

- Generative AI for clinical text summarization

nature medicine

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | Published: 27 February 2024

Adapted large language models can outperform medical experts in clinical text summarization

[Dave Van Veen](#) , [Cara Van Uden](#), [Louis Blankemeier](#), [Jean-Benoit Delbrouck](#), [Asad Aali](#), [Christian Bluethgen](#), [Anuj Pareek](#), [Malgorzata Polacin](#), [Eduardo Pontes Reis](#), [Anna Seehofnerová](#), [Nidhi Rohatgi](#), [Poonam Hosamani](#), [William Collins](#), [Neera Ahuja](#), [Curtis P. Langlotz](#), [Jason Hom](#), [Sergios Gatidis](#), [John Pauly](#) & [Akshay S. Chaudhari](#)

[Nature Medicine](#) **30**, 1134–1142 (2024) | [Cite this article](#)

- Integrating LLMs into clinical workflows could alleviate documentation burden, allowing clinicians to focus more on patient care^[1].

[1] Van Veen, D., Van Uden, C., Blankemeier, L. et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med 30, 1134–1142 (2024).

- Generative AI for clinical diagnosis

Radiology

ORIGINAL RESEARCH • COMPUTER APPLICATIONS

Feasibility of Differential Diagnosis Based on Imaging Patterns Using a Large Language Model

Jonathan Kottlors, MD • Grischa Bratke, MD • Philip Rauen, MD • Christoph Kabbasch, MD • Thorsten Persigehl, MD • Marc Schlamann, MD • Simon Lennartz, MD

From the Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Strasse 62, 50937 Cologne, Germany. Received May 6, 2023; revision requested May 16; revision received June 2; accepted June 15. Address correspondence to S.L. (email: simon.lennartz@uk-koeln.de).

Supported in part by the German Federal Ministry of Education and Research Network of University Medicine 2.0 (grant no. 01KX2121).

Conflicts of interest are listed at the end of this article.

Supplemental material is available for this article

Radiology 2023; 308(1):e231167 • <https://doi.org/10.1148/radiol.231167> • Content code: AI • © RSNA, 2023

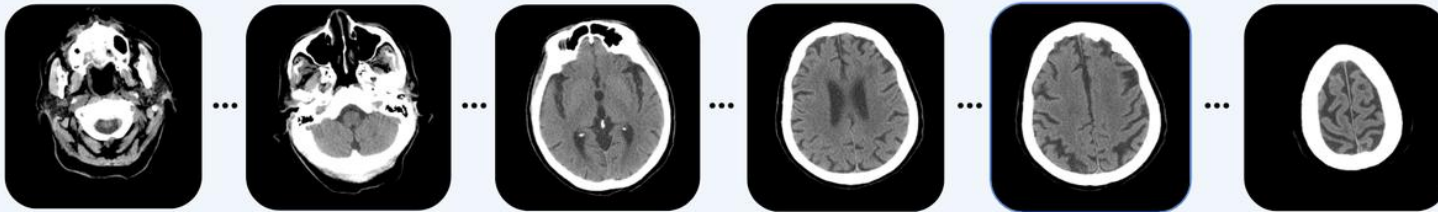
- GPT-4 achieved 93.8% expert acceptance rate when suggesting differential diagnoses for radiological imaging patterns.

- **Generative AI agent for real-world medical decision-making processes**
 - MDAgents, an adaptive multi-agent LLM framework for medical decisions, achieved best performance in all kinds of medical benchmarks.

- Generative AI for medical foundation model

Prompt

Given the image, what are the radiology findings?



Response

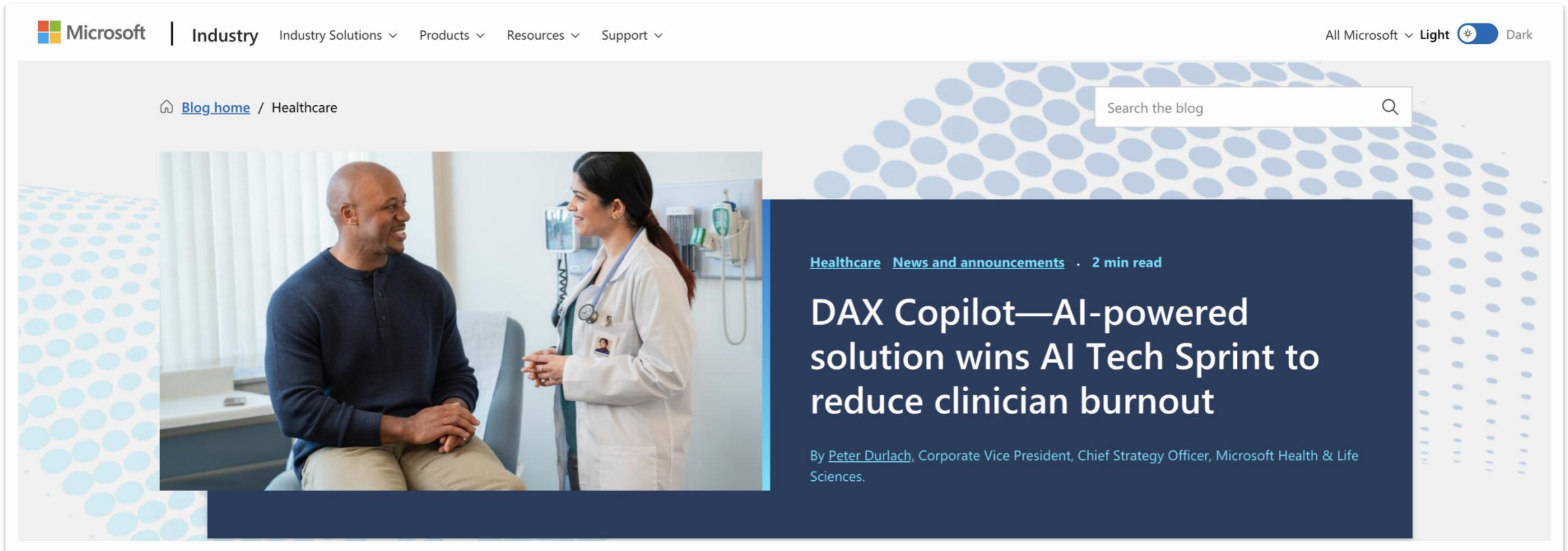
FINDINGS: mild cortical atrophy and periventricular white matter changes are noted compatible with microvasculopathy. there is proportionate ventriculomegaly. no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect. no abnormal extra-axial fluid collection is identified. no definitive evidence of intracranial hemorrhage or large territorial infarct within the limits of ct imaging. **old left basilar ganglia infarct**. no intracranial mass, swelling, hemorrhage, infarction or extra-axial collection is seen. visualized mastoid air cells: well pneumatized. visualized paranasal sinuses: well pneumatized. atherosclerotic calcification of vasculature at the skull base.

