

Research portfolio

Jin Kim, October 30, 2026

kimjin116@g.ucla.edu
jink-ucla.github.io

Research portfolio

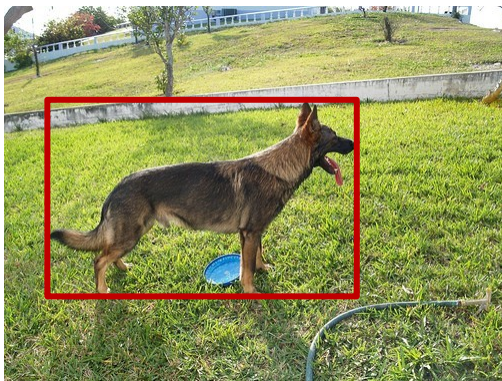
Selected work, 2025–2026

Amazon	CRAFT post-training and image generation	3–4
Med-R	Structured VLM outputs and GRPO	5–6
Phrase grounding	Grounding versus detection on chest X-rays	7
CXR-CurateBench	Controlled curation and grounded evaluation	8–9
JUN	Auditable radiology report validation	10–11
Q-LACE	Finding-specific longitudinal reasoning	12–13
DuEL-Med	Classification-guided report generation	14–15
TRACE	Visual affordance grounding	16–17
OpenManus-SimpleMind	Executable computer vision agents	18–19
Research profile	Education, availability and public links	20

Amazon applied-science internship

Computer Vision – Devices, June–October 2026

Q: *What animal is on the grass?*



Original photo: **dog**



repaint the box



CRAFT edit: **wolf**

Multimodal post-training

Co-developed CRAFT, combining counterfactual augmentation with GRPO-based VLM fine-tuning. Ran LoRA-based SFT and group-relative RL across eight nodes and 64 NVIDIA A100 GPUs in one data-parallel group.

Conditional image generation

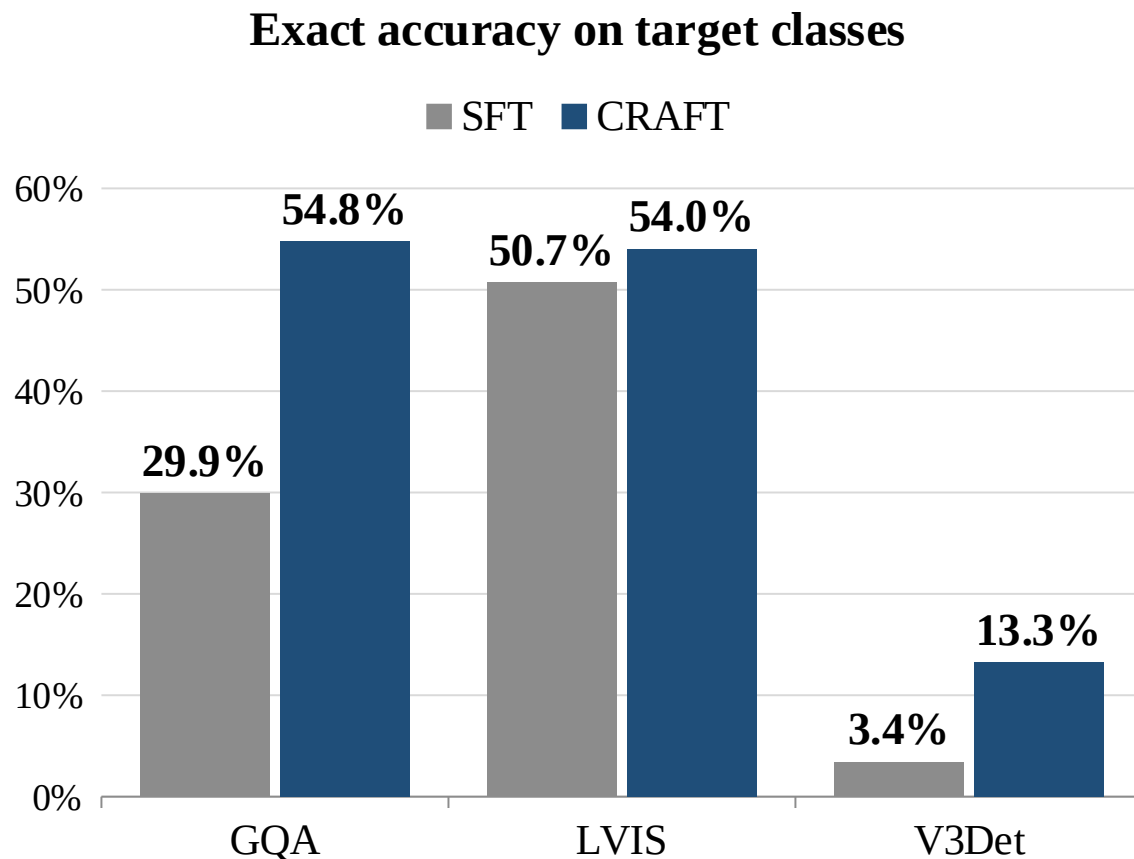
Implemented a single-copy conditional-generation path for a distilled image editor. Evaluated latency-quality tradeoffs across sampling budgets.

Training and evaluation reliability

Contributed to versioned-cache and checkpoint-evaluation integration. Debugged synchronization, activation checkpointing, attention-backend and precision issues.

