RaDialog: Large Vision-Language Models for X-Ray Reporting and Dialog-Driven Assistance

Chantal Pellegrini[1,3], Ege Özsoy[1,3], Benjamin Busam[1],
Benedikt Wiestler[2], Nassir Navab[1], and Matthias Keicher[1]

[1]Computer-Aided Medical Procedures and Augmented Reality, Technical University of Munich, Germany
[2]AI for Image-Guided Diagnosis and Therapy, Technical University of Munich, Germany
[3]Munich Center for Machine Learning (MCML), Garching, Germany

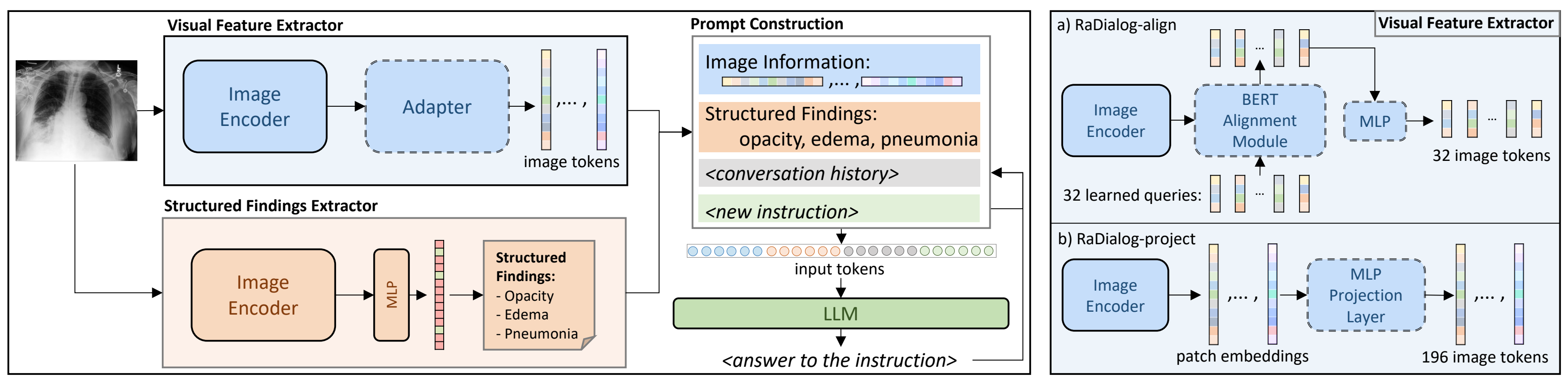
Abstract

Conversational AI tools for generating and discussing accurate radiology reports could transform radiology by **enabling collaborative, human-in-the-loop diagnostic processes**, saving time, and enhancing report quality.

- **Main Contributions**:
  - Propose a novel **dual-branch architecture** incorporating structured clinical findings with image embeddings.
  - Design an **interactive instruct dataset** to combat the issue of catastrophic forgetting and enable dialog-based human-AI collaboration.
  - Introduce a **context-dropping augmentation** that improves the model's attention to the image information.
  - **In-depth evaluation** on report generation and conversational downstream tasks.

