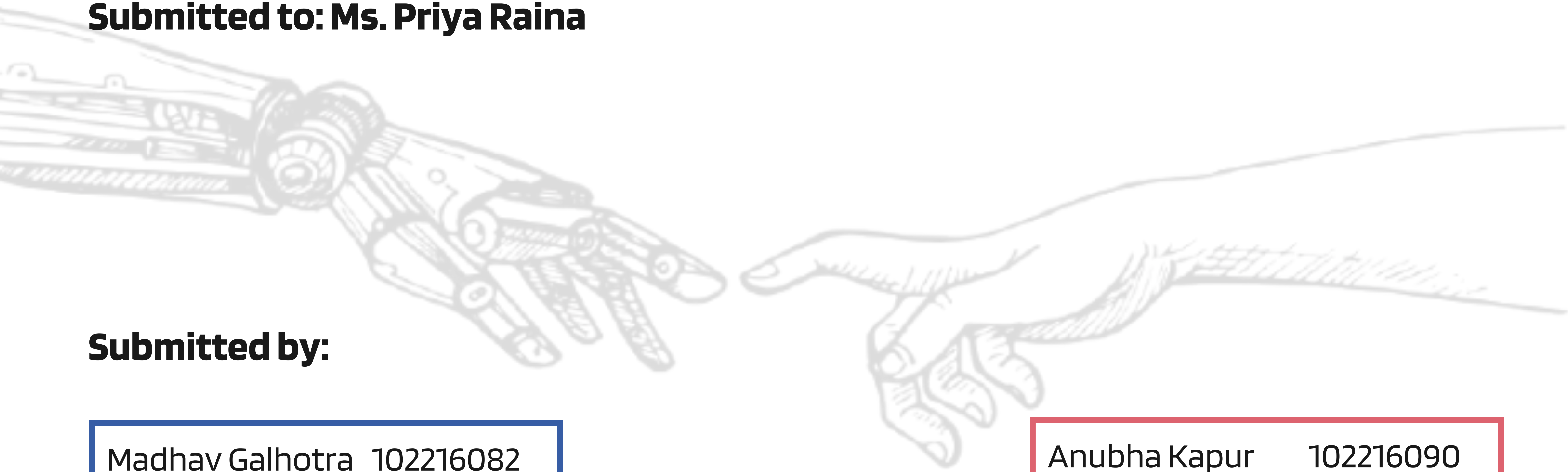
# Automatic Radiology Report Generator

**Submitted to: Ms. Priya Raina**



**Submitted by:**


Madhav Galhotra 102216082

Anubha Kapur 102216090

# Project Overview

01 Image feature extraction using a **Vision Transformer (vit_base_patch16_224)**

02 Nearest-neighbor retrieval of similar radiographs via **cosine similarity**

03 Prompt generation using **MeSH terms and findings**

04 Report synthesis using a **fine-tuned DistilGPT-2 model**

05 Performance evaluation through **BLEU, ROUGE, BERTScore, perplexity, and embedding visualizations**

# Need Analysis

1

**Increasing Diagnostic Workload**

Hospitals and diagnostic centers face a rapidly growing number of imaging studies, especially chest X-rays. Radiologists often handle hundreds of cases per day, leading to high cognitive load,report delays and risk of fatigue-related errors, An automated system can help reduce turnaround time and support radiologists in managing large case volumes.

