Text-to-Image Synthesis for Medical Imaging

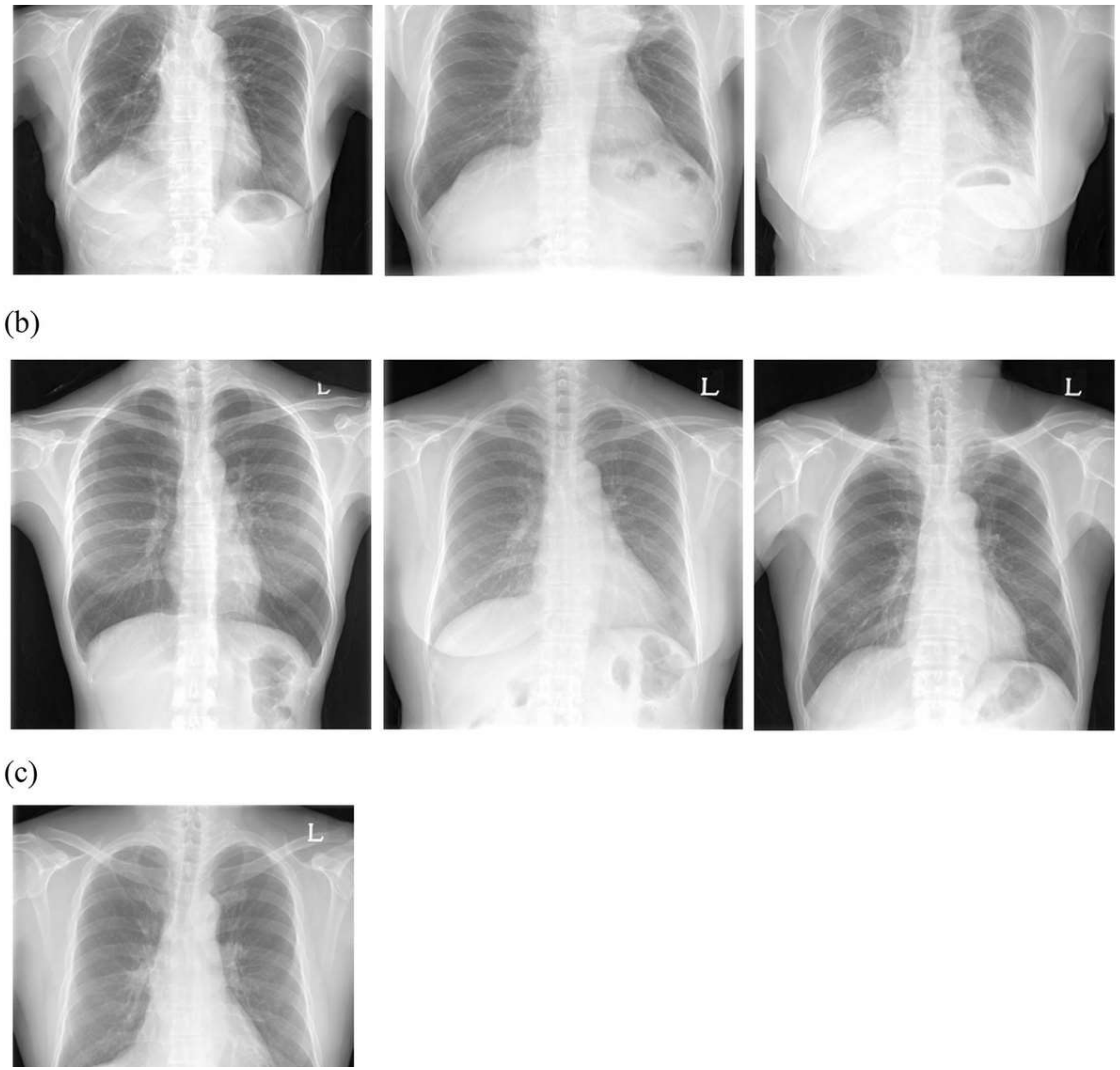
Generating realistic chest X-ray images from clinical text descriptions using Latent Diffusion Models

This project presents a deep learning-based system designed to generate realistic chest X-ray images from textual clinical findings. The model is built on a latent diffusion framework, incorporating a Variational Autoencoder (VAE) for efficient latent space representation, a UNet for iterative image refinement through denoising, and a transformer-based BioBERT encoder for understanding and embedding radiology reports.