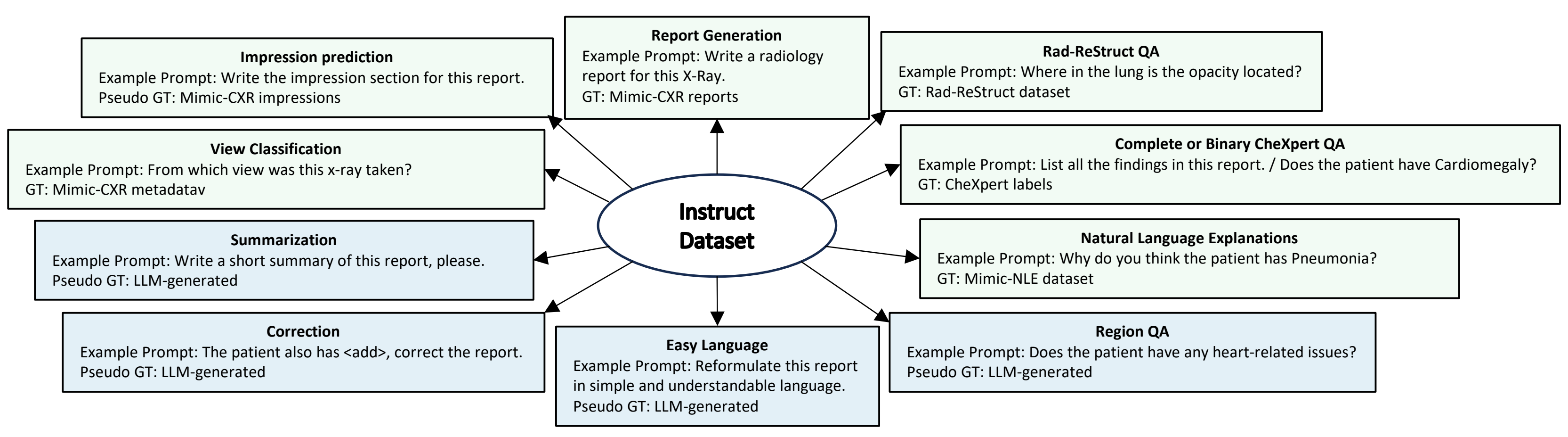
Method

Architecture



- **Visual Feature Extractor:** Extracts detailed visual encodings with a domain-specific image encoder and aligns them to the text space with the adapter module:
  - **RaDialog-align:** BERT-based alignment module, pre-trained on x-ray report pairs to extract 32 aligned image tokens
  - **RaDialog-project:** MLP projects all patch embeddings directly to language model tokens, trained end-to-end
- **Structured Findings Extractor:** Builds a structured representation of the main clinical findings in the image using a CLIP-based multi-label classifier, guiding the model on a higher level
- **Prompt Construction:** Image encoding, list of structured findings, conversation history, and instruction are converted into one prompt as input for the LLM
- **Language Model:** The LLM processes the prompt and produces an instruction-specific response. Fine-tuned using LoRA on x-ray report generation and auxiliary interactive downstream tasks.

Instruct Dataset



Impression prediction
Example Prompt: Write the impression section for this report.
Pseudo GT: Mimic-CXR impressions

Report Generation
Example Prompt: Write a radiology report for this X-Ray.
GT: Mimic-CXR reports

Rad-ReStruct QA
Example Prompt: Where in the lung is the opacity located?
GT: Rad-ReStruct dataset

View Classification
Example Prompt: From which view was this x-ray taken?
GT: Mimic-CXR metadatav

Complete or Binary CheXpert QA
Example Prompt: List all the findings in this report. / Does the patient have Cardiomegaly?
GT: CheXpert labels

Summarization
Example Prompt: Write a short summary of this report, please.
Pseudo GT: LLM-generated

Natural Language Explanations
Example Prompt: Why do you think the patient has Pneumonia?
GT: Mimic-NLE dataset

Correction
Example Prompt: The patient also has <add>, correct the report.
Pseudo GT: LLM-generated

Easy Language
Example Prompt: Reformulate this report in simple and understandable language.
Pseudo GT: LLM-generated

Region QA
Example Prompt: Does the patient have any heart-related issues?
Pseudo GT: LLM-generated

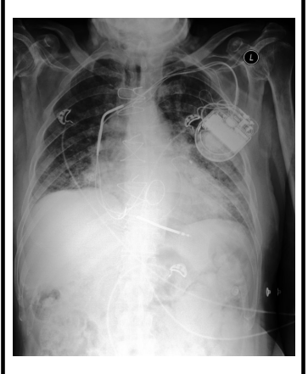
- **Catastrophic Forgetting:** Training only on image-report pairs degrades the LLM's performance on other tasks significantly, so a diverse dataset is essential.
- **Instruct Dataset:** New dataset with 580k samples across 10 tasks, each with diverse prompts, to ensure model retains flexibility and conversation abilities:
  - **Core Diagnostic Tasks:** Based on ground-truth data from different datasets, e.g. Findings QA or View Classification.
  - **Replay tasks**: Pseudo ground-truth generated by the non-fine-tuned LLM to prevent forgetting, e.g. Correction or Easy Language
- **Context Dropping Augmentation:** Randomly includes full, partial, or no report in training samples to force the model to adapt to varying levels of report completeness, improving its ability to rely on image features alone for downstream tasks when necessary.

Acknowledgements

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) under project DIVA (13GW0469C) and the Bavarian Ministry of Economic Affairs, Regional Development and Energy (StMWi) under project ThoraXAI (DIK-2302-0002).