2

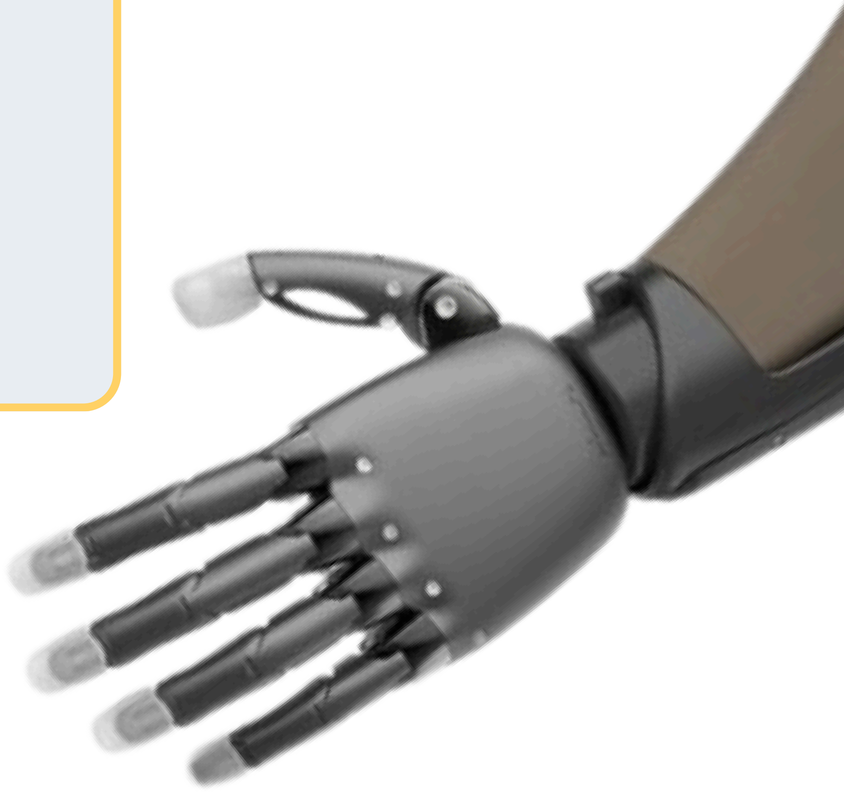
**Shortage of Trained Radiologists**

Many regions, especially developing countries, face a shortage of qualified radiologists. This gap leads to: delayed diagnoses, uneven quality of care and increased burden on existing professionals. A reliable AI-based system can offer decision support or preliminary reporting in such resource-constrained settings.

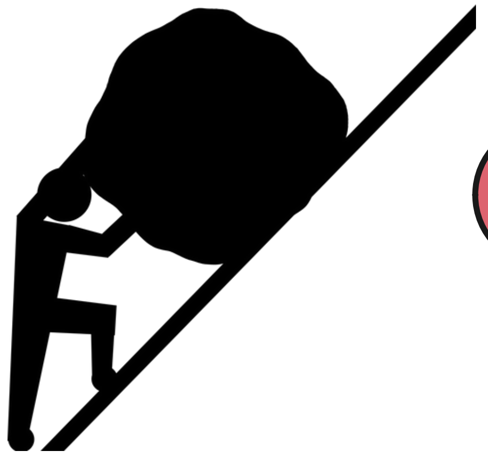
3

**Lack of Unified Vision-Language Frameworks**

Existing systems handle either image interpretation (vision models), or text generation (language models) but very few combine both effectively. A unified model is needed to bridge the gap between: pixel-level patterns in chest X-rays and high-level clinical reasoning expressed in radiology reports.



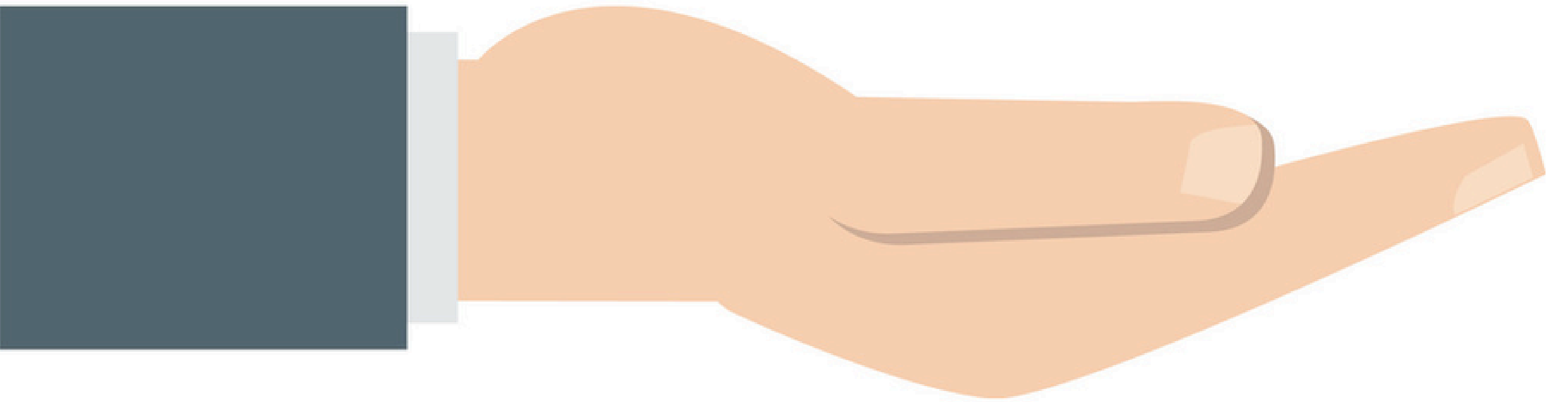
# Problem Statement



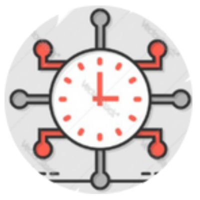
01 Existing ML models can classify abnormalities in chest X-rays but cannot generate reports.

02 Vision-only models fail to convert image features into coherent medical language, while language-only models lack grounding in the actual X-ray, leading to inaccurate or irrelevant reports..

03 Need for consistency,  standardization, and reduced variability



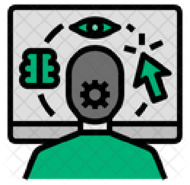
# Objectives



01 Generate complete, human-like radiology reports.

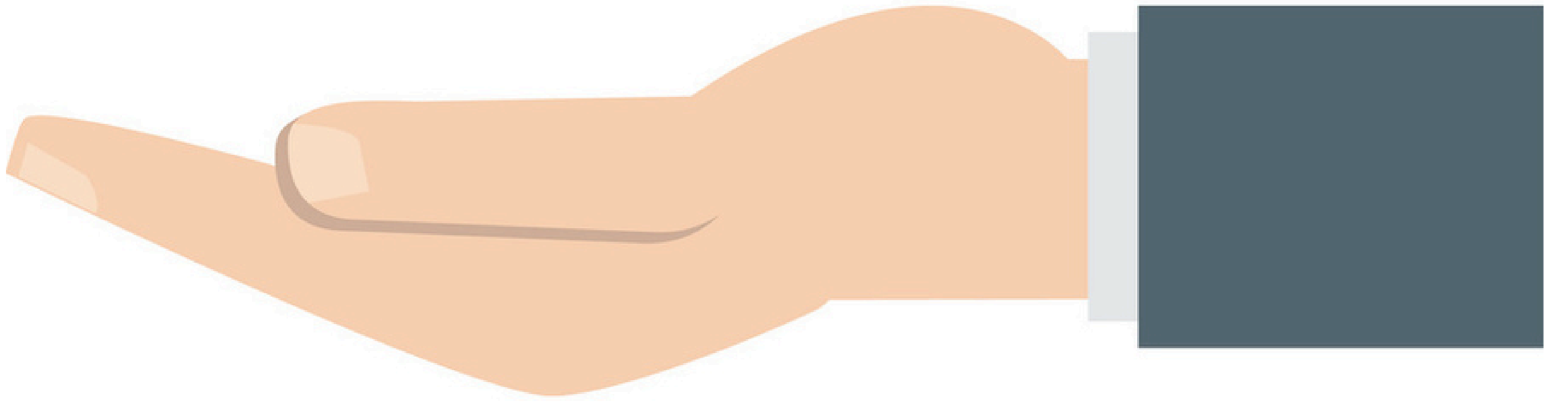
02 Extract deep visual features using a Vision Transformer

03 Construct meaningful prompts using MeSH terms

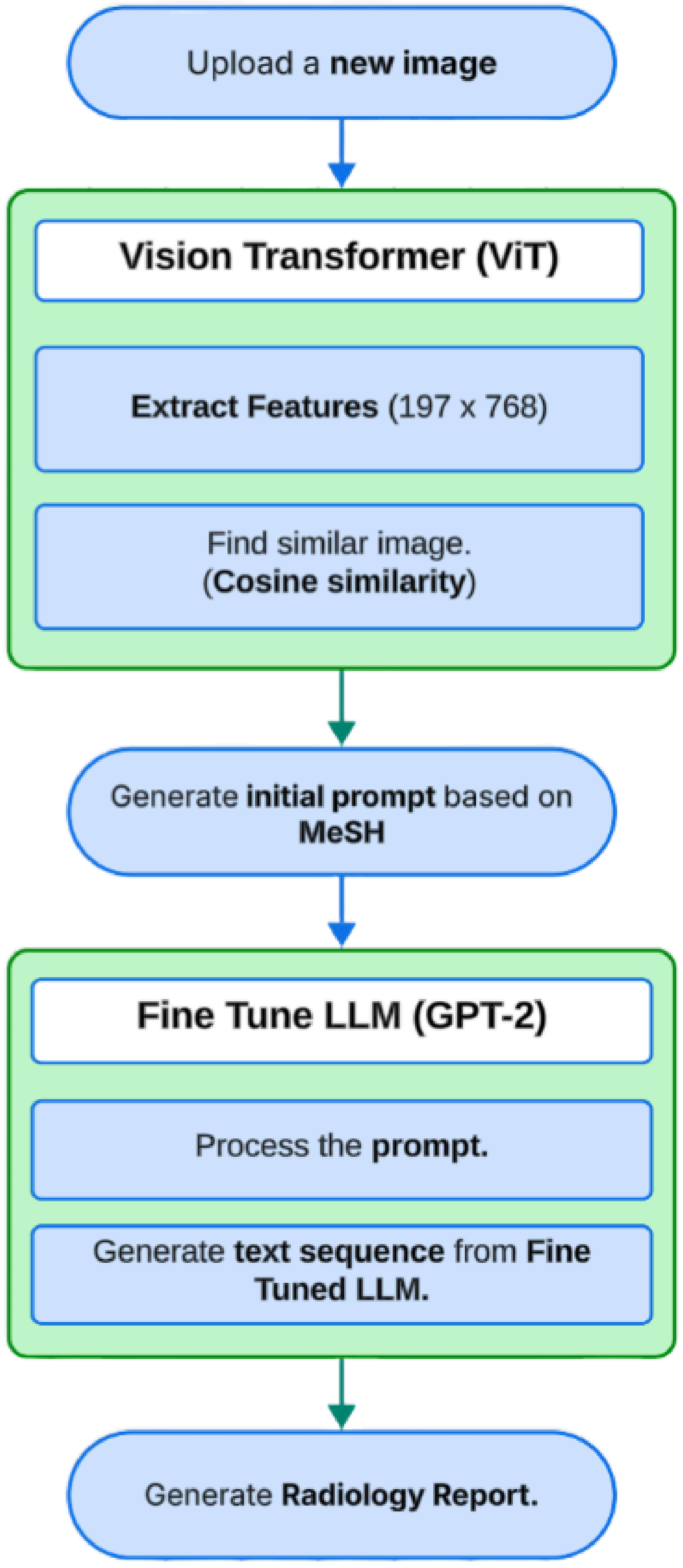


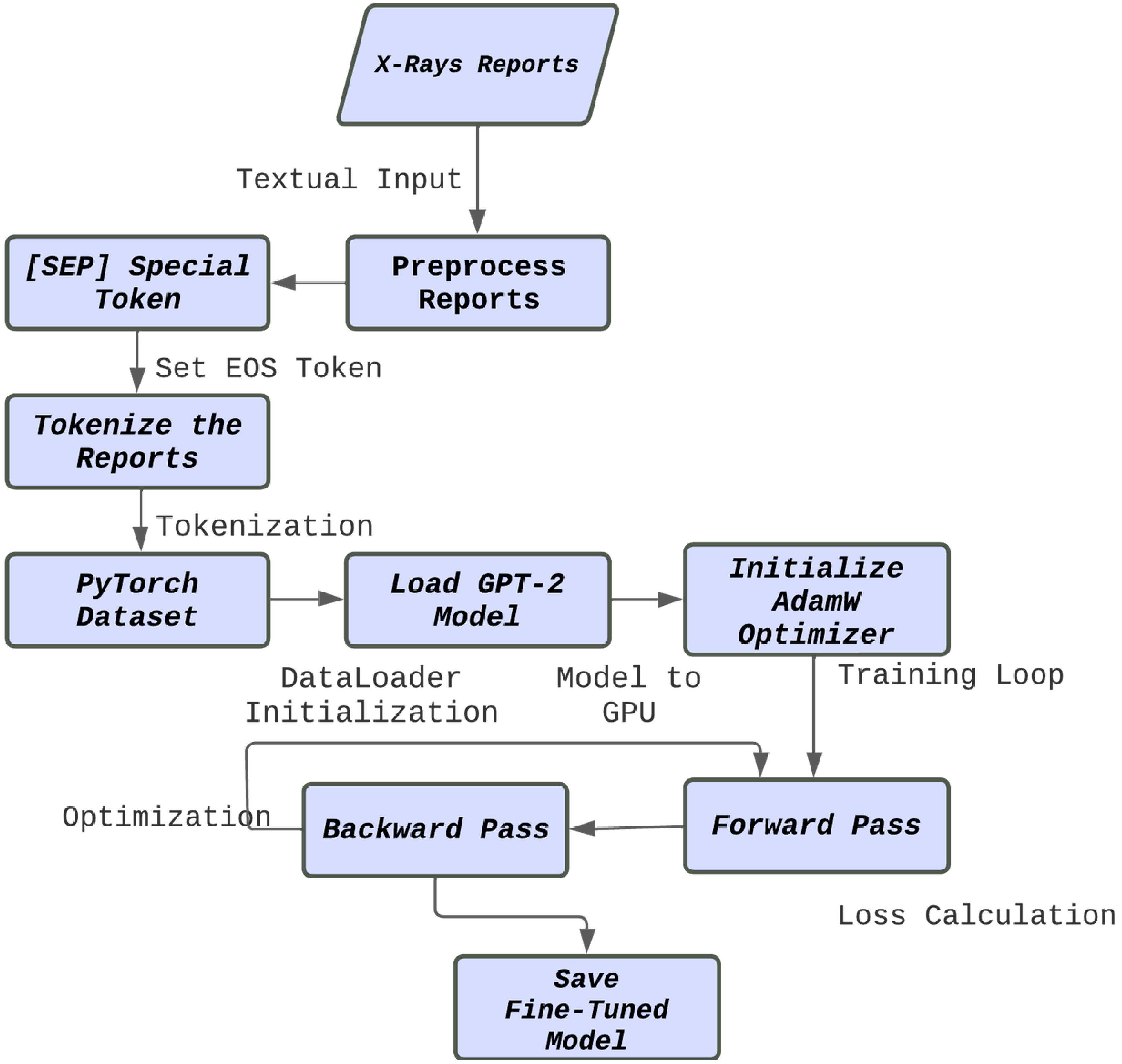
04 Fine-tune a DistilGPT-2 language model