Amazon applied-science internship

CRAFT versus SFT on target classes with low baseline accuracy



GQA targets

29.9% → 54.8%
+24.9 points

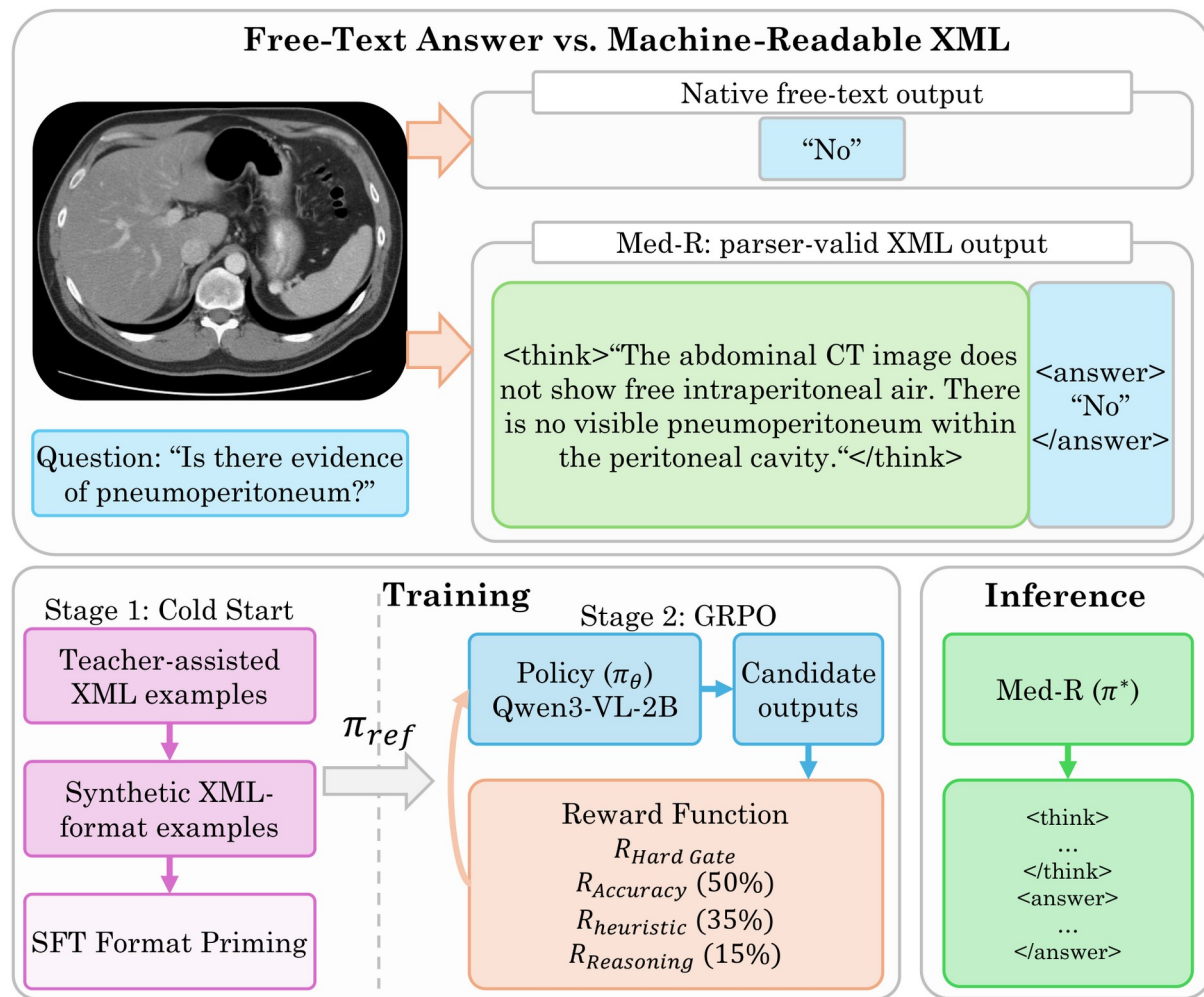
Gains also on LVIS and
V3Det targets.

Paired 95% intervals
exclude zero.

Overall GQA accuracy holds (62.3% vs. 62.4%). Lookalike-name questions drop 5.1 points.

1. Med-R

1.1. Hard-gated training for structured VLM outputs



SFT primes the output format.

GRPO gives malformed outputs zero reward.

Only valid outputs receive task-accuracy rewards.

One native generation pass.
No claim of faithful reasoning.

1. Med-R

1.2. Answer accuracy and format compliance

Native generation: strict accuracy / format compliance (%)

Method	Schema	Training scope	VQA-RAD Strict/Fmt (%)	PathVQA Strict/Fmt (%)	MIMIC Strict/Fmt (%)
SFT	full XML	full	68.20 / 99.62	51.10 / 98.33	60.80 / 99.00
Med-R-AO	answer-only	bounded ablation	61.30 / 94.64	51.85 / 86.76	56.53 / 91.20
Med-R full	full XML	full	69.35 / 99.62	51.28 / 98.75	62.40 / 99.20

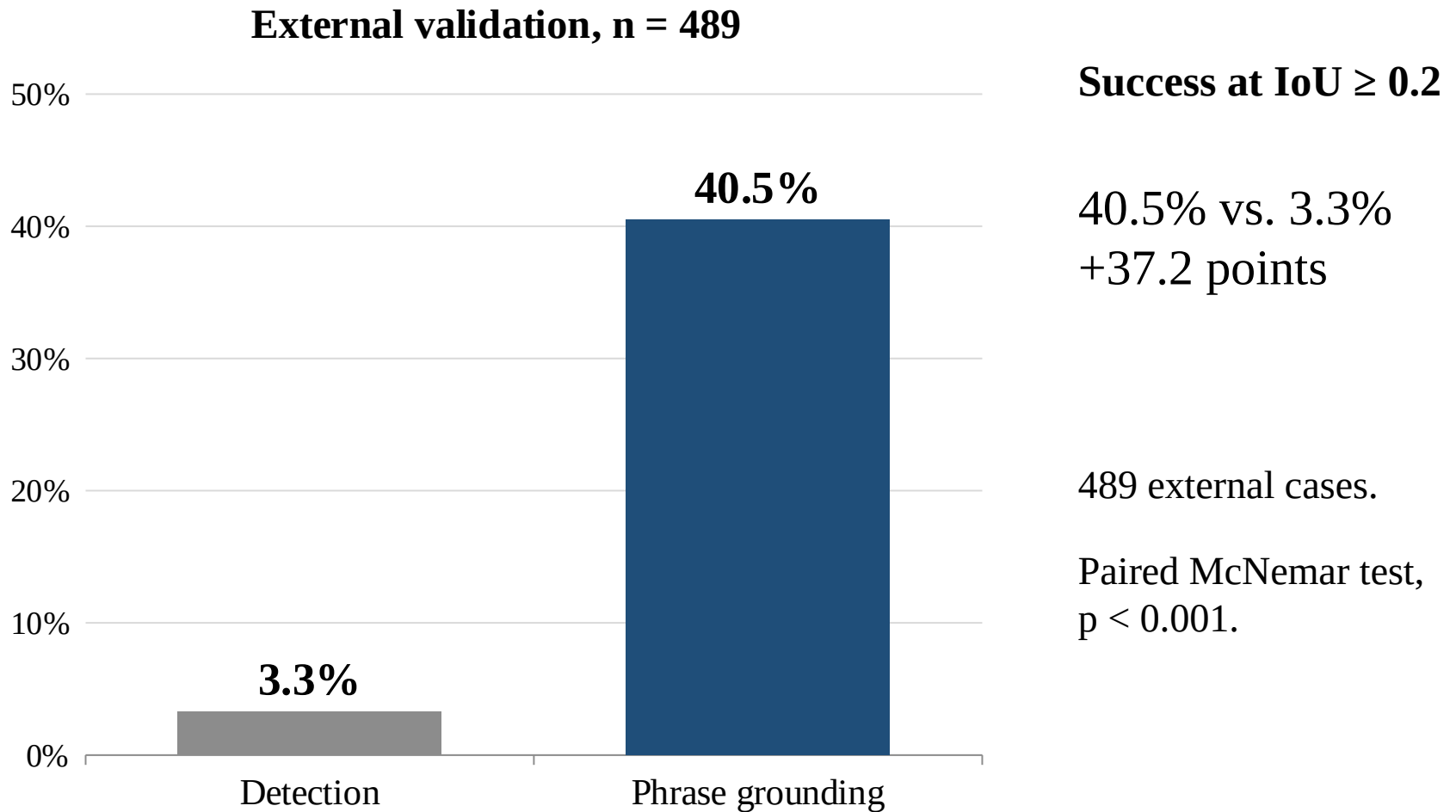
Native generation versus a constrained wrapper