Radiologist Written Report

FINDINGS: the fourth ventricle is midline. the remaining ventricular system is within normal limits. there is mild chronic white matter microangiopathic disease and moderate generalized cortical atrophy. no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. bone windows demonstrate an intact skull. the remaining visualized paranasal sinuses and mastoids are clear. chronic white matter microangiopathic disease, mild generalized cortical atrophy, not significantly changed since 2015. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect.

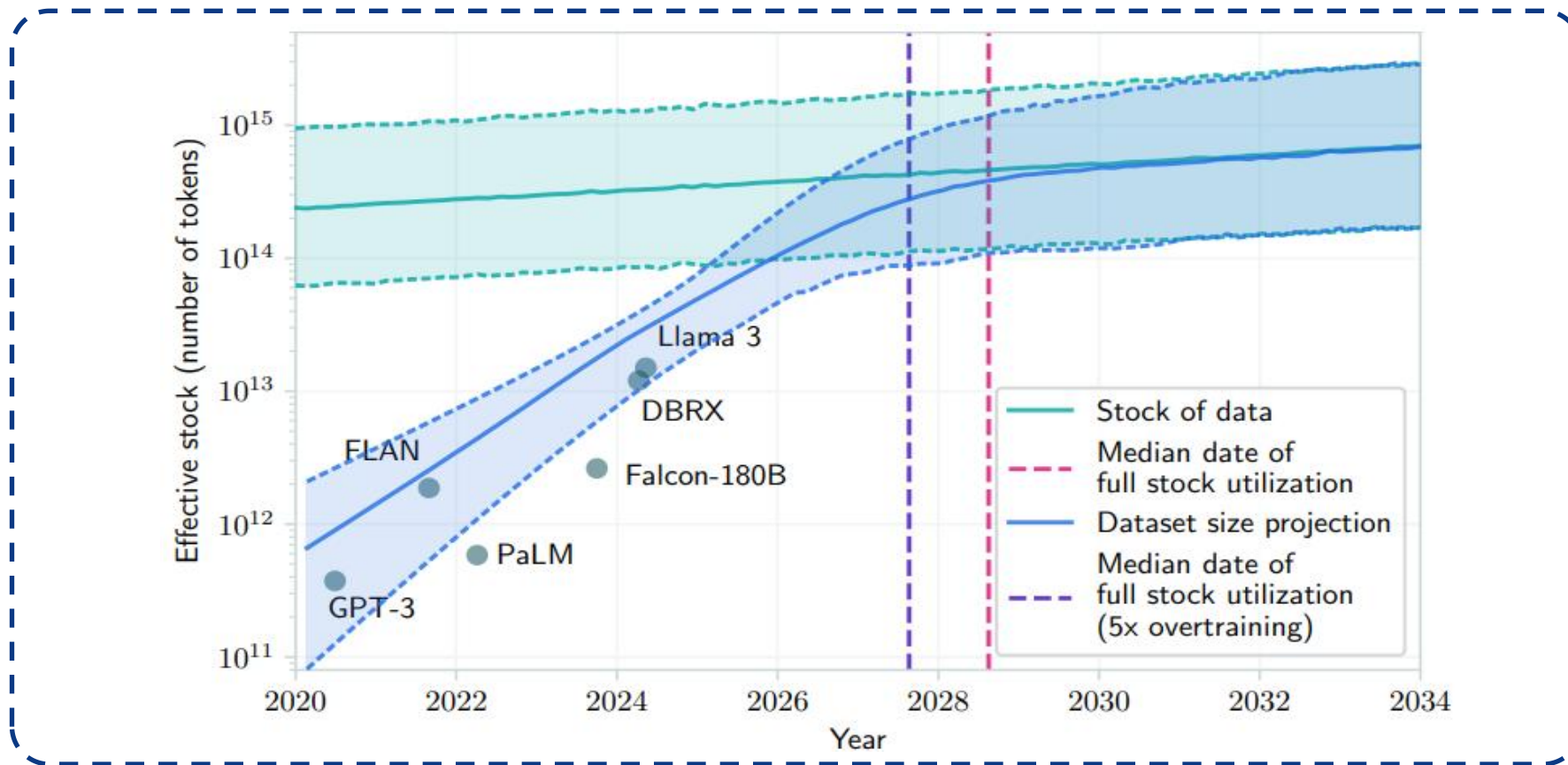
- Med-Gemini, a family of fine-tuned Gemini models for multimodal medical tasks, can analyze 2D/3D medical images, generate radiology reports, answer clinical questions, and predict disease risks from genomic data.

- Generative AI agent for doctor assistance



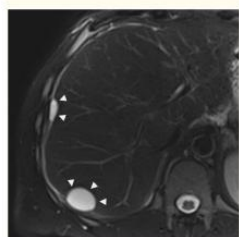
- DAX Copilot automates clinical documentation from patient conversations and assists clinicians with daily tasks to reduce administrative burden and burnout.

- **Problem: data shortage for foundation models**
 - Data will run out for large language models in the natural image field by 2028 [1].
 - It will happen much faster in the **medical domain**.



[1] Villalobos, Pablo, et al. "Position: Will we run out of data? Limits of LLM scaling based on human-generated data." *Forty-first International Conference on Machine Learning*.

- **Motivation: enhance multigranular alignment and effective scaling through synthetic data**



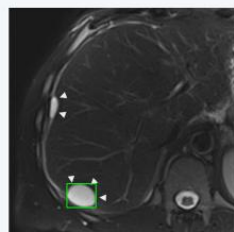
ROCO

A 49-year-old man presenting a pancreatic neoplasia with peritoneal carcinomatosis. Axial T2W TSE fat-suppressed MRI shows two hepatic pericapsular implants of peritoneal carcinomatosis (arrowheads), biconvex, in high signal intensity.

Modality

Structure
Detection

Lesion
Texture



Med-Trinity-25M
(Ours)

The image is an axial T2W TSE fat-suppressed MRI focusing on the liver and surrounding areas, highlighting two hepatic pericapsular implants indicative of peritoneal carcinomatosis, marked by their high signal intensity and biconvex shape. These abnormalities, located on the right side of the liver, are positioned horizontally to the left and vertically at the bottom of the image, occupying about 1.5% of the area. The region of interest reveals these unusual features, contrasting with the normal liver texture and appearance. These hepatic implants are significant as they suggest a spread from the primary pancreatic neoplasia, indicating a direct relationship where the primary disease has metastasized to adjacent organs, further complicating the patient's condition.