05 Develop a simple inference interface



# System Architecture



# Results

## Fine-Tuning Performance (Loss)

```
Using 1000 reports for fine-tuning
Using device: mps
Model loaded
Epoch: 1
Epoch 1/20, Loss: 0.8700
Epoch: 2
Epoch 2/20, Loss: 0.4816
Epoch: 3
Epoch 3/20, Loss: 0.3936
Epoch: 4
Epoch 4/20, Loss: 0.3298
Epoch: 5
Epoch 5/20, Loss: 0.2764
Epoch: 6
Epoch 6/20, Loss: 0.2342
Epoch: 7
Epoch 7/20, Loss: 0.1983
Epoch: 8
Epoch 8/20, Loss: 0.1695
Epoch: 9
Epoch 9/20, Loss: 0.1455
Epoch: 10
Epoch 10/20, Loss: 0.1289
Epoch: 11
Epoch 11/20, Loss: 0.1130
Epoch: 12
Epoch 12/20, Loss: 0.1016
Epoch: 13
Epoch 13/20, Loss: 0.0908
Epoch: 14
Epoch 14/20, Loss: 0.0837
Epoch: 15
Epoch 15/20, Loss: 0.0767
Epoch: 16
Epoch 16/20, Loss: 0.0699
Epoch: 17
Epoch 17/20, Loss: 0.0675
Epoch: 18
Epoch 18/20, Loss: 0.0639
Epoch: 19
Epoch 19/20, Loss: 0.0581
Epoch: 20
Epoch 20/20, Loss: 0.0555
Training complete. Model saved.
```

### Evaluation Metrics

**BLEU Score:** 0.2698

**ROUGE Scores:**

| Metric | Score |
|---|---|
| ROUGE-1 F1 | 0.7368 |
| ROUGE-2 F1 | 0.3529 |
| ROUGE-L F1 | 0.6315 |

**BERTScore:**

| Metric | Score |
|---|---|
| Precision | 0.943 |
| Recall | 0.963 |
| F1 | 0.953 |

# Constraints

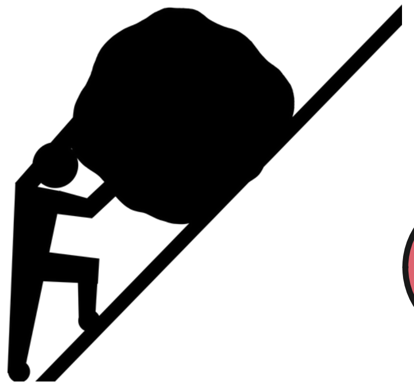
01

**Limited Computational Resources:**
Training ViT and GPT-2 requires high-performance GPUs; running them on CPU leads to significantly slower inference and training times.

02

**Single-Modality Dependence:** The system relies only on chest X-rays; lack of multi-modal clinical data (history, labs, vitals) may limit diagnostic accuracy.

# Future Scope



01

**Fine-tune on larger and more diverse radiology datasets:**
to improve robustness and clinical accuracy.

02

**Integrate multimodal data** such as CT/MRI scans, patient history, and lab reports, for more comprehensive diagnostic reporting.

03

**Introduce continual learning mechanisms** so the model adapts to new hospital data or evolving clinical guidelines without full retraining.

# Thank You!