Method	Dataset	Wrapper	Passes	Median latency (s)	Tokens	Peak VRAM (GB)	Strict (%)	Fmt (%)
SFT + constrained	VQA-RAD	yes	3	1.96	72.85	6.33	58.62	100.00
Med-R full native	VQA-RAD	no	1	1.61	70.73	4.05	69.35	99.62
SFT + constrained	PathVQA	yes	3	1.72	65.00	4.87	62.43	100.00
Med-R full native	PathVQA	no	1	1.74	76.99	4.05	51.28	98.75
SFT + constrained	MIMIC-CXR	yes	3	4.91	65.00	9.60	47.40	100.00
Med-R full native	MIMIC-CXR	no	1	1.63	71.39	4.05	62.40	99.20

SFT is a strong baseline. The wrapper guarantees validity and leads on PathVQA accuracy.

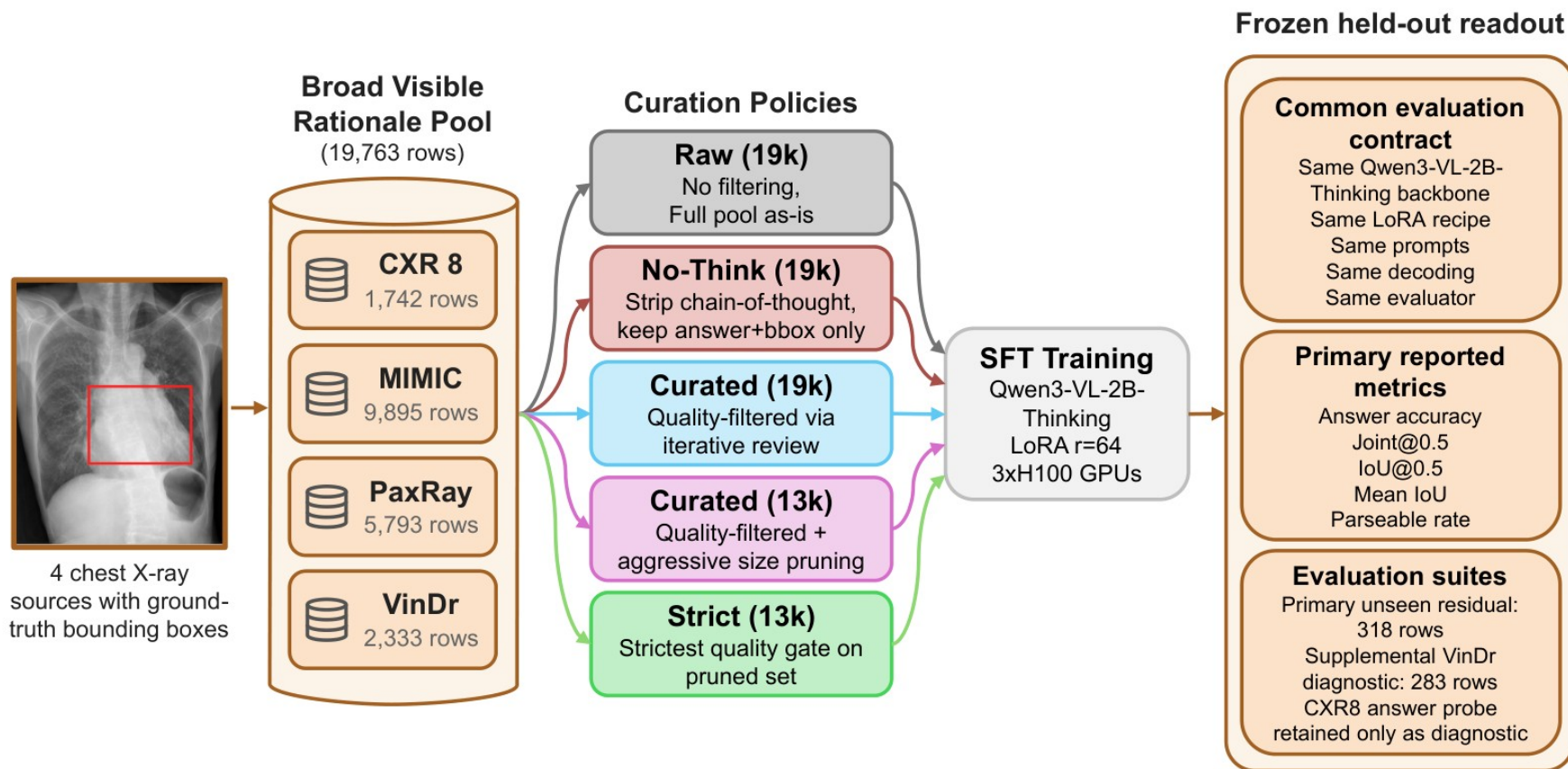
2. Phrase grounding

2.1. Grounding versus detection for conditional chest X-ray localization



3. CXR-CurateBench

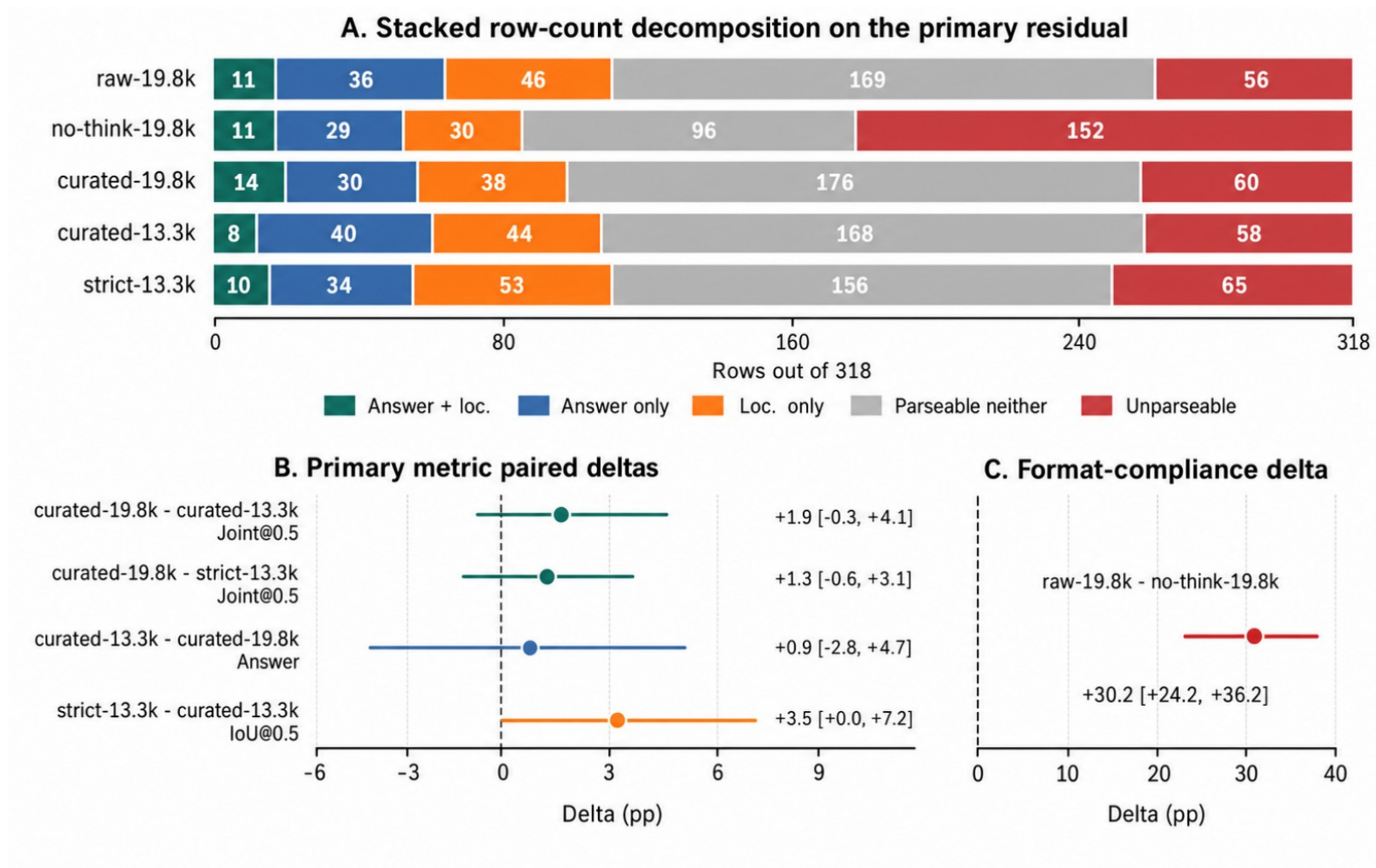
3.1. Controlled supervision for grounded medical VLMs



I compare curation policies while holding the model and evaluation recipe fixed.

3. CXR-CurateBench

3.2. Curation changes distinct failure modes



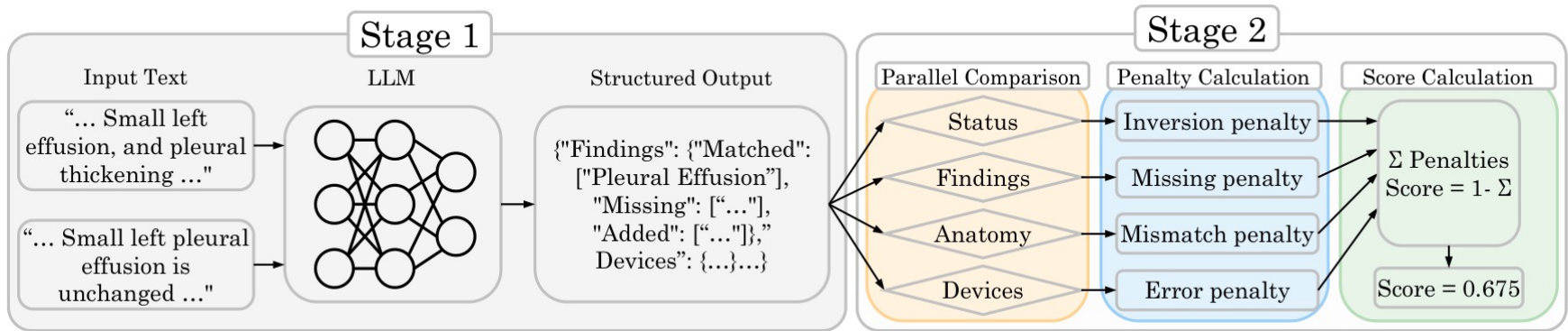
Format improves with rationale retention. Joint-success gains remain uncertain.

Manuscript in preparation (2026). Original Fig. 2, Apr. 28 draft.

4. JUN

4.1. Auditable radiology report validation

Which report discrepancy caused a score deduction?



Neural extraction aligns reports into structured findings.

Deterministic rules expose the discrepancy behind each deduction.

Missed findings can still bypass the scoring rules.