Modality

Structure
Detection

ROI
Analysis

Lesion
Texture

Local-global
Relation

existing multimodal annotation:

- lack of ROI in images
- lack of detailed description
- lack of local-global relation

lack of
multigranular
alignment



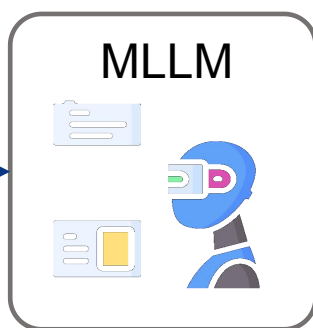
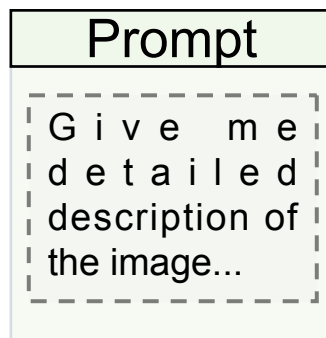
Multigranular annotation:

- rich information
- multigranular attributes

Datasets Medtrinity-25M

Naive Prompt

- Naive prompt MLLM to generate fine-grained descriptions will normally fail.

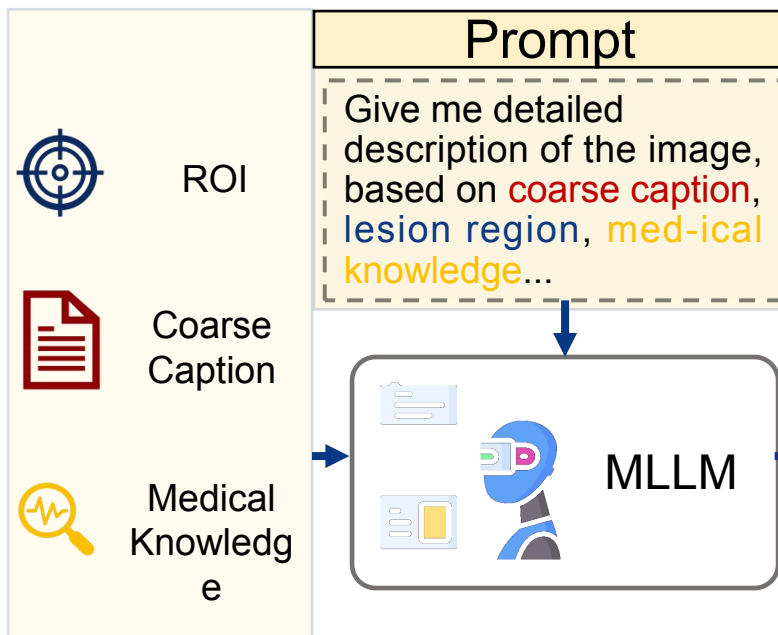


The image is a chest-ray showing the thoracic region, including the lungs and heart, primarily located in the central to lower thoracic cavity...This X-ray is typically used to assess the condition of these organs, looking for signs of diseases such as pneumonia or other pulmonary conditions.

Short Description

Undefine Disease

Ours



A chest X-ray.

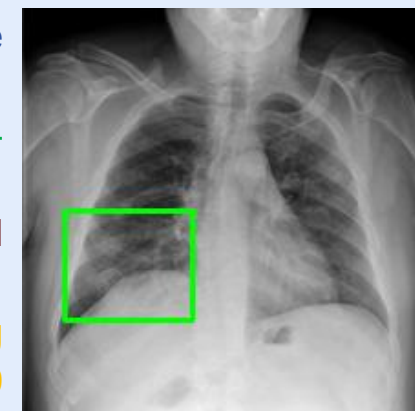
Showing lungs centrally located within the thoracic cavity.

ROI is positioned horizontally at the left-center and vertically at ...,

Region exhibits increased opacity and irregular texture, indicating ...

Showing a pattern of bilateral lung involvement typically seen in COVID-19 cases.

Image-ROI-Description Triplets



- Equip MLLM with our pipeline to successfully generate fine-grained descriptions.

- Method



Without coarse caption:

The image is a chest X-ray showing the thoracic region, including the lungs and heart, primarily located in the central to lower thoracic cavity...This X-ray is typically used to assess the condition of these organs, looking for signs of diseases such as pneumonia or other pulmonary conditions.



Coarse Caption

“A chest X-Ray image with COVID-19 in the lungs”