Results

Report Generation

Input X-ray	Ground Truth	Prediction
	**CP Findings:** Lung Opacity, Cardiomegaly, Edema **Report:** Patient is status post median sternotomy, CABG, and mitral valve replacement. A left-sided AICD device is noted with leads terminating in the right atrium, right ventricle, and coronary sinus. Mild enlargement of the cardiac silhouette is redemonstrated, with unchanged tortuosity of the thoracic aorta. There is perihilar haziness with vascular indistinctness and diffuse alveolar opacities compatible with moderate pulmonary edema. No large pleural effusion or pneumothorax is seen. There are no acute osseous abnormalities.	**CP Findings:** Lung Opacity, Cardiomegaly, Edema **Report:** The patient is status post median sternotomy and CABG. Left-sided AICD/pacemaker device is noted with leads terminating in the right atrium, right ventricle, and coronary sinus . Moderate cardiomegaly is re- demonstrated. Mediastinal contours are unchanged. There is mild pulmonary edema, slightly improved compared to the previous exam. Ill-defined opacities in the lung bases may reflect areas of atelectasis. No pleural effusion or pneumothorax is identified. There are no acute osseous abnormalities.

Comparison to SOTA domain-specific VLLMs:

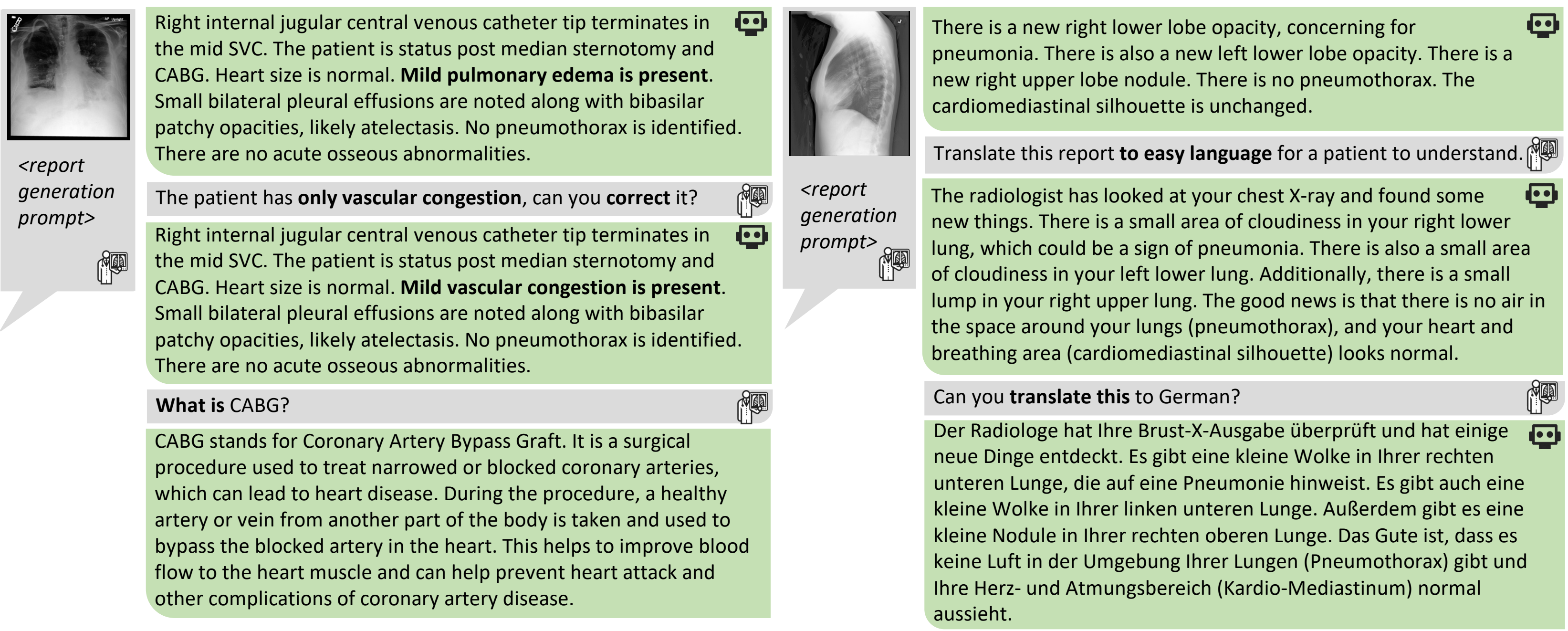
	MIMIC-CXR				IU-Xray (OOD)			
Method	FT	CE	BS	R-L	FT	CE	BS	R-L
LLaVA-Med	×	*10.7*	*0.19*	*15.1*	×	5.0	0.20	15.8
Rad-FM	✓	15.4	0.22	15.6	×	5.9	0.20	13.8
XrayGPT	✓	19.3	0.33	22.0	×	9.9	0.39	25.7
LLM-CXR	✓	21.1	-	-	-	-	-	-
CheXagent	✓	22.2	0.36	25.9	✓	*14.1*	*0.51*	*34.6*
R2GenGPT	✓	24.7	0.36	**27.6**	-	-	-	-
RaDialog$_{\text{align-rep}}$	✓	39.4	**0.40**	27.1	×	22.6	**0.47**	**31.0**
RaDialog$_{\text{align-ins}}$	✓	38.6	0.39	27.0	×	22.9	0.46	30.2
RaDialog$_{\text{project-rep}}$	✓	**39.7**	0.36	25.6	×	23.0	0.45	29.6
RaDialog$_{\text{project-ins}}$	✓	39.2	0.37	26.7	×	**23.1**	0.45	30.4