4. JUN

4.2. Factual failures in fluent reports

Constructed stress-test examples

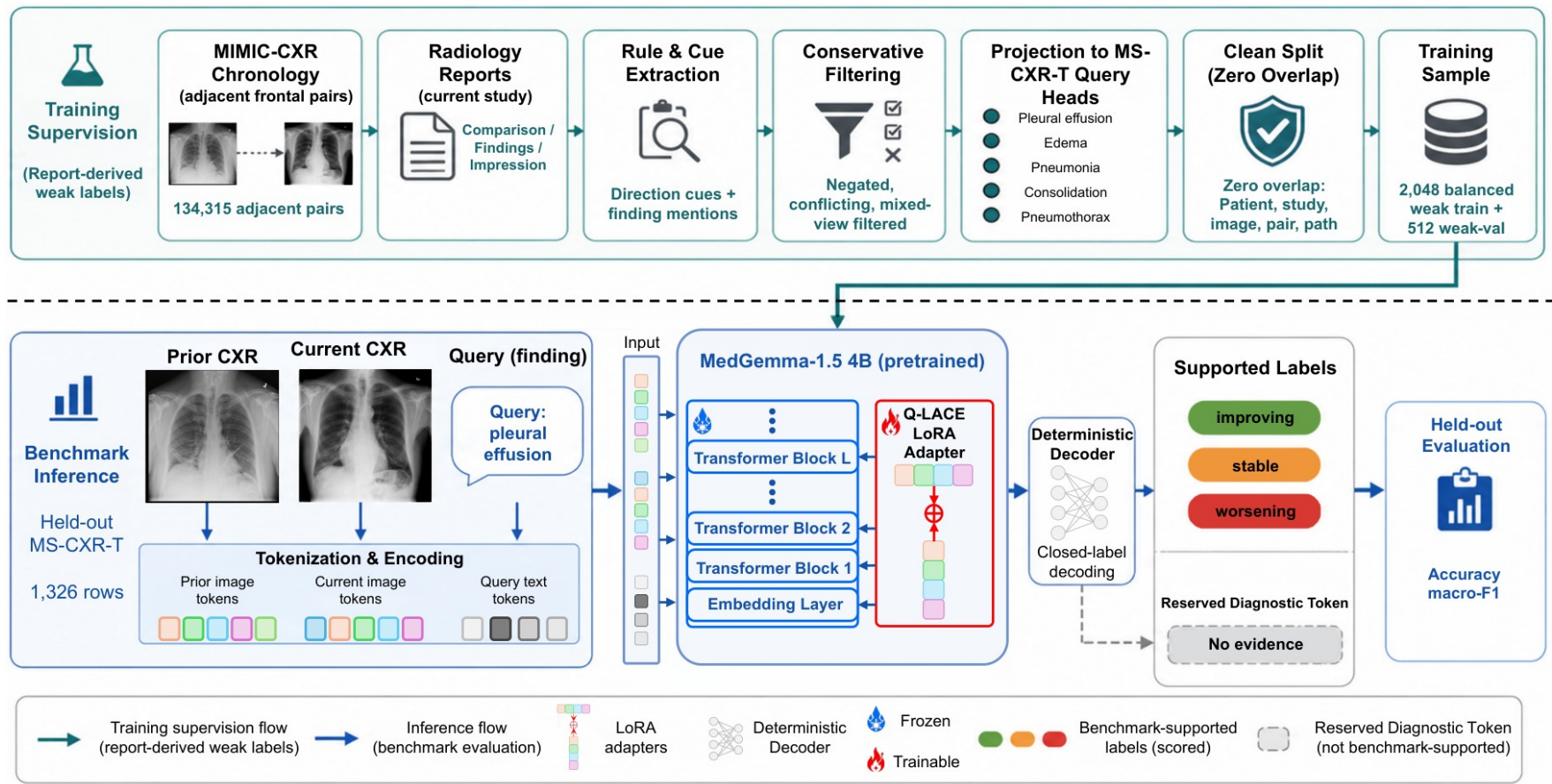
Safety Trojan		Critical Fabrication		Internal Logic Check	
Reference		Reference		Reference	
“No acute infiltrate”	JUN: 0.04	Hyperexpanded lungs consistent with emphysema. No evidence of active disease	JUN: 0.04	...pulmonary parenchyma appears entirely clear ... no evidence of focal consolidation...	JUN: 0.02
	CREPE: 0.28		CREPE: 0.19		CREPE: 0.79
	FineRad: 0.13		FineRad: 0.25		FineRad: 0.50
Candidate		Candidate		Candidate	
“...globular 'water-bottle' configuration... Large pericardial effusion with signs of tamponade...”	GREEN: 1.00	...Chronic emphysematous changes... Large pericardial effusion with signs of tamponade.	GREEN: 1.00	...dense alveolar consolidation ... highly suggestive of acute bacterial pneumonia...	GREEN: 0.75
	RadCliQ: 0.29		RadCliQ: 0.39		RadCliQ: 0.65
	RaTE: 0.30		RaTE: 0.46		RaTE: 0.47

Inspect fabricated findings and internally inconsistent statements.

Reliable scoring still depends on accurate finding extraction.

5. Q-LACE

5.1. Finding-specific longitudinal change extraction



Report-derived weak labels supervise change in a specific finding.
Figure thumbnails illustrate the workflow.

5. Q-LACE

5.2. Query and history controls

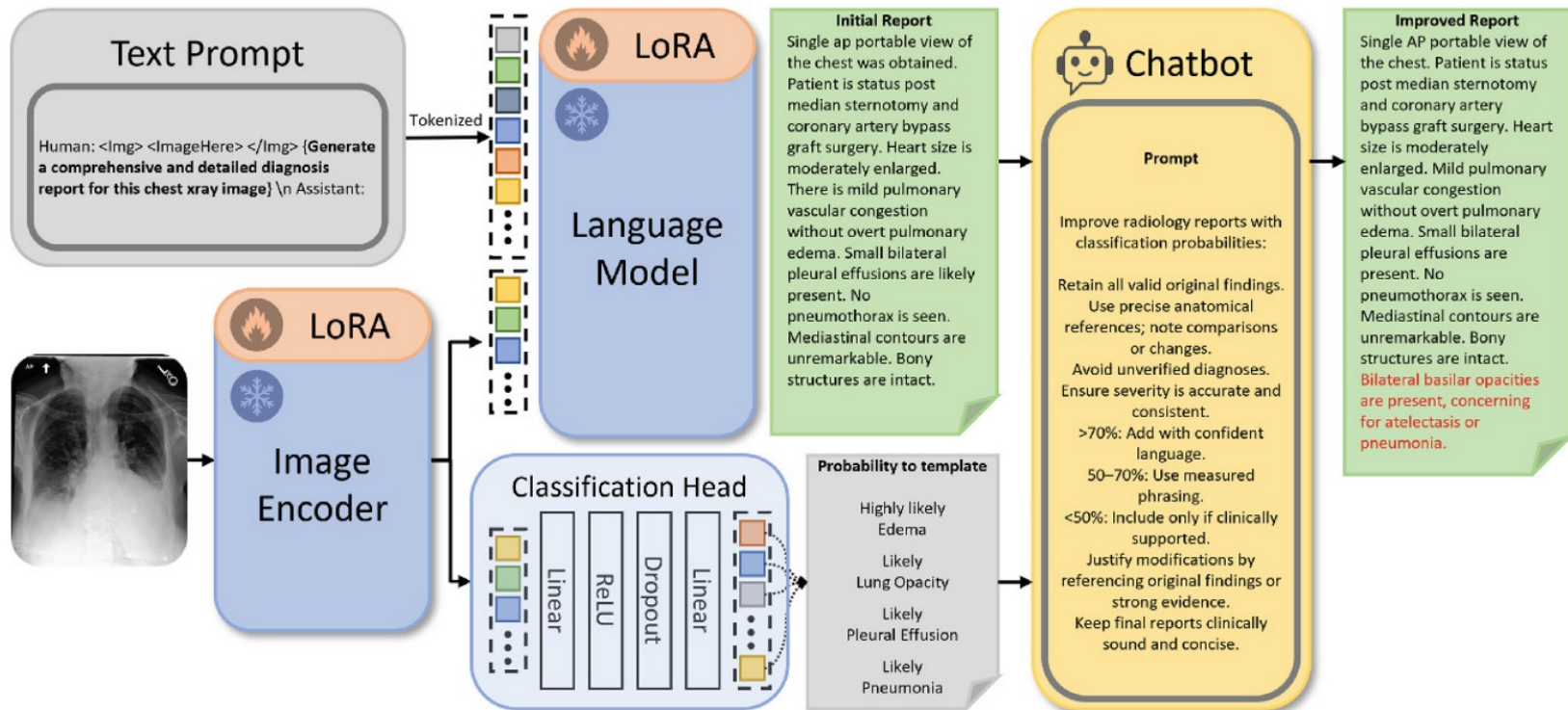
System	Protocol	No-prior/current Acc. / F1	Prior+current Acc. / F1	Δ supp. F1	same_view Acc. / F1	95% CI Acc.
Always-stable heuristic	input-free shortcut	0.400 / 0.190	0.400 / 0.190	+0.000	0.415 / 0.196	[0.369, 0.432]
Always-worsening heuristic	input-free shortcut	0.419 / 0.197	0.419 / 0.197	+0.000	0.407 / 0.193	[0.385, 0.452]
Finding-majority prior	input-free finding prior	0.488 / 0.458	0.488 / 0.458	+0.000	0.477 / 0.450	[0.462, 0.514]
MedSigLIP (Søllergren et al., 2025)	frozen head	0.452 / 0.432	0.541 / 0.489	+0.057	0.529 / 0.478	[0.512, 0.570]
BioViL-T (Bannur et al., 2023)	frozen head	0.517 / 0.474	0.554 / 0.527	+0.053	0.536 / 0.512	[0.524, 0.583]
TorchXRayVision DenseNet (Cohen et al., 2022)	frozen head	0.501 / 0.453	0.520 / 0.458	+0.005	0.507 / 0.437	[0.491, 0.548]
MedGemma prompt-only (Søllergren et al., 2026)	prompt-only	0.314 / 0.286	0.336 / 0.285	-0.001	0.329 / 0.271	[0.304, 0.367]
Qwen2.5-VL-7B (Bai et al., 2025)	fixed label prompt	0.097 / 0.135	0.238 / 0.228	+0.093	0.229 / 0.219	[0.210, 0.265]
CheXagent side-by-side MC [†] (Chen et al., 2024b)	fixed MC prompt	0.461 / 0.442	0.441 / 0.394	-0.048	0.454 / 0.407	[0.414, 0.470]
Lingshu-7B (Xu et al., 2025)	fixed label prompt	0.399 / 0.231	0.382 / 0.227	-0.004	0.379 / 0.217	[0.349, 0.413]
Lingshu-I-8B (Xu et al., 2025)	fixed label prompt	0.397 / 0.240	0.388 / 0.221	-0.019	0.379 / 0.213	[0.356, 0.420]
HuatuoGPT-Vision-7B (Chen et al., 2024a)	fixed label prompt	0.387 / 0.203	0.404 / 0.242	+0.039	0.398 / 0.238	[0.371, 0.436]
SmolVLM2-2.2B (Marafioti et al., 2025)	fixed label prompt	0.333 / 0.205	0.343 / 0.196	-0.009	0.336 / 0.190	[0.310, 0.373]
Q-LACE (ours)	query-conditioned VLM	0.490 / 0.408	0.592 / 0.562	+0.154	0.580 / 0.555	[0.566, 0.621]

