

# Broca: a Multimodal Language Model to Draft Radiology Reports for X-rays

Leveraging multimodal prompt tuning and large language model for Automated Radiology Reporting from X-ray Imagery.

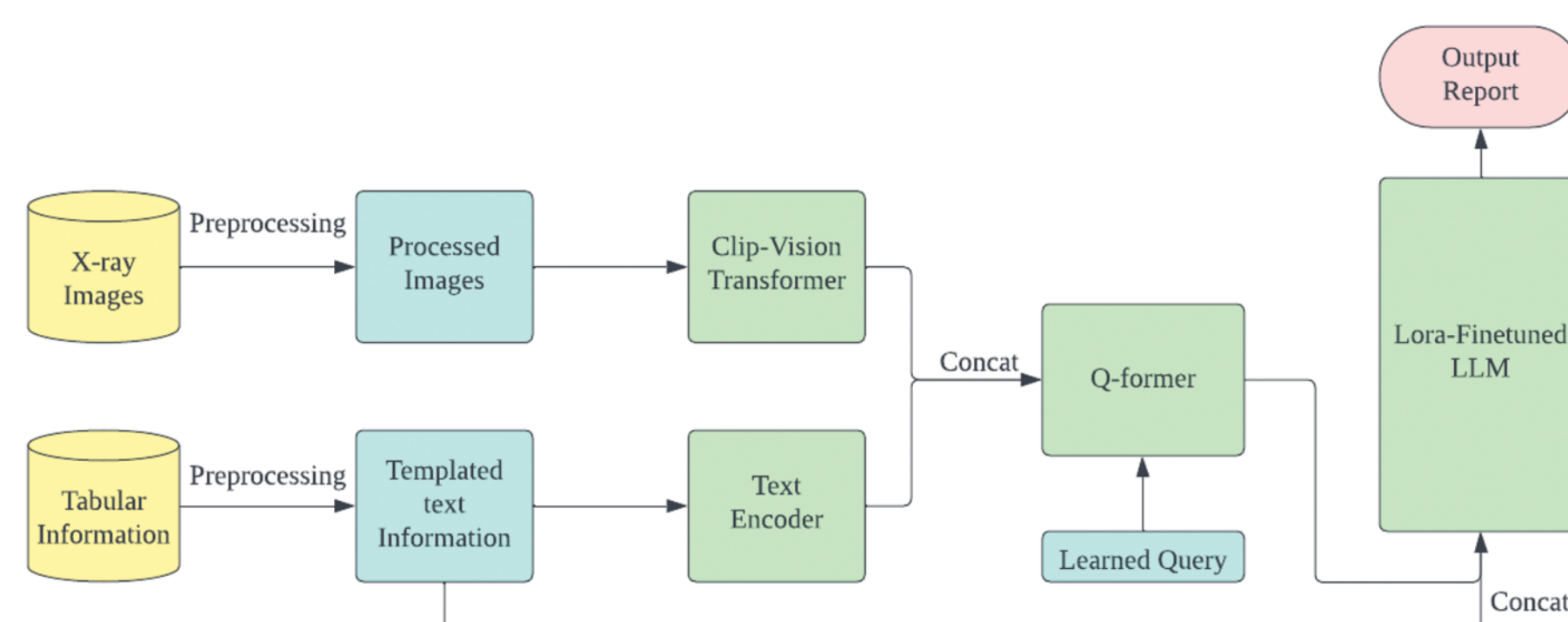
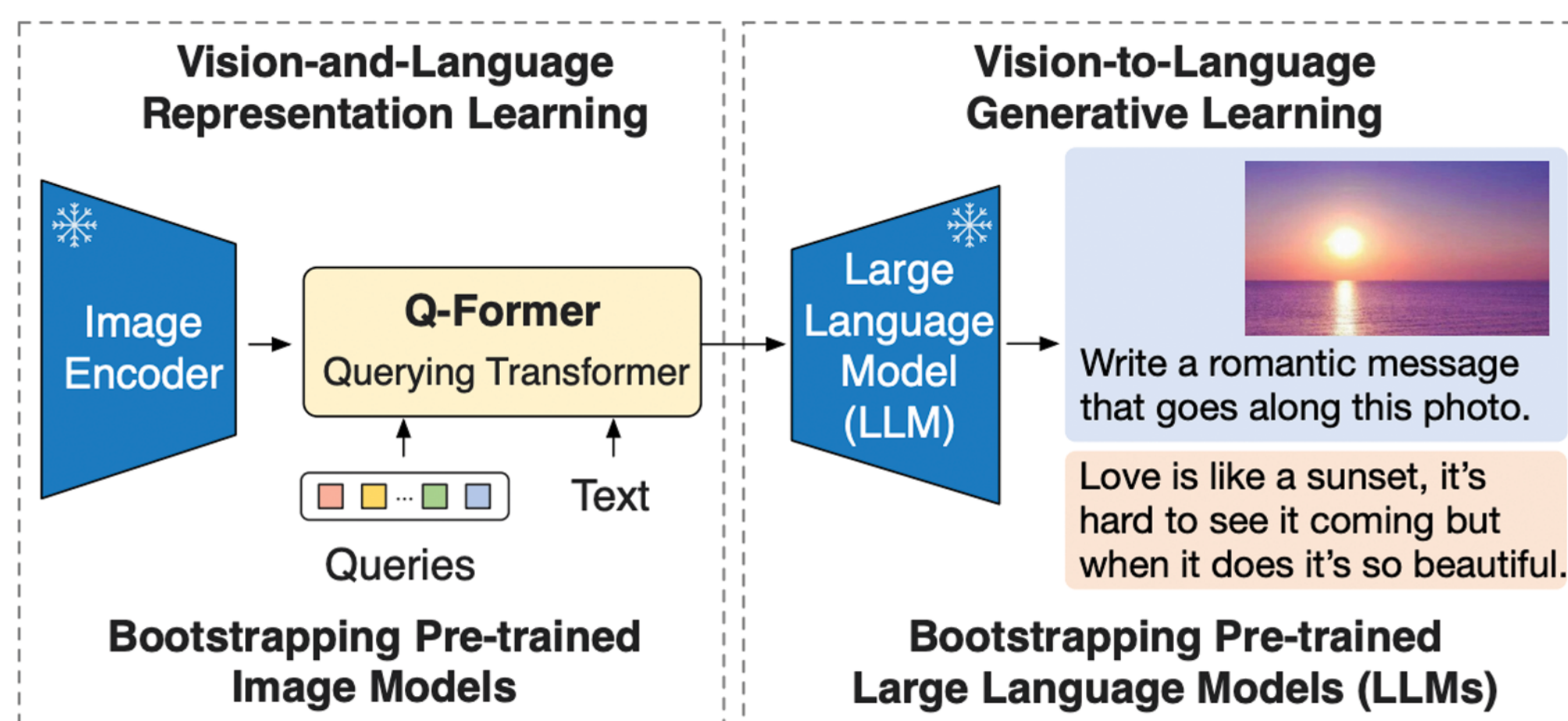
**Yuxuan Wang**