Specify Disease

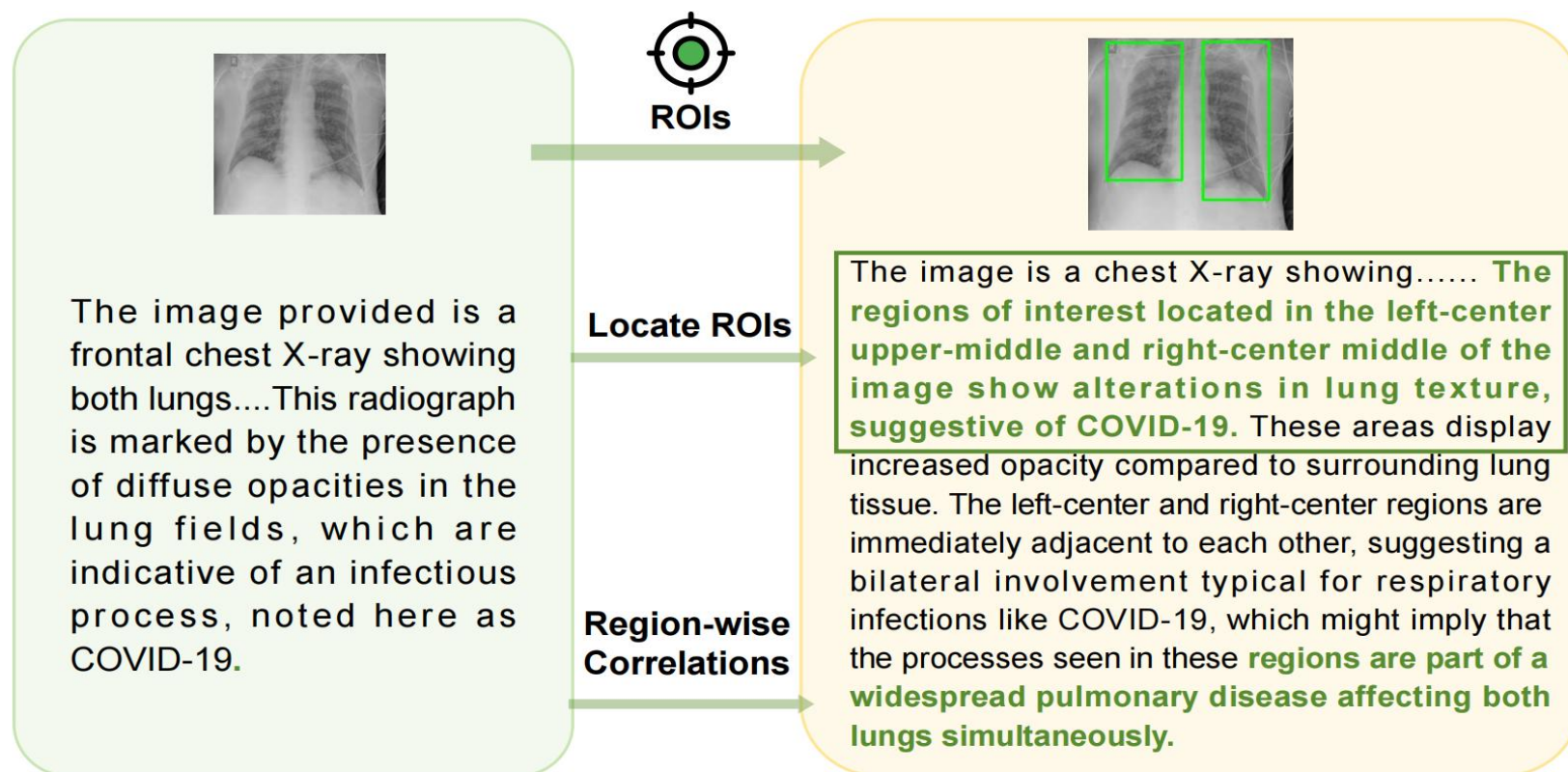


With coarse caption:

The image provided is a frontal chest X-ray showing both lungs....This radiograph is marked by the presence of diffuse opacities in the lung fields, which are indicative of an infectious process, noted here as COVID-19.