Query intervention	Acc. $\uparrow$	Supported macro-F1 $\uparrow$	Δ supp. F1
Canonical finding query	0.592	0.562	—
Blank query	0.557	0.429	-0.133
Generic change query	0.556	0.442	-0.120
Random wrong finding query	0.485	0.449	-0.113
Paraphrased finding query	0.590	0.552	-0.010

Query and prior-image interventions lower F1. Query binding remains incomplete.

6. DuEL-Med

6.1. Classification-guided radiology report generation



I combine report generation with pathology probabilities to refine findings.

6. DuEL-Med

6.2. A report refinement example



Classifier: Lung Opacity, Pneumonia

GT: Lung Opacity, Pneumonia

GREEN Score:

Report from MambaXray-VL: 0.429, Report from DuEL-Med: 0.600

METEOR Score:

Report from MambaXray-VL: 0.140, Report from DuEL-Med: 0.132

Ground-Truth Report

again seen are nonspecific **bibasilar opacities** which have increased from. the apices of lungs are clear. there is no evidence of **pulmonary edema** **pleural effusion** or pneumothorax. **cardio-mediastinal** and **hilar** contours are unremarkable. no acute displaced rib fracture identified.

Report from MambaXray-VL

frontal and lateral views of the chest were obtained. there is no **focal consolidation** **pleural effusion** or pneumothorax. the **cardiomediastinal** silhouette is within normal limits. imaged osseous structures are intact. no free air below the right hemidiaphragm is seen. surgical clips project over the left upper abdomen.

Report from DuEL-Med

frontal and lateral views of the chest were obtained. there is no **focal consolidation**, **pleural effusion**, or pneumothorax. the **cardiomediastinal** silhouette is normal. imaged osseous structures are intact. no free air is seen below the right hemidiaphragm. surgical clips project over the left upper abdomen. **there is a subtle opacity in the right lower lobe that may represent early pneumonia, clinical correlation is recommended.**

