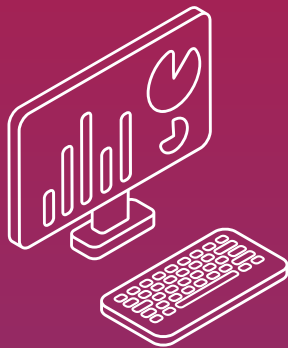


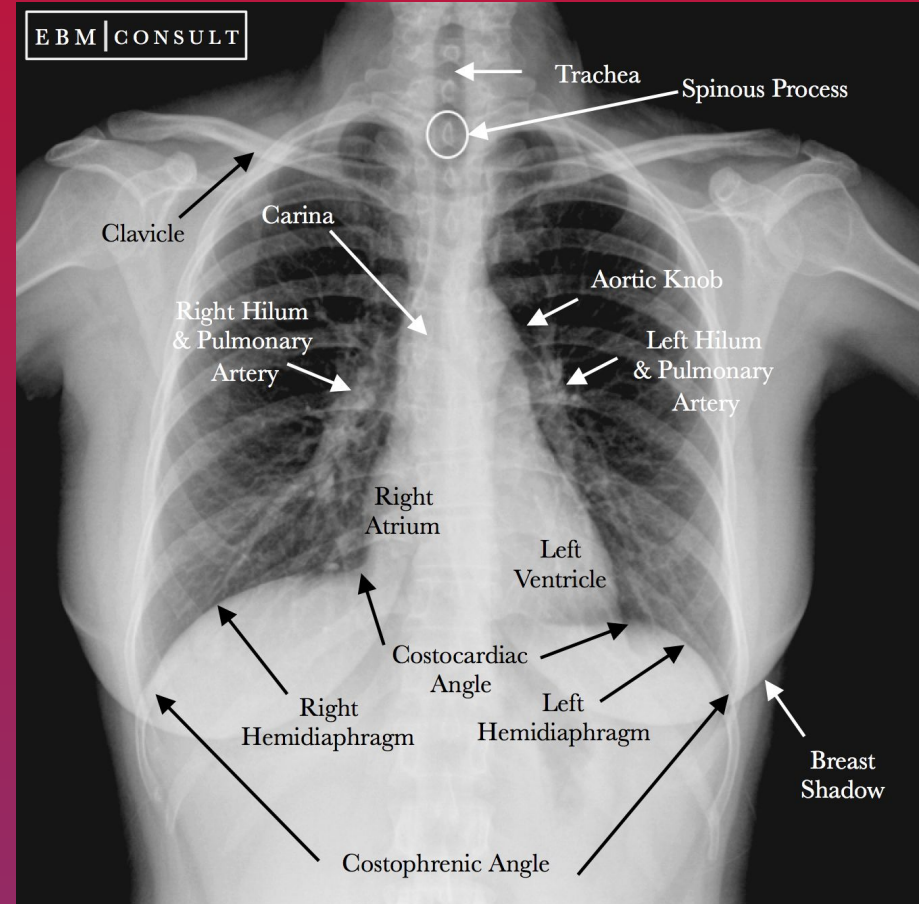
Region-guided Radiology Report Generation