Coarse caption help to specify disease type

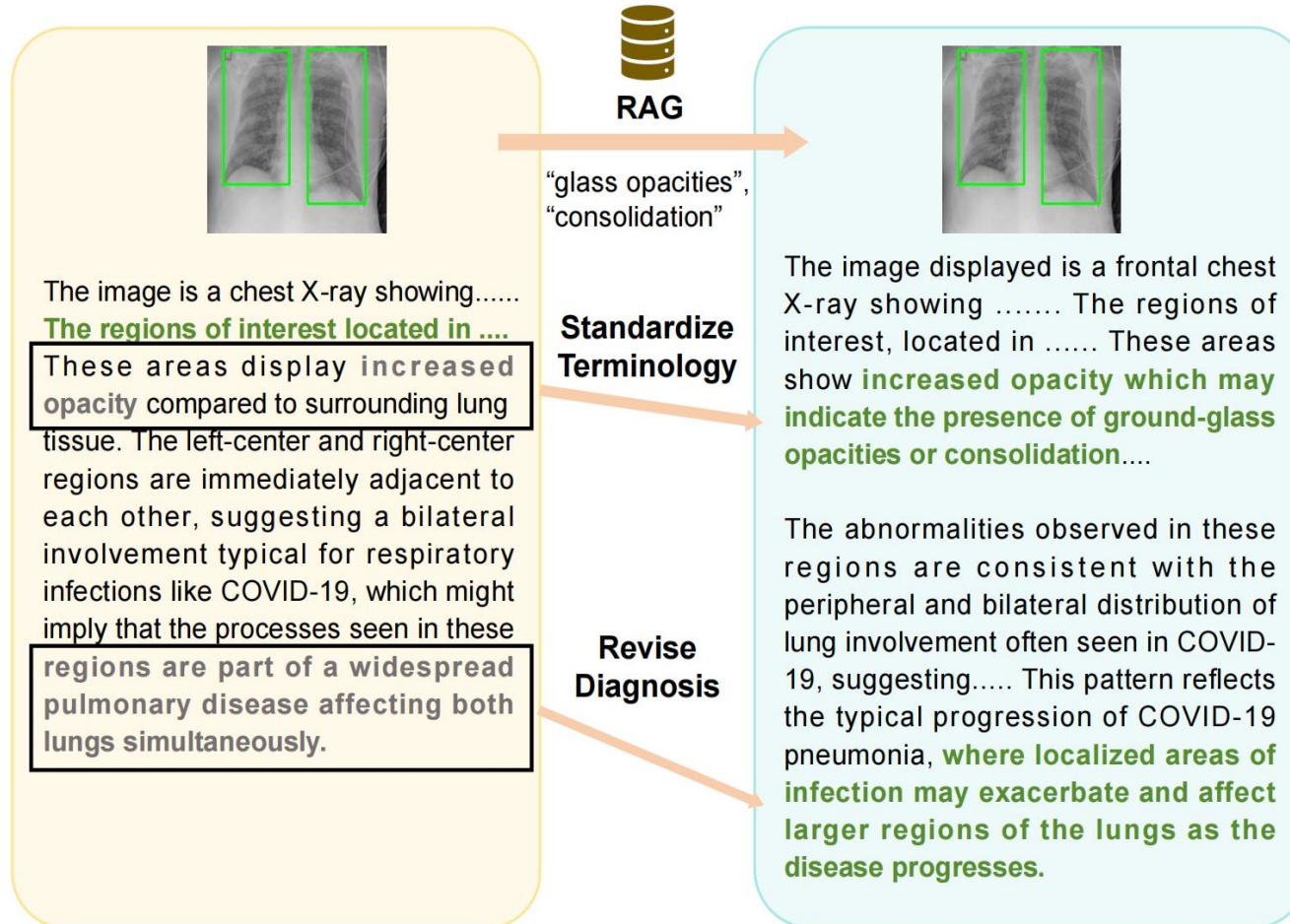
- Method



ROIs guide MLLM to

- focus on local lesion area
- access impact of ROIs on adjacent normal regions

- Method

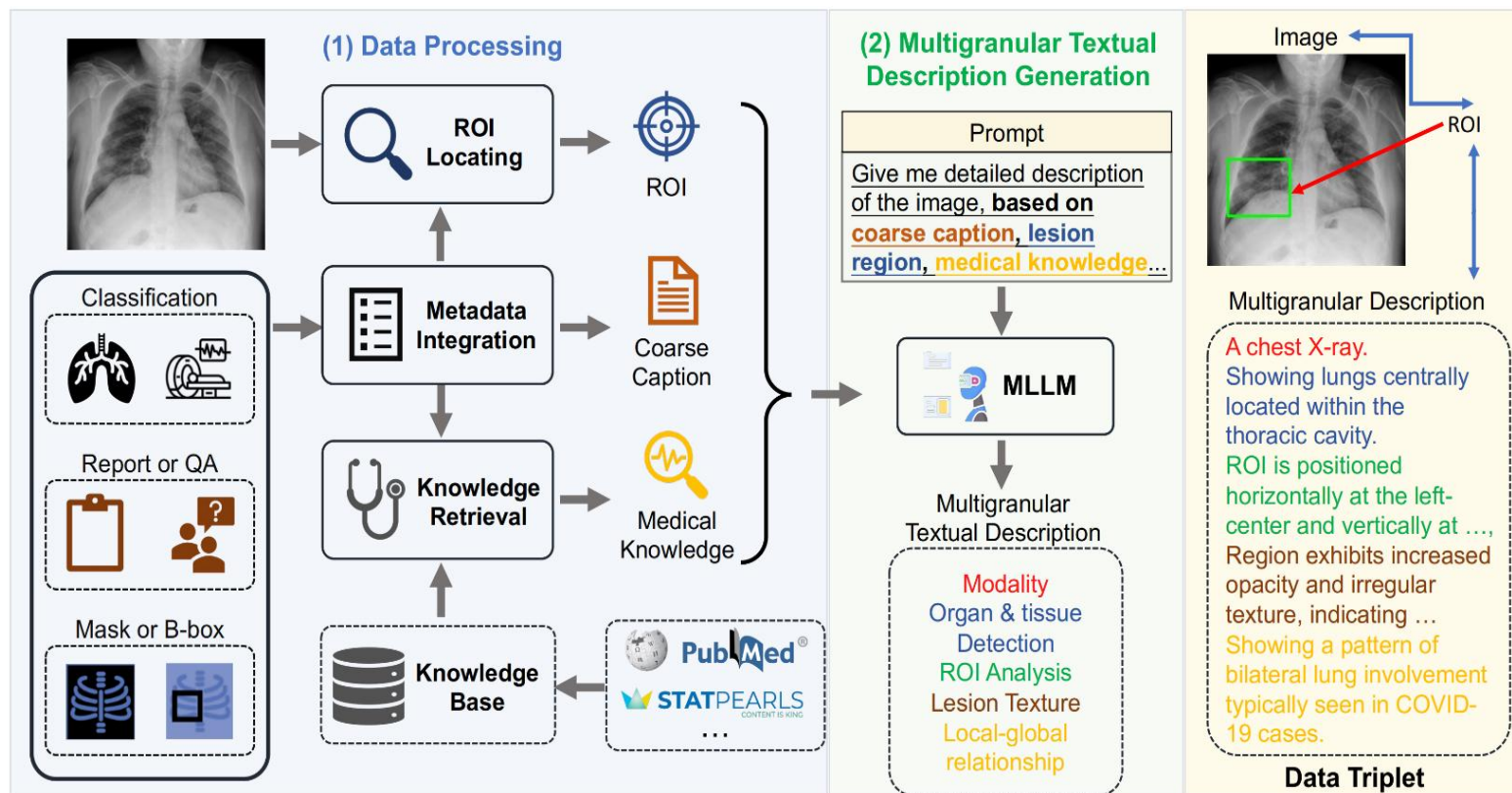


RAG retrieves external medical knowledge

- standardize medical terminology
- revise diagnosis

Datasets Medtrinity-25M

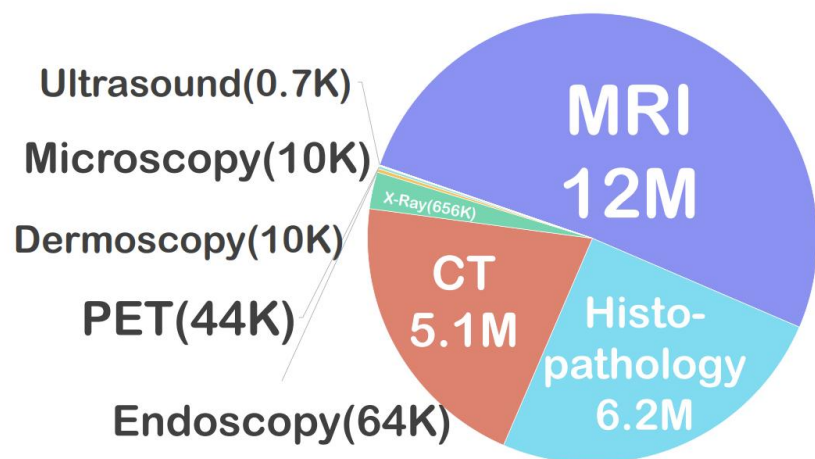
- A step-by-step pipeline for generating multigranular annotations for medical images
 - RAG and metadata for data trustworthiness
 - ROI for local focus
 - step-by-step prompt for generating local and global multigranular annotations



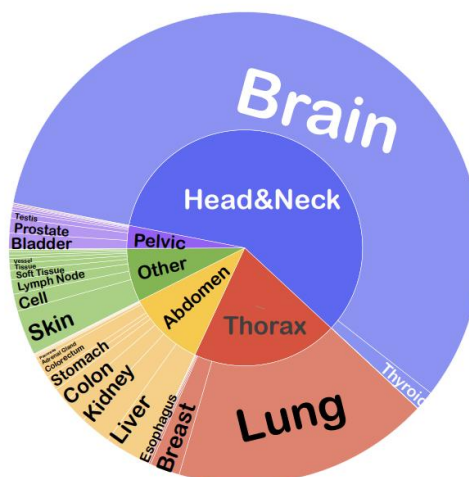
Datasets Medtrinity-25M

• Statistics Overview

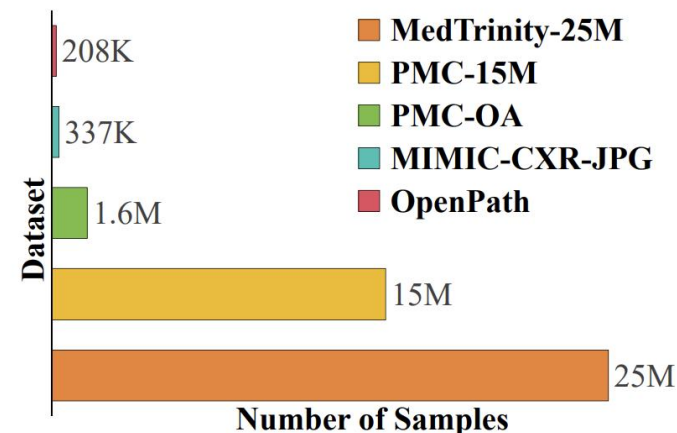
- 25 million images with ROIs and multigranular text descriptions
- 10 imaging modalities
- 65 kinds of disease
- 48 kinds of anatomical structures



(a) Modality distribution.