**Farzad Khalvat**

ACADEMIC SUPERVISOR

**Mark Cicero and Alexander Bilbily**

INDUSTRY SUPERVISORS



## PROJECT SUMMARY

Our project is focused on developing a model that can produce radiology reports from X-ray images and patients' information. We utilized a multimodal encoder combined with a Q-former to generate prompts from both text and X-ray images, achieving a cross-modality approach. The multimodal encoder comprises a text encoder initialized from BERT base and an vision transformer initialized from CLIP-384. We incorporated the Q-former based on insights from the BLIP-2 paper, which is pretrained through two stages with four tasks. This Q-former can extract embeddings from the CLIP modality and create soft prompts suitable for large language models. We then linked these prompts to a large language model, finetuning it using techniques like quantization and low-rank adaptation (LoRA). This allowed us to tune the prompts and adapt the model to radiology-specific needs. After extensive hyper-parameter experiments and adjustments, we settled on our model's structure. Currently, we're implementing alignment using reinforcement learning through human feedback (RLHF) to minimize inaccuracies and reduce the problem of hallucination. Preliminary tests show that our model performs well on our internal company X-ray datasets.

## REFERENCES

- [1] Li, Junnan, Dongxu Li, Silvio Savarese and Steven C. H. Hoi. "BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models." ArXiv abs/2301.12597 (2023): n. pag.
- [2] Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry et al. "Learning transferable visual models from natural language supervision." In International conference on machine learning, pp. 8748-8763. PMLR, 2021.
- [3] Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. "Lora: Low-rank adaptation of large language models." arXiv preprint arXiv:2106.09685 (2021).