**MIMIC-CXR
GREEN**

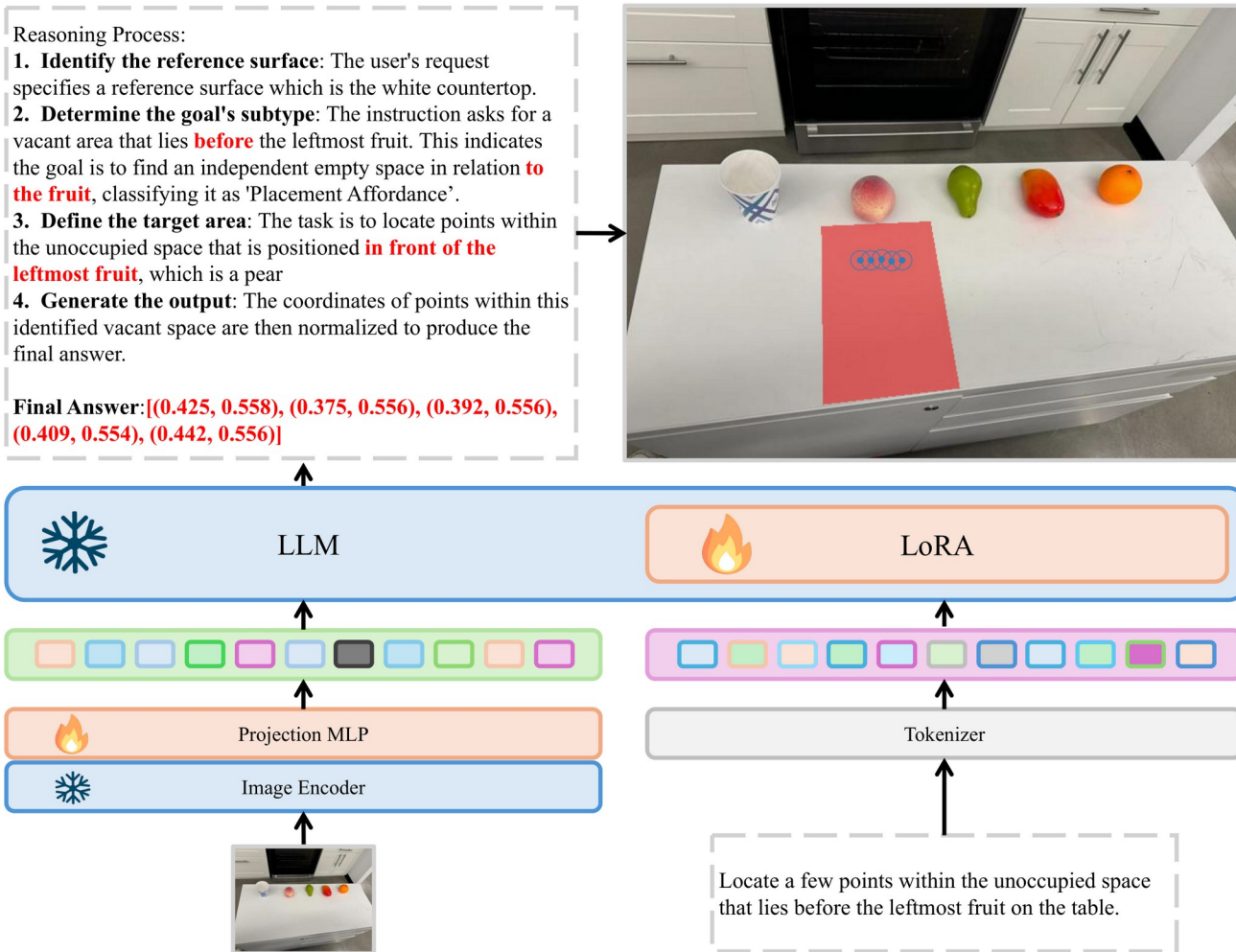
0.262 to 0.273
+4.2% relative

The figure shows one report.

Aggregate gains do not guarantee every finding is correct.

7. TRACE

7.1. Textual reasoning for affordance grounding

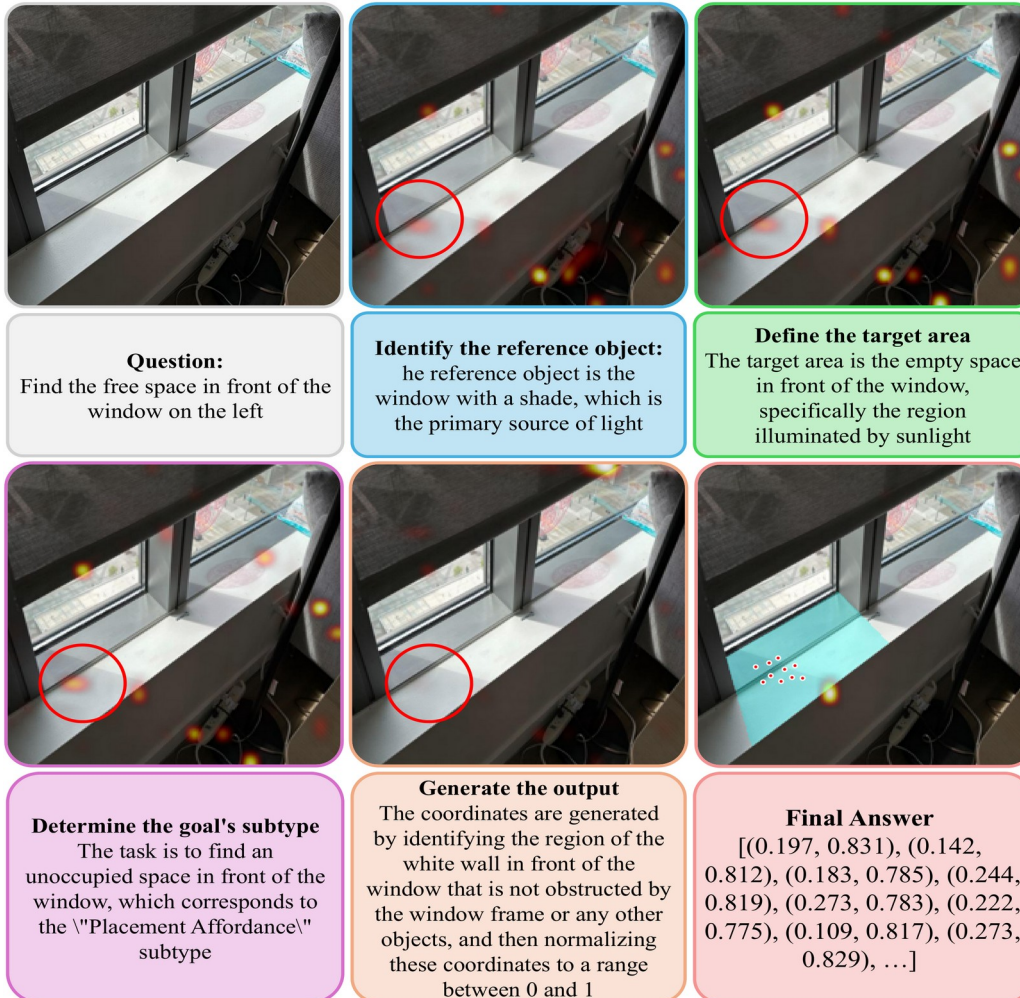


I co-developed VLM fine-tuning with reasoning supervision before coordinate prediction.

Nonmedical visual reasoning with grounded spatial outputs.

7. TRACE

7.2. Grounded outputs and attention inspection



Where2Place accuracy

43.9% to 48.1%

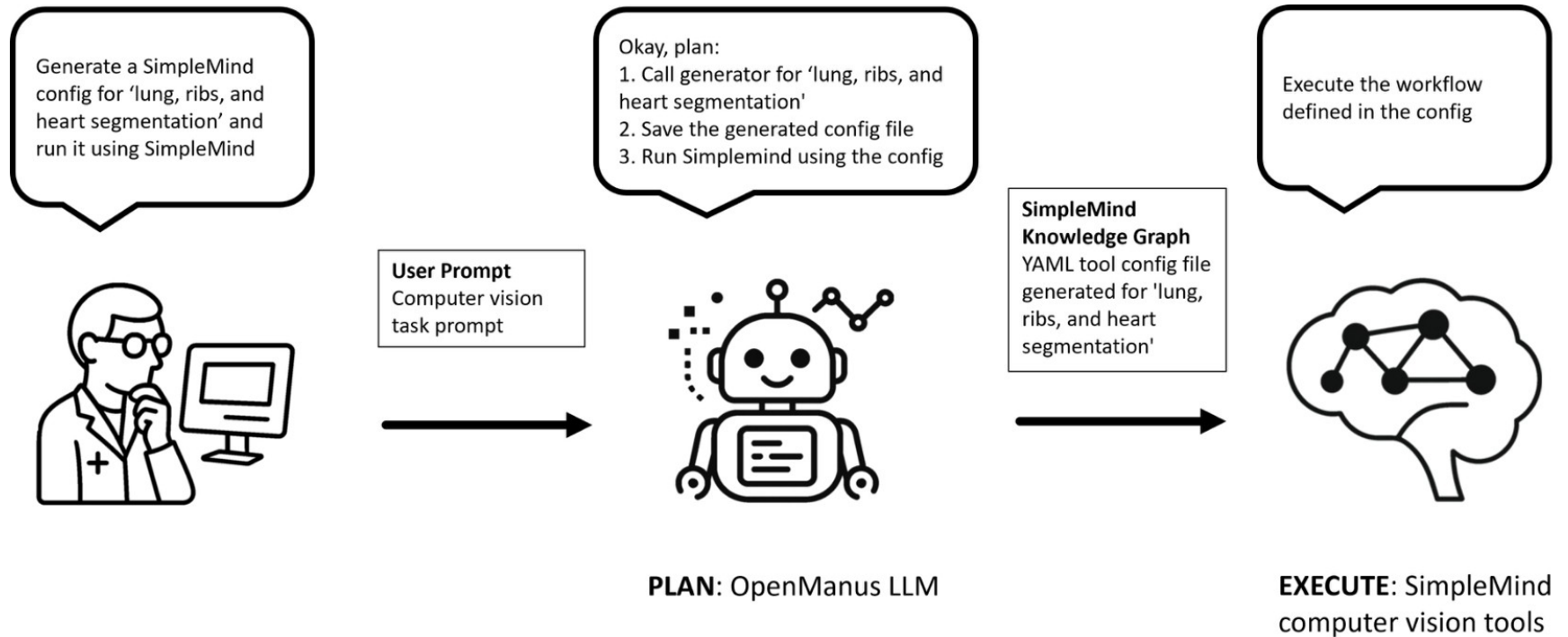
+4.2 percentage points

Hard subset: 55.0%

Attention maps help inspect behavior. They do not prove faithful reasoning.

8. OpenManus-SimpleMind

8.1. Executable computer vision workflows

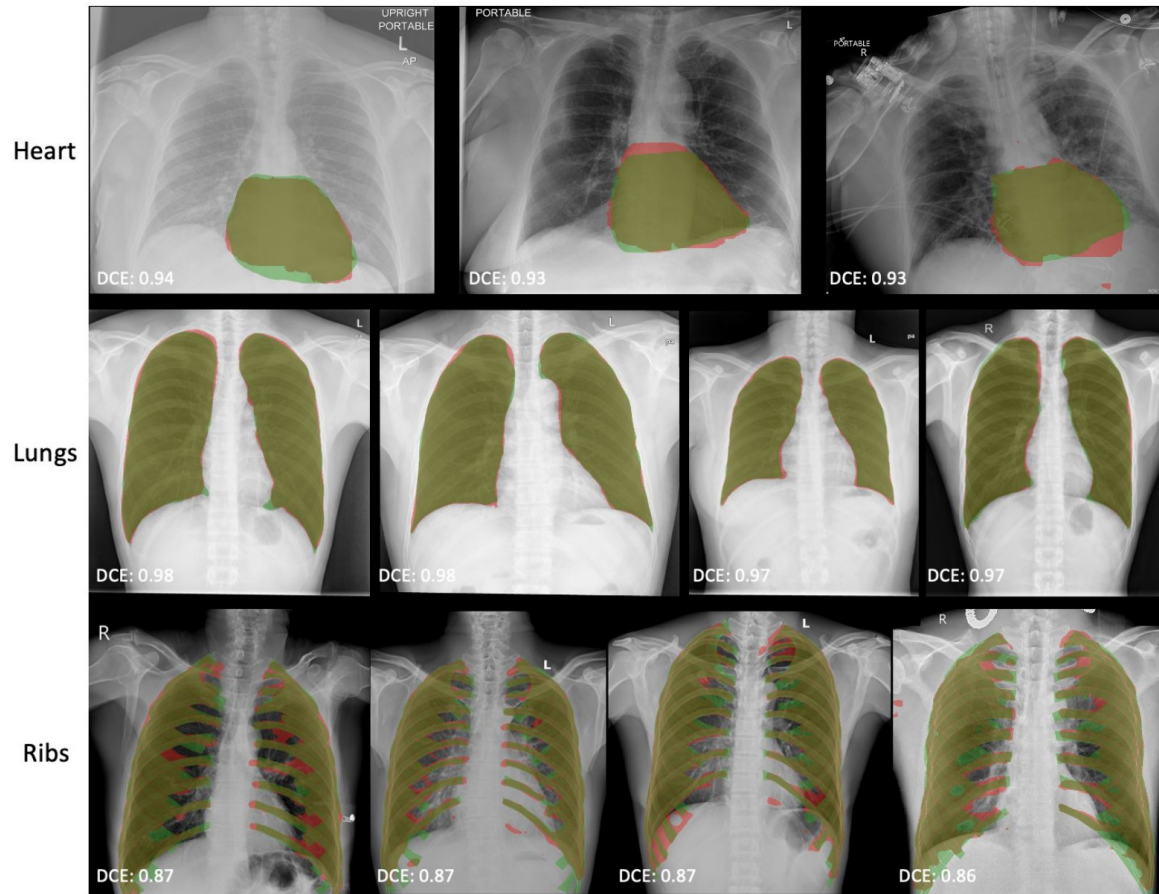


I connect configuration generation, verifier feedback and training/inference tools.

8. OpenManus-SimpleMind

8.2. Segmentation outputs from configured workflows

Anatomical Segmentation Result Visualization



Mean Dice

Lung 0.96

Heart 0.82

Ribs 0.83

Test images

50 / 49 / 50

Three prescribed tasks.
Scores measure
segmentation quality.

Research profile

Jin Kim, Los Angeles, CA

Multimodal post-training and grounded reasoning
Executable agent workflows for computer vision

UCLA Ph.D. candidate

Physics and Biology in Medicine, September 2022–present

Expected graduation: 2027

Full-time availability: early 2027, contingent on degree completion

Seoul National University, M.S. in Bioengineering

Attendance: June 2020–September 2022

Degree conferred: February 2024

[Portfolio website](#) [GitHub](#) [Google Scholar](#)