(b) Biological structures.



(c) Data size comparison.

- Main Result

Method	VQA-RAD			SLAKE			PathVQA		
	Open	Closed	Average	Open	Closed	Average	Open	Closed	Average
GPT-4V (Achiam et al., 2023)	39.5	78.9	59.2	33.6	43.6	38.6	-	-	-
VL Encoder-Decoder (Bazi et al., 2023)	71.5*	82.5	77.0	-	-	-	71.5*	85.6	78.6
Q2ATransformer (Liu et al., 2023)	79.2*	81.2	80.2	-	-	-	54.9*	88.9	71.9
Prefix T. Medical LM (Sonsbeek et al., 2023)	-	-	-	84.3*	82.0	83.2	40.0*	87.0	63.5
PubMedCLIP (Eslami et al., 2023)	60.1*	80.0	70.1	78.4*	82.5	80.5	-	-	-
BiomedCLIP (Zhang et al., 2023a)	67.6*	79.8	73.7	82.1*	89.7	85.9	-	-	-
M2I2 (Li et al., 2023)	66.5*	83.5	75.0	74.7*	91.1	82.9	36.3*	88.0	62.2
LLaVA (Liu et al., 2024)	50.0	65.1	57.6	78.2	63.2	70.7	7.7	63.2	35.5
LLaVA-Med (Li et al., 2024b)	61.5	84.2	72.8	83.1	85.3	84.1	37.9	91.2	64.5
LLaVA-Tri	77.1	86.0	81.6	86.2	89.3	87.8	66.5	99.0	82.8

- After alignment pretraining on Medtrinity-25M, our LLaVA-Tri achieves SOTA in all three VQA benchmarks.

- Ablation Study

Table 4: **Comparison of different models with or without alignment pretraining with MedTrinity-25M.** The notation **w/** and **w/o** indicate models with and without pretraining on MedTrinity-25M, respectively.

Model	Dataset Use	VQA-RAD			SLAKE			PathVQA		
		open	close	average	open	close	average	open	close	average
LLaVA-Tri	w/o	64.6	77.0	70.8	79.3	84.0	81.7	55.0	94.0	74.5
	w/	77.1 + (12.5)	86.0 + (9.0)	81.6 + (10.8)	86.2 + (6.9)	89.3 + (5.3)	87.8 + (6.1)	66.5 + (11.5)	99.0 + (5.0)	82.8 + (8.3)
MiniCPM-V-2.6-8B (Yao et al., 2024)	w/o	48.5	86.4	67.5	57.2	80.0	68.6	31.2	90.5	60.9
	w/	50.5 + (2.0)	87.6 + (1.2)	69.1 + (1.6)	65.3 + (8.1)	80.6 + (0.6)	73.0 + (4.4)	34.2 + (3.0)	94.8 + (4.3)	64.5 + (3.6)
InternVL2-8B (Chen et al., 2024)	w/o	38.2	76.2	57.2	61.7	77.8	69.8	16.8	86.4	51.6
	w/	40.7 + (2.5)	80.0 + (3.8)	60.4 + (3.2)	66.4 + (4.7)	78.8 + (1.0)	72.6 + (2.8)	23.6 + (6.8)	87.4 + (1.0)	55.5 + (3.9)
PubMedCLIP (Eslami et al., 2023)	w/o	55.6	79.3	67.5	-	-	-	-	-	-
	w/	60.6 + (5.0)	79.7 + (0.4)	70.2 + (2.7)	-	-	-	-	-	-

- Multigranular alignment pretraining on Medtrinity-25M can improve performance on all kinds of VLMs and CLIP-based models.