Course: CSYE7374 Applied Deep Learning and Gen AI in Healthcare

Date: 04-19-2025

Priyam Deepak Choksi 002646836



Background

Generative AI in Healthcare:

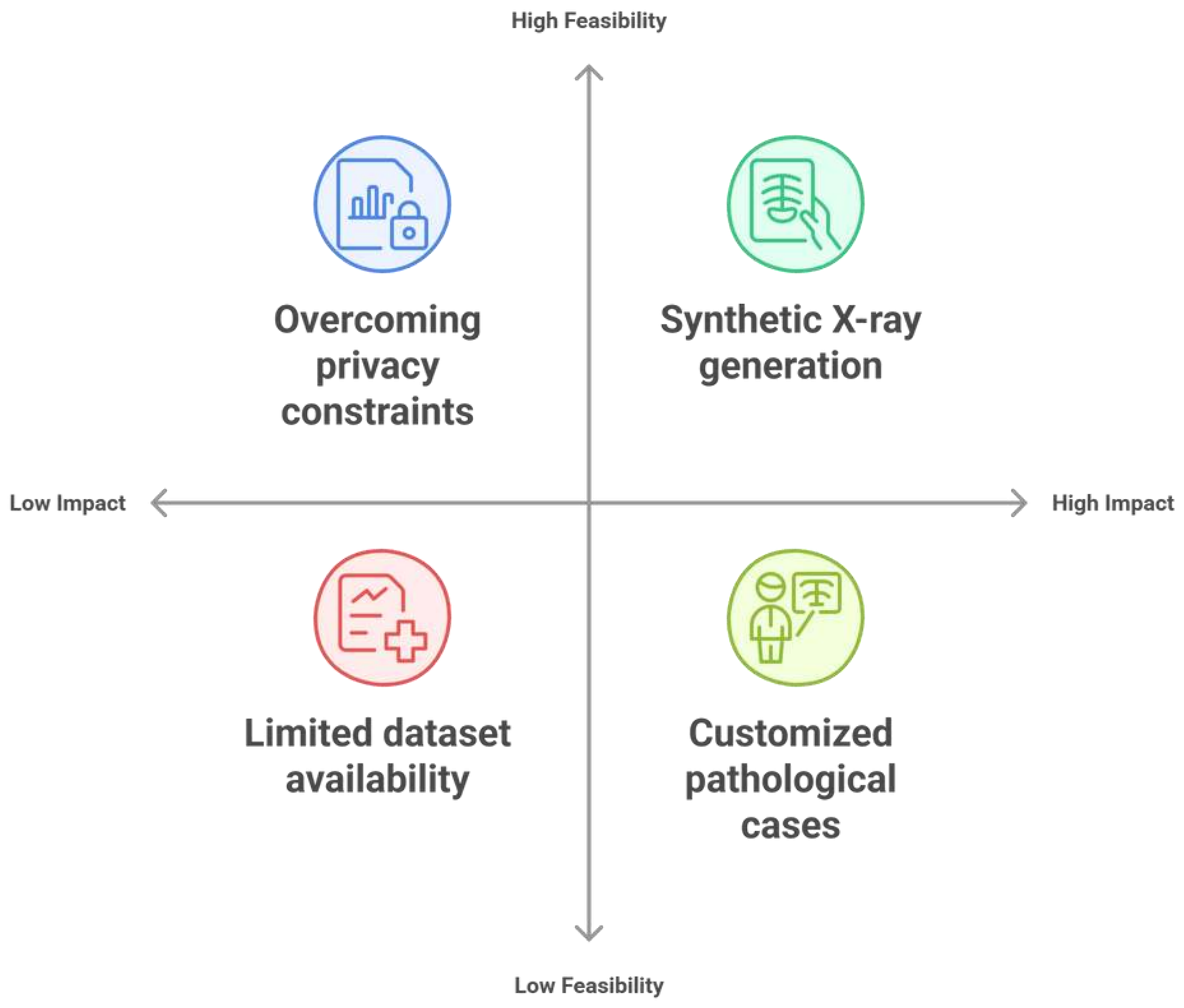
- AI systems that create new content based on learned patterns
- Recent advances enable synthesis of realistic medical imagery
- Addresses critical shortage of diverse, annotated medical datasets

