records, of which 32 remain visible and 22 are explicitly redacted.

Two patterns govern the frontier. First, under $\ell = 1$, increasing k yields the expected smooth privacy–utility trade-off: evidence utility decreases gradually as linkage exposure falls. Second, imposing $\ell = 2$ is the decisive step for this release schema: once diversity is enforced, all tested values $k \in \{3, 5, 7, 9\}$ collapse to the same operating regime. At $\ell = 3$, the visible surface collapses entirely. The resulting picture is therefore not that privacy monotonically improves with every stronger setting, but that the diversity constraint is the binding variable on this release surface.

These release-surface diagnostics should be read together with the artifact-level attacks in the main paper. The supplement clarifies *why* the frontier has the shape shown in Fig. 2, while the main paper's cross-method attack table evaluates the residual disclosure that remains once the selected release surface is turned into final exported evidence.

2. Proofs of Theoretical Guarantees

This section provides the formal proofs for the theoretical guarantees established in Section 3 of the main text. We formalize the generative reporting process as a constrained Markov chain and leverage set-theoretic logic and probability theory to bound the test-time behavior of the Large Language Model (LLM).

2.1. Proof of Theorem 1: Bounded Generative Faithfulness

Definition 1 (Atomic Claim Set). Let $\phi : \mathcal{R} \to 2^{\mathcal{S}}$ be an extraction function that maps a natural language report $R \in \mathcal{R}$ to a set of discrete, atomic semantic claims $S_c \subset \mathcal{S}$.

Definition 2 (Strict Barrier Critic). Let $\mathcal{C}_{E(x)} = \{R \in \mathcal{R} \mid \phi(R) \subseteq E(x)\}$ define the admissible constraint set of reports fully grounded in the evidence state $E(x)$. A critic energy function $E_{\text{critic}}(R)$ acts as a strict barrier if it assigns an infinite penalty to any report outside this set:

$$E_{\text{critic}}(R) = \begin{cases} 0, & \text{if } R \in \mathcal{C}_{E(x)} \\ \infty, & \text{otherwise} \end{cases}$$

Proof. The test-time optimization loop operates over the grounded evidence state $E(x)$, which is completely defined by the deterministic visual and tabular differentials: $\Delta_j = (A(x) \cap A_j, A(x) \setminus A_j, A_j \setminus A(x), \Delta_j^{\text{tab}})$.

Let R^* be the accepted report returned by the Scribe-Critic search algorithm. By the definition of the algorithm's termination criteria, the report must satisfy $E_{\text{critic}}(R^*) \leq \epsilon$. For a strictly deterministic barrier critic ($\epsilon = 0$), this implies $E_{\text{critic}}(R^*) = 0$.

Following Definition 2, $E_{\text{critic}}(R^*) = 0 \implies R^* \in \mathcal{C}_{E(x)} \implies \phi(R^*) \subseteq E(x)$. Because the comparative evidence state $E(x)$ is strictly bounded by Δ_j, it holds that the atomic claims within the report are a strict subset of the pre-computed set-theoretic differentials. Therefore, the hypothesis space of the accepted report is mathematically restricted to the neuro-symbolic bottleneck, and the probability of the LLM introducing an ungrounded hallucinated comparison relative to this bottleneck is exactly 0. ■

2.2. Proof of Theorem 2: Information-Theoretic Privacy Bound

To prove the privacy bounds, we model the framework as a Markov chain where information flows from the raw training prototype to the final generated report. We assume an adversary $\mathcal{A}$ whose prior belief matches the empirical distribution of the sanitized equivalence class (i.e., no auxiliary background knowledge linking identities to attributes).

Proof. Let the flow of data be represented by the following sequence of random variables forming a Markov chain:

$$S \to P \to \Sigma \to E \to R^*$$

where S is a sensitive tabular attribute contained within the raw training prototype P, $\Sigma = \Psi(P)$ is the resulting non-differentiable sanitized semantic signature, E is the grounded prompt context, and R^* is the final generated report. Let lowercase letters denote their respective realizations (e.g., $S = s$, $R^* = r$).

Let $\mathcal{A}$ attempt to guess the sensitive attribute S from the final report $R^* = r$. We seek to bound the maximum posterior probability $\max_s \Pr(S = s \mid R^* = r)$.

By the Law of Total Probability, we marginalize over all possible released signatures σ:

$$\Pr(S = s \mid R^* = r) = \sum_{\sigma} \Pr(S = s \mid \sigma, R^* = r) \Pr(\sigma \mid R^* = r)$$

Because the system enforces a strict Markov chain, the raw attribute S and the final report R^* are conditionally independent given the bottleneck signature Σ. Therefore, $\Pr(S = s \mid \sigma, R^* = r) = \Pr(S = s \mid \sigma)$. Substituting this yields:

$$\Pr(S = s \mid R^* = r) = \sum_{\sigma} \Pr(S = s \mid \sigma) \Pr(\sigma \mid R^* = r)$$

The privacy gate Ψ enforces strict ℓ-diversity on the visible signature σ, which guarantees that the empirical frequency of the most common sensitive attribute within any released equivalence class is at most $1/\ell$. Given $\mathcal{A}$ possesses no auxiliary knowledge to skew this distribution, $\max_s \Pr(S = s \mid \sigma) \leq \frac{1}{\ell}$. Applying this bound:

$$\Pr(S = s \mid R^* = r) \leq \sum_{\sigma} \left(\frac{1}{\ell}\right) \Pr(\sigma \mid R^* = r) = \frac{1}{\ell} \sum_{\sigma} \Pr(\sigma \mid R^* = r)$$

Since $\sum_{\sigma} \Pr(\sigma \mid R^* = r) = 1$, we conclude:

$$\max_s \Pr(S = s \mid R^* = r) \leq \frac{1}{\ell}$$

Similarly, the k-anonymity constraint ensures that any released signature σ corresponds to an equivalence class of at least k distinct prototype records. By the identical Markov property argument, the posterior probability of an adversary successfully executing an exact record linkage attack (isolating a unique individual i from R^*) is bounded by $\frac{1}{k}$.

Because the generative model never observes P directly, the artifact R^* is mathematically secured against attribute inference and linkage attacks beyond these established thresholds. ■

3. Extended clinical report gallery

The main paper shows one final report per class. This section extends that view with additional same-class studies, restricted-comparison behavior, and harder threshold-sensitive examples. Each panel is written in the same final report voice as the main figure so that the supplement reads as an expanded gallery of outputs rather than as a case audit.

3.1. Additional routine reports

The cases below extend the main figure with routine reports that stay clinically direct while varying in density reserve, structured risk burden, and comparison difficulty.

Strongly preserved normal lumbar study

Query patient: 86-year-old woman; lumbar DXA image plus tabular clinical record.

ProtoMedX output: NORMAL, $T = 2.3$, Confidence 99.2%; p1=`normal_proto_1` [N], p2/p3 = N/N.

ProtoMedAgent report: Lumbar bone density is well preserved and clearly within normal limits. The scan follows a strong preserved-density pattern, with the lower lumbar vertebrae standing out as denser than the upper spine and no convincing sign of generalized mineral loss. Mild degenerative change and small overlying artifact may accentuate that distal preservation, but they do not change the overall normal impression. Rheumatoid arthritis is the main structured risk factor in an otherwise light-burden profile. The study aligns with the high-preservation normal reference rather than any boundary osteopenic prototype.

Early osteopenia with low non-age risk burden

Query patient: 95-year-old woman; lumbar DXA image plus tabular clinical record.

ProtoMedX output: OSTEOPENIA, $T = -1.0$, Confidence 94.2%; p1=`osteopenia_proto_0` [Op], p2/p3 = Op/Op.

ProtoMedAgent report: Low bone mass consistent with early osteopenia is present. The estimate sits at the osteopenic threshold, and the scan no longer shows the smooth preserved-density appearance expected of a normal lumbar spine. Some lower-lumbar reserve remains, but the overall impression is one of mild distributed attenuation rather than advanced depletion. Advanced age keeps the finding clinically relevant even without additional major structured risk factors. The nearest prototype comparison is an early-loss osteopenic reference rather than a truly preserved normal card.

Advanced osteoporosis with prior fragility fracture

Query patient: 84-year-old woman; lumbar DXA image plus tabular clinical record.

ProtoMedX output: OSTEOPOROSIS, $T = -4.0$, Confidence 97.7%; p1=`osteoporosis_proto_2` [Os], p2/p3 = Op/Op.

ProtoMedAgent report: Bone density is in the osteoporotic range with advanced mineral loss. The estimate lies far below the osteopenic boundary, and the lumbar spine reads as diffusely depleted rather than residually borderline. Previous fragility fracture places the patient in a high-risk context for future fracture. The study is best matched to a severe low-density osteoporotic reference and is clearly separated from any early-loss osteopenic prototype.

3.2. Edge-Case reports

These reports show that the final report style remains usable when only part of the comparison surface is visible or when the density evidence sits close to a diagnostic threshold.

Near-threshold normal study under restricted comparison release

Query patient: 88-year-old woman; lumbar DXA image plus tabular clinical record.

ProtoMedX output: NORMAL, $T = -0.2$, Confidence 86.5%; p1=`normal_proto_1` [N], p2/p3 = N/Op.

ProtoMedAgent report: Lumbar bone density remains within normal limits. The estimate lies close to the normal/osteopenia boundary, but the visible comparison surface still favors preserved density over early mineral loss. The study retains a preserved lower-lumbar-greater-than-upper pattern rather than the flatter, weaker appearance of an osteopenic reference. Previous fragility fracture keeps the case clinically important despite the normal density impression, so the report remains normal in class while using careful near-threshold wording.

Borderline osteopenia with fracture and rheumatoid arthritis

Query patient: 74-year-old woman; lumbar DXA image plus tabular clinical record.

ProtoMedX output: OSTEOPENIA, $T = -0.6$, Confidence 61.2%; p1=`osteopenia_proto_0` [Op], p2/p3 = Op/N.

ProtoMedAgent report: Early low bone mass is favored despite a near-normal estimated T-score. The calibrated interval crosses the normal/osteopenia boundary, and the scan reads closer to an early mineral-loss pattern than to a convincingly preserved normal spine. Previous fragility fracture and rheumatoid arthritis make this borderline presentation clinically significant. Comparison with lower-boundary osteopenic references therefore supports cautious osteopenic wording and follow-up rather than dismissal as preserved density.

Low-density study straddling osteopenia and osteoporosis

Query patient: 85-year-old woman; lumbar DXA image plus tabular clinical record.

ProtoMedX output: OSTEOPOROSIS, $T = -2.0$, Confidence 61.8%; p1=`osteoporosis_proto_2` [Os], p2/p3 = Op/Op.

ProtoMedAgent report: Low bone density is present, but the final impression remains clinically measured. Image features raise concern for osteoporosis, yet the calibrated estimate remains in the osteopenic range and spans the osteopenia/osteoporosis threshold. The visible comparison neighborhood is weaker than preserved or early-loss normal references, but it still does not reproduce the most depleted severe osteoporotic patterns across the entire lumbar column. With age as the main risk driver and no stronger structured burden, correlation with prior DXA studies or interval follow-up is appropriate before labeling the scan unequivocal advanced osteoporosis.

4. Prototype atlas (ProtoCards)

The main paper highlights three reference ProtoCards. The atlas below organizes the full learned library by diagnostic class and makes explicit what each prototype contributes: the density regime it occupies, the scan pattern that tends to recur, the surrounding clinical context, and the caveat that most often changes interpretation.

The atlas below enumerates the full learned prototype library in the same narrated style as the main figure. Each card is written as a usable clinical reference: what the lumbar scan tends to look like, which clinical backdrop tends to recur, and why that prototype matters when it is contrasted against neighboring cards.

4.1. Normal reference prototypes

The normal atlas ranges from strongly preserved spines to cautious near-threshold normal studies in which fracture history or inflammatory context can coexist with preserved density.

Heterogeneous mild preserved-density normal prototype

Reference prototype: `normal_proto_0`.
Typical study: density remains preserved overall, but the lumbar column is visibly uneven from vertebra to vertebra. Lower lumbar vertebrae often read slightly denser than the upper spine, while unreadable labels, incomplete coverage, and mild lower-lumbar degenerative change repeatedly force measured wording instead of an emphatically pristine normal report.

Robust preserved-density normal prototype

Reference prototype: `normal_proto_1`.
Typical study: the lower lumbar spine remains distinctly denser than the upper lumbar vertebrae and the whole column reads comfortably normal across a broad preserved range. Prior fragility fracture can coexist with this prototype, but unlike the boundary normal or osteopenic references the preserved structure remains convincing even when mild sclerosis or overlay exaggerates the distal vertebrae.

Mid-range preserved-density normal prototype

Reference prototype: `normal_proto_2`.
Typical study: density is clearly preserved but without the excess reserve of the strongest normal anchor. This prototype fits quieter normal scans, usually with lighter lifestyle or body-composition risk context and fewer recurring artifact cues than the more heterogeneous normal references.

Boundary-side normal prototype with fracture context

Reference prototype: `normal_proto_3`.
Typical study: the scan stays on the preserved side of the normal/osteopenia boundary, but the surrounding clinical context is heavier than in the tighter normal references. Prior fragility fracture recurs often enough that this prototype is useful when clinical risk is elevated even though the lumbar density pattern still reads normal.

Tight near-threshold normal prototype with inflammatory risk

Reference prototype: `normal_proto_4`.
Typical study: density remains technically normal, but only with a small reserve above the osteopenic boundary. Low BMI, fracture history, or inflammatory context often make this a cautious preserved-density report rather than a strongly reassuring one.

Compact mild-normal prototype with fracture history

Reference prototype: `normal_proto_5`.
Typical study: a mild preserved-density pattern centered close to the boundary of normality, without the stronger reserve of the high-preservation references. Everyday fracture or lifestyle context may still accompany this prototype, so it supports restrained normal wording rather than an emphatic statement of high skeletal reserve.

4.2. Osteopenic reference prototypes

The osteopenic atlas is organized around lower- versus upper-boundary low bone mass, with additional separation between fracture-linked, lifestyle-linked, and especially cohesive early-loss patterns.

Early mineral-loss osteopenic prototype with vertebral heterogeneity

Reference prototype: `osteopenia_proto_0`.
Typical study: lower lumbar vertebrae may retain relatively greater density, but preservation becomes uneven across the visible lumbar column and the upper lumbar spine shows reduced density relative to the lower vertebrae. Rotation, limited coverage, unreadable vertebral labels, degenerative change, and focal overlay recur often enough that the report should sound measured, yet the prototype still reads clearly weaker than a normal spine.

Tight low-burden lower-boundary osteopenia

Reference prototype: `osteopenia_proto_1`.
Typical study: real but subtle low bone mass clustered near the lower osteopenic boundary, with little dramatic visual flourish. This prototype is useful when the class decision comes from a consistent borderline deficit rather than a memorable structural cue, often in patients with leaner body habitus or prior fracture.

Broader fracture-linked lower-boundary osteopenia

Reference prototype: `osteopenia_proto_2`.
Typical study: a lower-boundary osteopenic scan that stays close to normal on density alone, but appears in a clinically heavier fracture-linked context. The visual family is mixed rather than tidy, so this prototype supports osteopenic wording for heterogeneous early-loss presentations rather than textbook boundary examples.

Lifestyle-linked lower-boundary osteopenia

Reference prototype: `osteopenia_proto_3`.
Typical study: mild lower-boundary osteopenia accompanied more often by lifestyle-associated risk, especially heavier alcohol exposure, than by recurrent fracture. The density loss is gentle and the image family broad, making this a useful prototype for clinically lighter osteopenia that still deserves recognition as low bone mass.

Upper-boundary osteopenia near preserved density

Reference prototype: `osteopenia_proto_4`.
Typical study: a subtly weakened lumbar spine that leans toward preserved density more than any other osteopenic reference. This prototype is most helpful when a study nearly passes as normal but still shows enough diffuse weakening to justify osteopenic wording.

Cohesive fracture-associated lower-boundary osteopenia

Reference prototype: `osteopenia_proto_5`.
Typical study: definite early mineral loss with one of the most stable image patterns in the osteopenic class. Previous fragility fracture may recur, but the main value of this prototype is that it gives a crisp visual anchor when a study is clearly osteopenic even though it is not yet osteoporotic.

4.3. Osteoporotic reference prototypes

The osteoporotic atlas separates moderate versus deeper low-density anchors and distinguishes relatively isolated density loss from the highest-burden severe disease.

Moderate low-density osteoporosis with residual lower-lumbar contrast

Reference prototype: `osteoporosis_proto_0`.
Typical study: overall osteoporotic density loss with residual lower-lumbar-greater-than-upper contrast still visible. That apparent preservation must be read cautiously because degenerative change, overlay, and incomplete coverage recur here; this prototype fits scans that are osteoporotic overall but not uniformly depleted across all visible vertebrae.

Broad severe osteoporosis with secondary-cause context

Reference prototype: `osteoporosis_proto_1`.
Typical study: unmistakably low density, but with mixed secondary-cause context and a visually diverse severe support set. This prototype is most useful when advanced skeletal fragility is clear even though no single picture-perfect osteoporotic template dominates the presentation.

Deep low-density osteoporotic prototype

Reference prototype: `osteoporosis_proto_2`.
Typical study: deeper low density that sits farther from the osteopenic boundary than the moderate severe references. The defining feature is depth of mineral loss rather than a dense accompanying risk profile, so the image itself carries most of the diagnostic weight.

High-burden severe osteoporosis prototype

Reference prototype: `osteoporosis_proto_3`.
Typical study: marked low density accompanied by the richest clinical burden in the atlas, including recurrent fracture, smoking, and inflammatory risk. This prototype represents the part of the osteoporotic class in which both the scan and the clinical history point strongly toward advanced skeletal fragility.

Tight deep-low-density osteoporotic subset

Reference prototype: `osteoporosis_proto_4`.
Typical study: a narrow deep-low-density subset in which the T-score band is extremely severe even when other structured risks are relatively sparse. This prototype is the right reference when the scan is unequivocally osteoporotic on density alone and the report should emphasize severity rather than risk accumulation.

Cohesive low-density osteoporosis with prior-fracture motif

Reference prototype: `osteoporosis_proto_5`.
Typical study: consistently depleted lumbar mineralization with recurrent prior fragility fracture across a very cohesive severe support set. This prototype provides the cleanest decisive comparison when a case is clearly beyond early low bone mass and no longer sits in a threshold-sensitive regime.

References

[1] OFM Riaz Rahman Aranya and Kevin Desai. Trace: Temporal radiology with anatomical change explanation for grounded x-ray report generation, 2026. 1