Comparison to SOTA report generation methods:

Method	CE	BS	B-4	MTR	R-L
R2Gen (Chen et al., 2020)	27.6	0.27	10.3	14.2	27.7
Kiut (Wang et al., 2022)	32.1	-	11.3	16.0	28.5
COMG (Gu et al., 2024a)	34.5	-	10.4	13.7	27.9
HKRG (Wang et al., 2025)	33.9	-	**14.3**	**16.7**	**31.0**
ORID (Gu et al., 2024b)	35.2	-	11.6	15.0	28.4
MPO (Xiao et al., 2024)	35.3	-	13.9	16.2	30.9
RaDialog-project-instruct	**39.2**	**0.37**	9.4	14.2	26.7

- Our architecture design leads to fine-grained image understanding and improved clinical correctness compared to prior methods.
- Further, RaDialog is the first LVLM to maintain SOTA report generation performance and outperforms all other methods significantly.

Conversational Downstream Tasks



Right internal jugular central venous catheter tip terminates in the mid SVC. The patient is status post median sternotomy and CABG. Heart size is normal. **Mild pulmonary edema is present**. Small bilateral pleural effusions are noted along with bibasilar patchy opacities, likely atelectasis. No pneumothorax is identified. There are no acute osseous abnormalities.

The patient has **only vascular congestion**, can you **correct** it?

Right internal jugular central venous catheter tip terminates in the mid SVC. The patient is status post median sternotomy and CABG. Heart size is normal. **Mild vascular congestion is present**. Small bilateral pleural effusions are noted along with bibasilar patchy opacities, likely atelectasis. No pneumothorax is identified. There are no acute osseous abnormalities.

What is CABG?

CABG stands for Coronary Artery Bypass Graft. It is a surgical procedure used to treat narrowed or blocked coronary arteries, which can lead to heart disease. During the procedure, a healthy artery or vein from another part of the body is taken and used to bypass the blocked artery in the heart. This helps to improve blood flow to the heart muscle and can help prevent heart attack and other complications of coronary artery disease.

There is a new right lower lobe opacity, concerning for pneumonia. There is also a new left lower lobe opacity. There is a new right upper lobe nodule. There is no pneumothorax. The cardiomediastinal silhouette is unchanged.

Translate this report **to easy language** for a patient to understand.

The radiologist has looked at your chest X-ray and found some new things. There is a small area of cloudiness in your right lower lung, which could be a sign of pneumonia. There is also a small area of cloudiness in your left lower lung. Additionally, there is a small lump in your right upper lung. The good news is that there is no air in the space around your lungs (pneumothorax), and your heart and breathing area (cardiomediastinal silhouette) looks normal.

Can you **translate this** to German?

Der Radiologe hat Ihre Brust-X-Ausgabe überprüft und hat einige neue Dinge entdeckt. Es gibt eine kleine Wolke in Ihrer rechten unteren Lunge, die auf eine Pneumonie hinweist. Es gibt auch eine kleine Wolke in Ihrer linken unteren Lunge. Außerdem gibt es eine kleine Nodule in Ihrer rechten oberen Lunge. Das Gute ist, dass es keine Luft in der Umgebung Ihrer Lungen (Pneumothorax) gibt und Ihre Herz- und Atmungsbereich (Kardio-Mediastinum) normal aussieht.

Performance on Downstream Tasks:

Task	Metric	Comp. Method	Comp. Result	RaDialog
Report Correction	CE improvement	XrayGPT	10.0	**33.4**
Finding Prediction	F1	XrayGPT	20.6	**39.7**
Impression Generation	R-L	CheXagent	40.3	**45.8**
View Classification	Accuracy	CheXagent	**97.5**	95.9

- Our instruct dataset and training setup successfully enables dialog-based assistance by avoiding catastrophic forgetting and teaching domain-specific conversational tasks.
- In a human evaluation comparing RaDialog against XRayGPT, the radiologist preferred RaDialog in 84% of cases for report generation and 71% in conversational tasks.

Code: Paper: