

1

Conventional Caption-Pair Corpus

🔥 Retained for training

Figure 1:

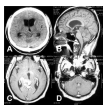
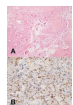


Figure 2:



Caption 1:

CT scan showing a hypointense mass in the pineal region.

Caption 2:

Histopathology shows an anaplastic astrocytic tumor with necrosis.

📄 Source article context (discarded)

A pineal lesion extended to the midbrain and thalamus with ependymal dissemination (Figure 1).

GFAP-positive immunohistochemistry confirmed glioblastoma after subtotal resection (Figure 2).



Body-text evidence discarded



Cross-figure links lost



Limited grounding beyond captions

2

Source-Grounded Recovery

1. Recover missing captions

"No caption found" →  → recovered caption

2. Remove XML tags

"Figure <xref ref-type='fig' rid='1'...>2</xref>" →  → "Figure 2"

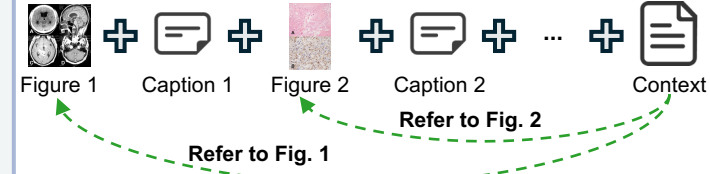
3. Remove duplicate contexts

"Common bile duct measured 0.5 cm ..."
Common bile duct measured 0.5 cm ..."
→  → "Common bile duct measured 0.5 cm ..."

3

Reference-constrained interleaving

1. Construction



2. Refinement

- ①  Figure Numbering Normalization
- ②  Context Coherence Repair
- ③  Minimum-Length Filtering

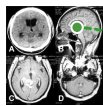
6

PMC-InterCPT

🔥 Interleaved training data

Cleaned interleaved context

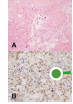
Figure 1:



Caption 1:

CT scan showing a hypointense mass in the pineal region.

Figure 2:



Caption 2:

Histopathology shows an anaplastic astrocytic tumor with necrosis.

GFAP-positive immunohistochemistry confirmed glioblastoma after subtotal resection (Figure 2).



Traceable image-text grounding



Coherent CPT sequences

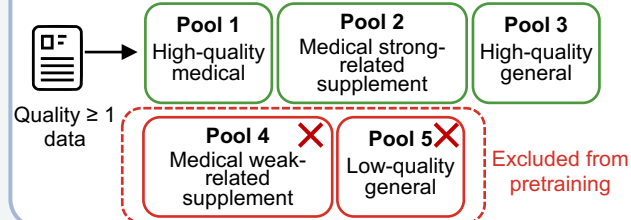


Reduced unsupported pairings

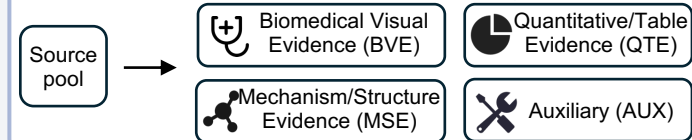
5

Evidence-Aware Allocation

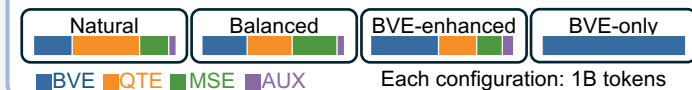
1. Data pool construction



2. Evidence bucket taxonomy



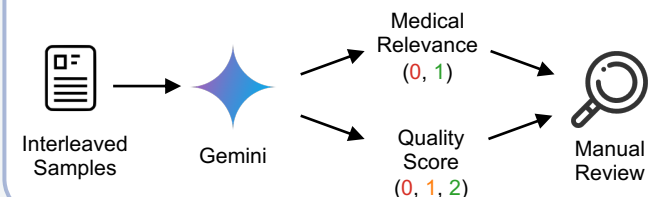
3. Mixture configurations



4

Medical Relevance and Quality Classification

1. LLM annotation



2. Scalable classifier inference