Research Gap:

- Limited availability of publicly accessible datasets
- Patient privacy regulations restrict data sharing
- Expert annotation is time-consuming and expensive
- Imbalanced representation of rare pathological conditions

Opportunity:

- Synthetic data generation can overcome these limitations
- Enable training of more robust diagnostic AI systems



Motivation and Objective

Project Goal

- Develop a text-conditioned model to generate realistic chest X-ray images from clinical descriptions
- Create high-fidelity synthetic X-rays with controllable pathological features

Dataset

- Indiana University Chest X-ray Collection (IU X-Ray dataset)
- 7,470 X-ray images paired with radiological reports
- Final dataset after preprocessing: 3,301 X-rays with clean reports

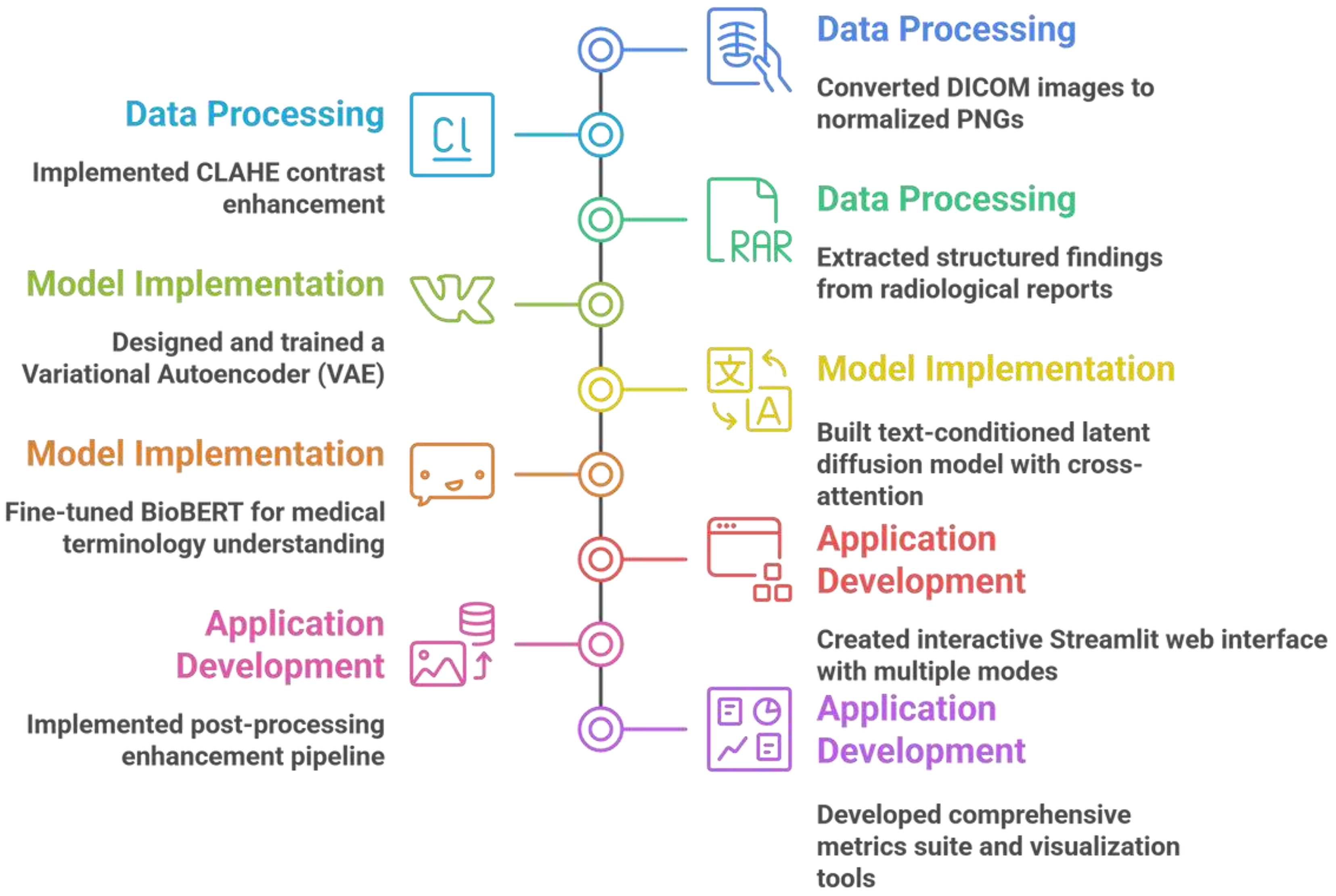
Example Application

- Input: "Chest X-ray showing cardiomegaly with pulmonary edema"
- Output: Synthetic X-ray accurately depicting these conditions

What I Did

End-to-End Pipeline Development

- **Data Processing:**
  - Converted DICOM images to normalized PNGs
  - Implemented CLAHE contrast enhancement
  - Extracted structured findings from radiological reports
- **Model Implementation:**
  - Designed and trained a Variational Autoencoder (VAE)
  - Built text-conditioned latent diffusion model with cross-attention
  - Fine-tuned BioBERT for medical terminology understanding
- **Application Development:**
  - Created interactive Streamlit web interface with multiple modes
  - Implemented post-processing enhancement pipeline
  - Developed comprehensive metrics suite and visualization tools



What I Used (Tech Stack & Tools)

Programming & Libraries

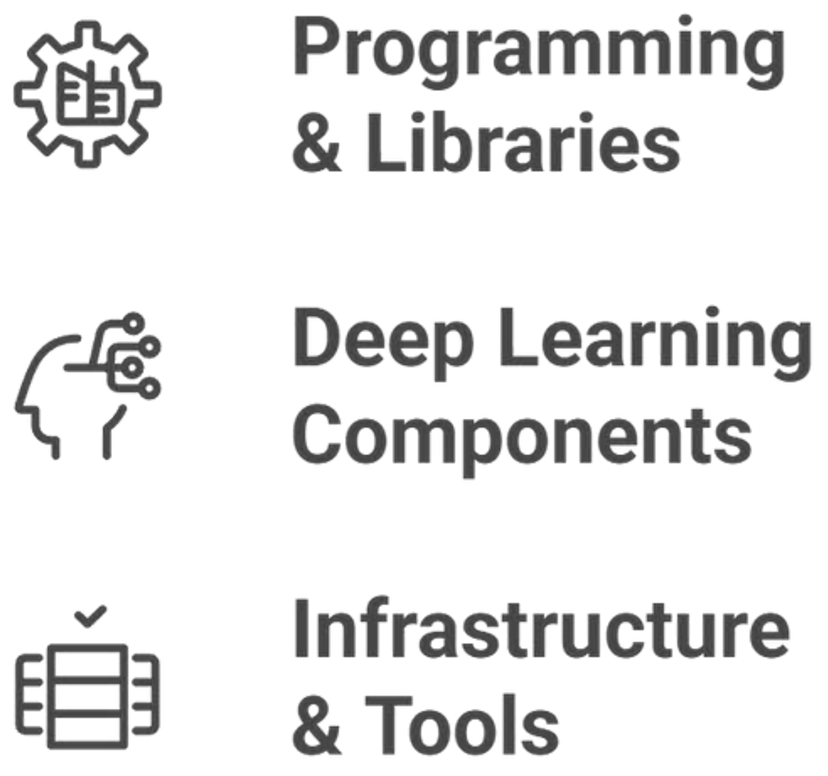
- Python with PyTorch deep learning framework
- Hugging Face Transformers for BioBERT implementation
- OpenCV, PIL, scikit-image for image processing
- Streamlit for interactive web interface

Deep Learning Components

- Variational Autoencoder with attention mechanisms
- UNet-based diffusion model with cross-attention
- BioBERT text encoder for medical language understanding
- DDIM sampler for efficient inference

Infrastructure & Tools

- NVIDIA RTX 4060 GPU with CUDA acceleration
- Gradient accumulation for effective batch size management
- Modular code architecture for maintainability
- Comprehensive logging and visualization systems