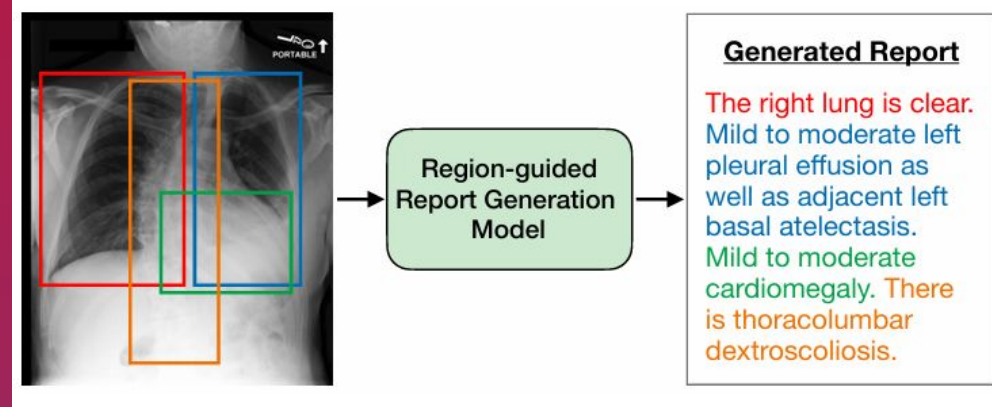
Parsa Sharifi

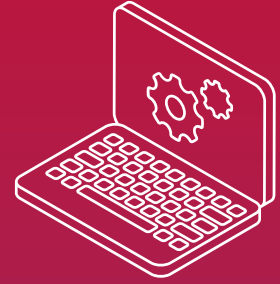


- Chest radiography is widely used
- Whole picture

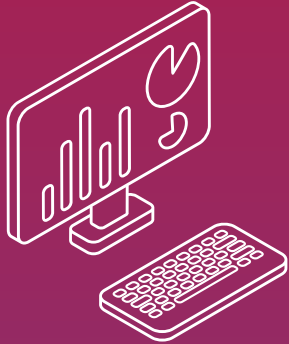


- Region-Guided Radiology Report Generation (RGRG)
- detects anatomical regions
- generates individual descriptions

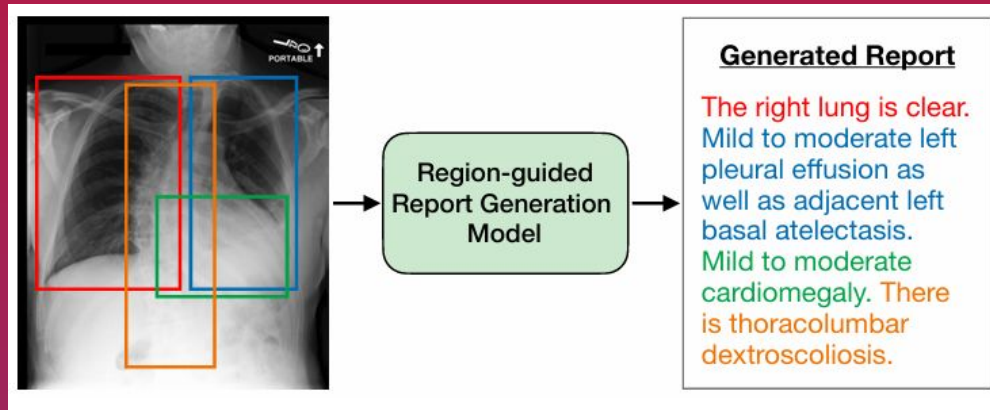


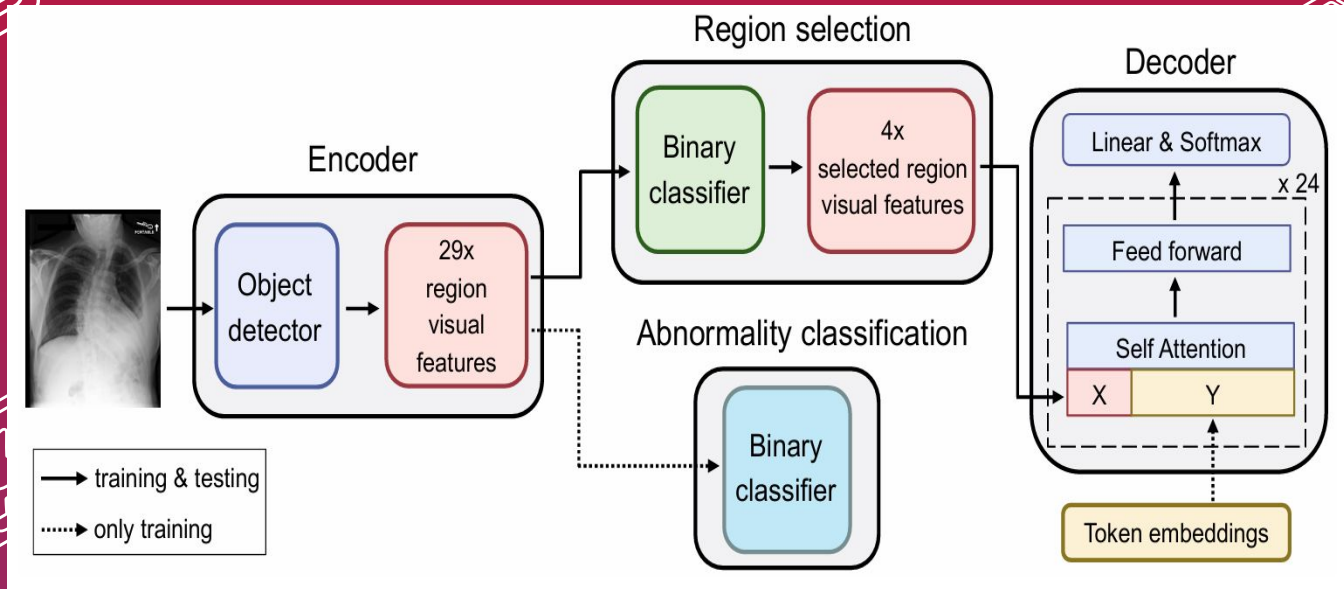


Method!

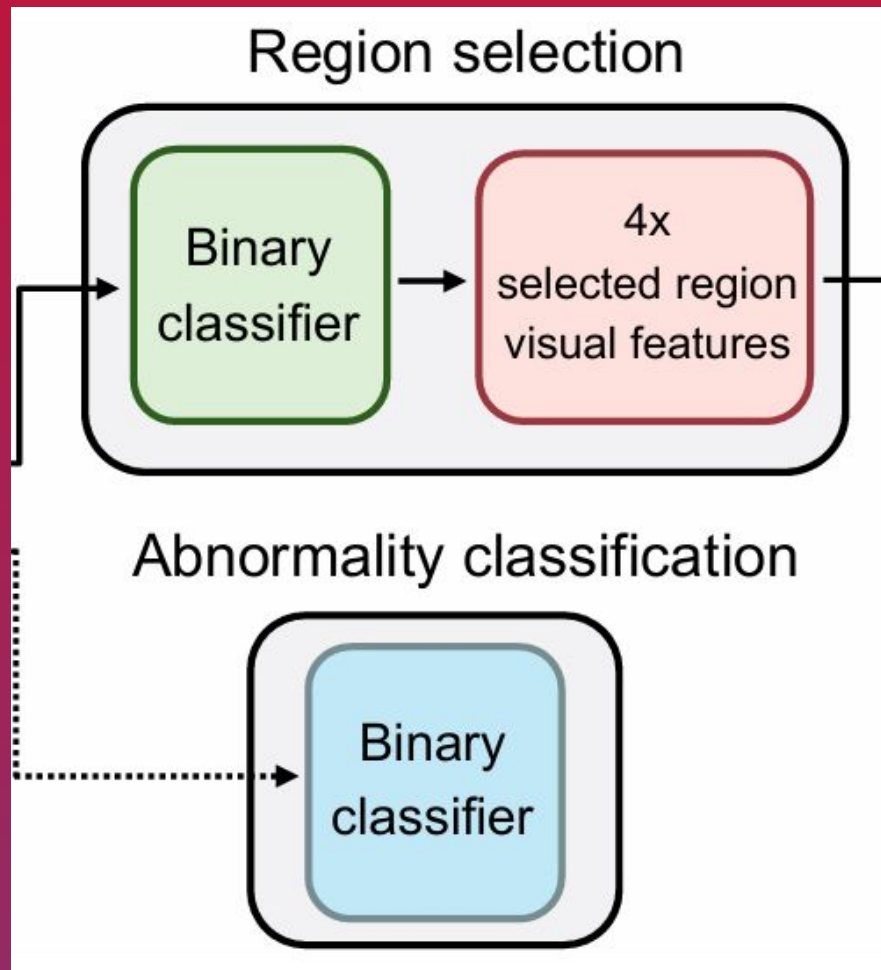


- Pre-trained language model on anatomical region independently.
- Anatomy-based sentence generation
- Flexible use for radiologists
- Selection-based