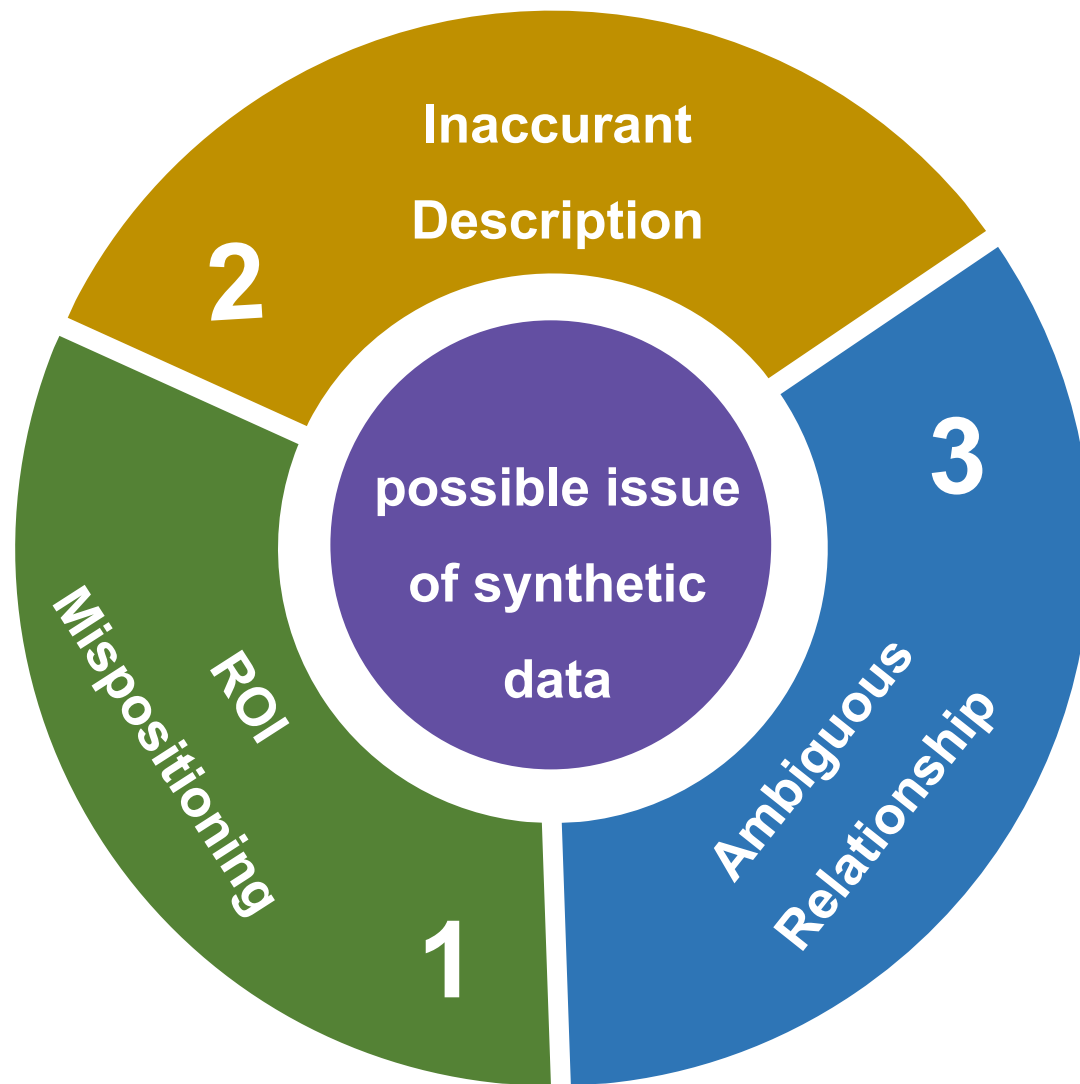
- Limitation

Evaluator	Attributes					Avg.
	Modality	Organ Detection	ROI Analysis	Lesion Texture	Region-wise Correlations	
LLM	1.00/1.00	0.90/1.00	0.90/1.00	0.80/1.00	0.70/1.00	0.86/1.00
Expert	1.00/1.00	0.90/1.00	0.90/1.00	0.70/1.00	0.80/1.00	0.85/1.00

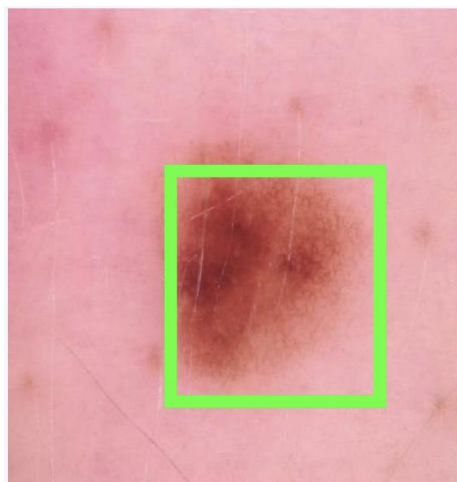
Noise in annotation

- expert and LLM evaluation show relatively high accuracy of synthetic data
- noise still exists particularly in complex attributes (e.g., region-wise correlations)
- large scale datasets are able to improve model despite the presence of noise

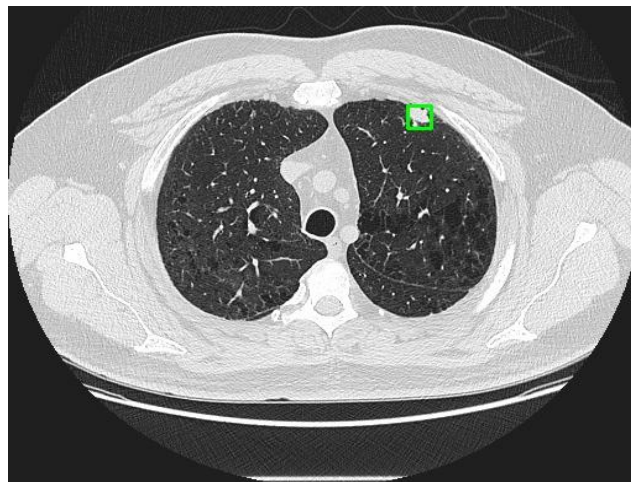
- Limitation



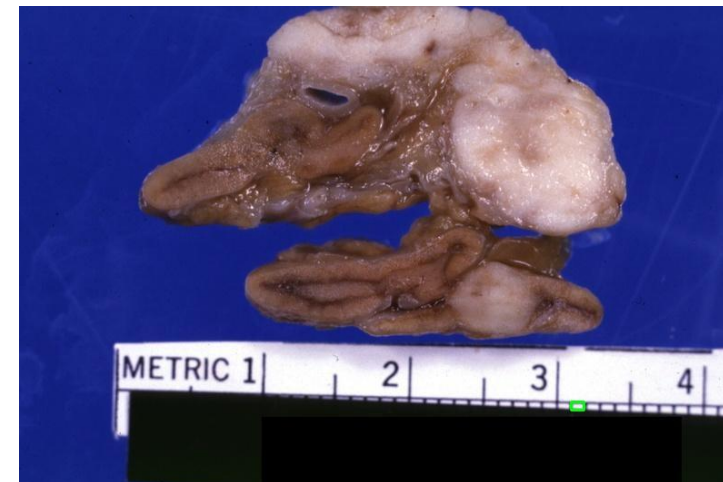
1 ROI Mispositioning



A natural image
of skin disease



A radiology image
of lung cancer

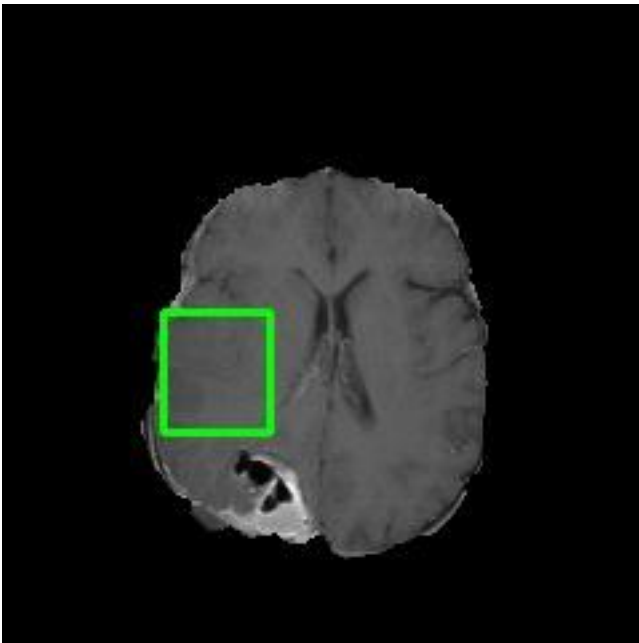


A natural image
of metastatic carcinoma

- Leverage expert models to identify ROI for lesion localization.
 - **Challenges:** detecting lesions when applied to rare diseases or natural images with out-of-distribution objects

2 Inaccurant Description

A correct example:



An image of brain tumor

Modality: The image is a magnetic resonance imaging (MRI) scan of the brain.

Anatomical Structural: ...showing a cross-sectional view with various brain structures visible, such as the cerebral hemispheres.

ROI locations: The region of interest, located left-center and in the middle vertically, occupies approximately 3.1% of the area.