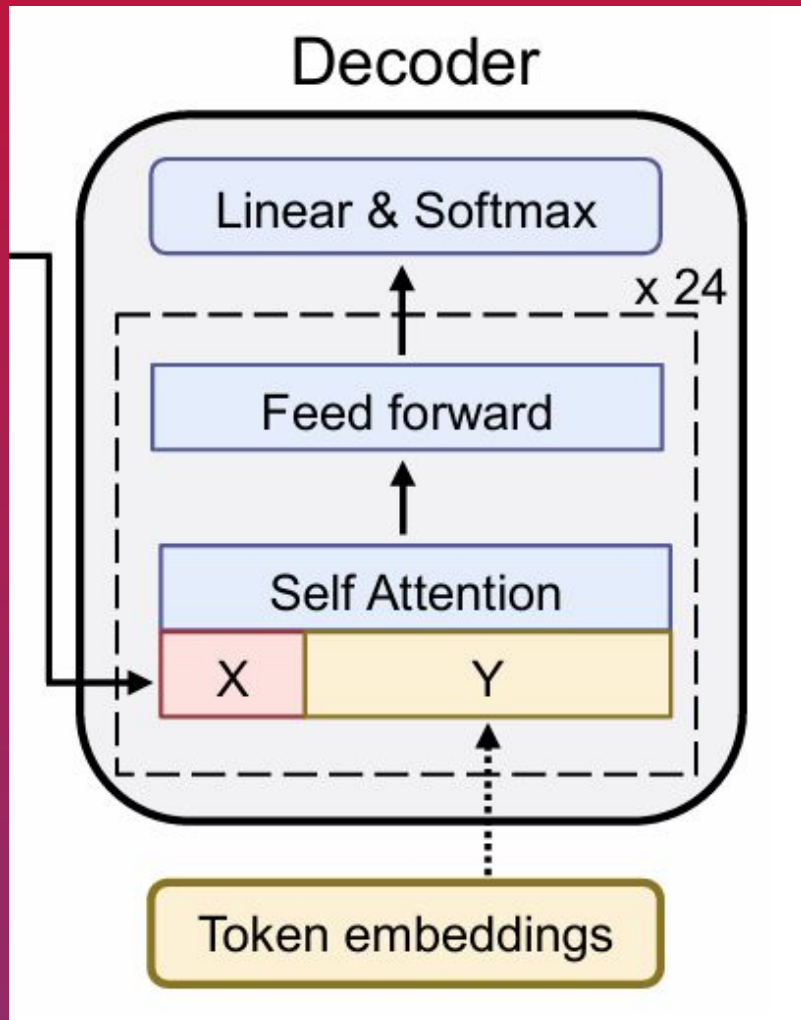




- Object detector extracts features
- For 29 unique anatomical regions
- Binary classifiers select region



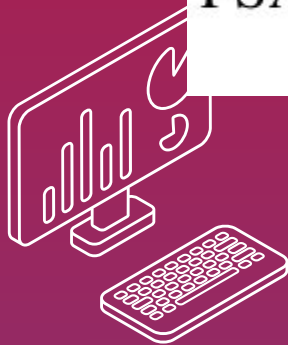
- 355M-parameter model GPT-2 Medium
- Fine-tuned on PubMed abstracts
- Based on self-attention



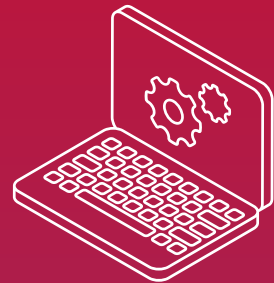


$$\text{SA}(Y) = \text{softmax}((YW_q)(YW_k)^\top)(YW_v),$$

$$\text{PSA}(X, Y) = \text{softmax} \left((YW_q) \begin{bmatrix} XU_k \\ YW_k \end{bmatrix}^\top \right) \begin{bmatrix} XU_v \\ YW_v \end{bmatrix}$$



- Y represents the token embedding
- W_q , W_k , W_v are the query, key, value projection
- X is the region visual features
- U_k and U_v are the corresponding (newly initialized) key and value projection parameters.



$$\mathcal{L} = \lambda_{\text{obj}} \cdot \mathcal{L}_{\text{obj}} + \lambda_{\text{select}} \cdot \mathcal{L}_{\text{select}} \\ + \lambda_{\text{abnormal}} \cdot \mathcal{L}_{\text{abnormal}} + \lambda_{\text{language}} \cdot \mathcal{L}_{\text{language}}$$