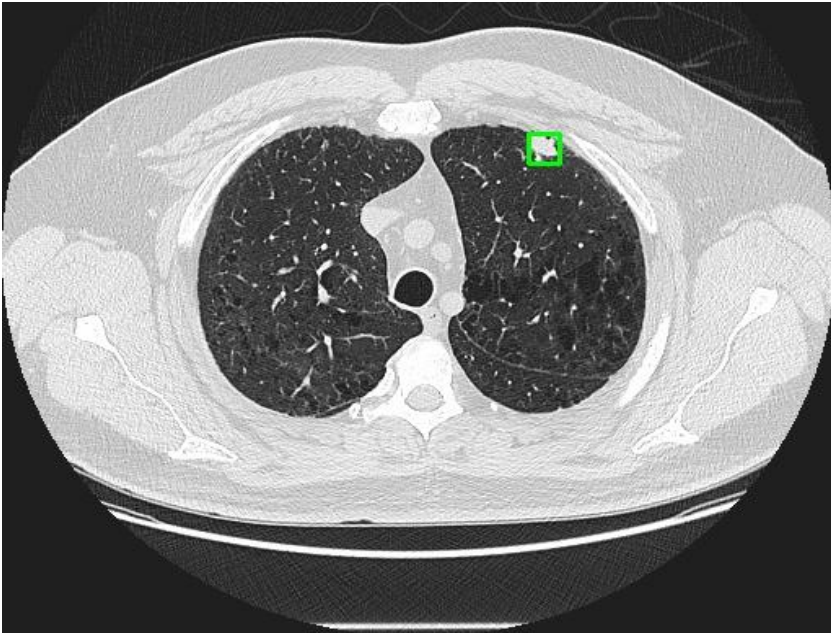
ROI description: It appears to have an abnormal texture or signal intensity compared to the surrounding brain tissue, which may indicate the presence of a disease of tumor.

Local and global relation: This abnormal area could potentially be affecting or be affected by adjacent brain structures due to...

- **ROI description:** Identify the abnormality and determine its associated disease.

2 Inaccurant Description

An incorrect example:



An image of lung cancer

Modality: The image is a transverse section of a thoracic CT scan, showing the lungs, heart, and part of the spine.

Anatomical Structural: The lungs are the primary organs visible, with the heart centrally located between them and the spine visible posteriorly.

ROI locations: The region of interest, located in the right-center and upper-middle of the image, occupies approximately 0.1% of the area and is situated in the lung tissue.

Inaccuracy: nonspecific disease

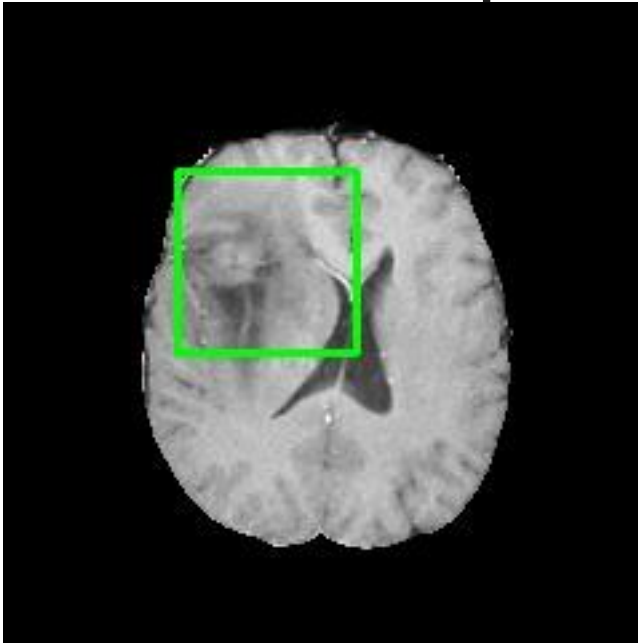
ROI description: This region exhibits an unusual appearance, potentially indicating a disease process, **such as an abnormal texture or density compared to the surrounding lung parenchyma.**

Local and global relation: The relationship between the content in the region of interest and other regions could be....

- Lack of sufficient medical knowledge to specify of disease with accurant terminology.

3 Ambiguous Relationship

A correct example:



An image of brain tumor

Global: The image is a magnetic resonance imaging (MRI) scan of the brain...

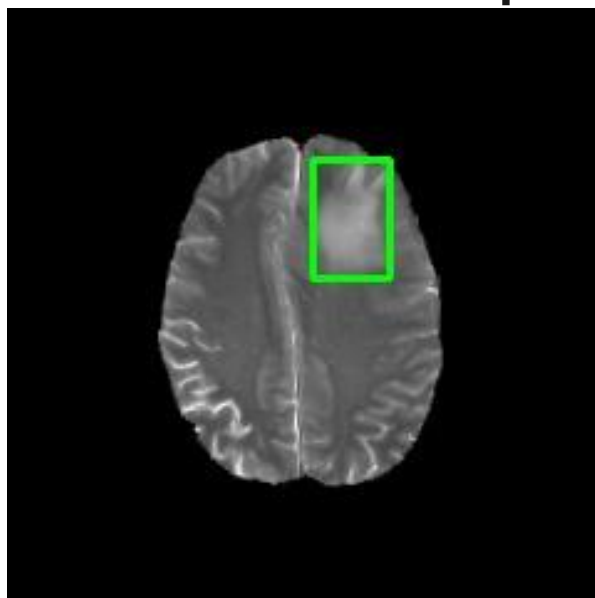
Lobal: The region of interest, located centrally and in the middle of the image, encompasses an area of the brain that appears to have abnormal tissue characteristics, possibly indicating a brain tumor. This region is characterized by a difference in texture and intensity compared to the surrounding brain tissue, which may suggest the presence of a pathological process

Local-global Relationship: The abnormal area could be affecting or be affected by adjacent brain structures due to its central location, potentially influencing nearby tissue either by direct invasion or secondary effects such as edema or mass effect

- **Local-global Relationship:** How local abnormal area may affect the global anatomical structure as the disease progresses?

3 Ambiguous Relationship

An incorrect example:



An image of brain edema

Global: The image is a magnetic resonance imaging (MRI) scan of the brain...

Lobal: The region of interest, located centrally and in the upper-middle area of the image shows a brain edema in the brain tissue, which could be indicative of a disease process given its distinct appearance compared to the surrounding brain matter.

Ambiguity: too coarse to learn the effect of local abnormality

Local-global Relationship: This abnormality is likely related to the surrounding brain structures, **potentially affecting them or being affected by a pathological process occurring in the brain.**

- Generative models sometimes struggle to reason about how localized abnormalities contribute to global disease:
 - Insufficient data availability to train models on complex pathological reasoning relationships.