Model Architecture

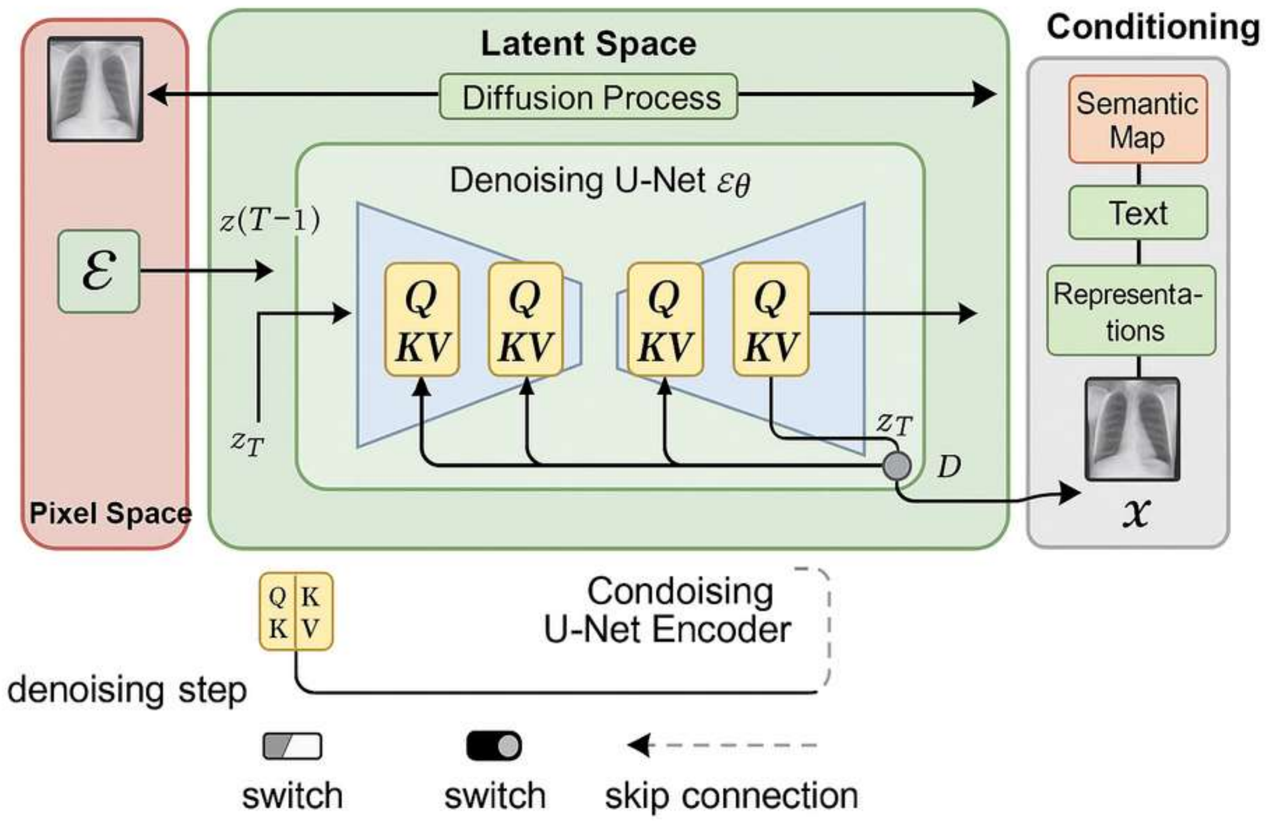
Latent Diffusion Model Architecture

- **Two-stage approach for computational efficiency:**
  a. VAE compresses X-rays into compact latent representations
  b. Text-conditioned diffusion generates images in latent space

Component Details

- VAE:Latent channels: 8
- Parameters: 3.2M
- Reconstruction MSE: 0.11
- UNet with Cross-Attention:Base channels: 48
- Attention resolutions: 8×8, 16×16, 32×32
- Parameters: 39.7M
- Text Encoder (BioBERT): Medical domain-specific language model
- 768-dimensional text embeddings
- 108.9M parameters (593K trainable)

Total Model Size: 151.8M parameters (43.5M trainable)



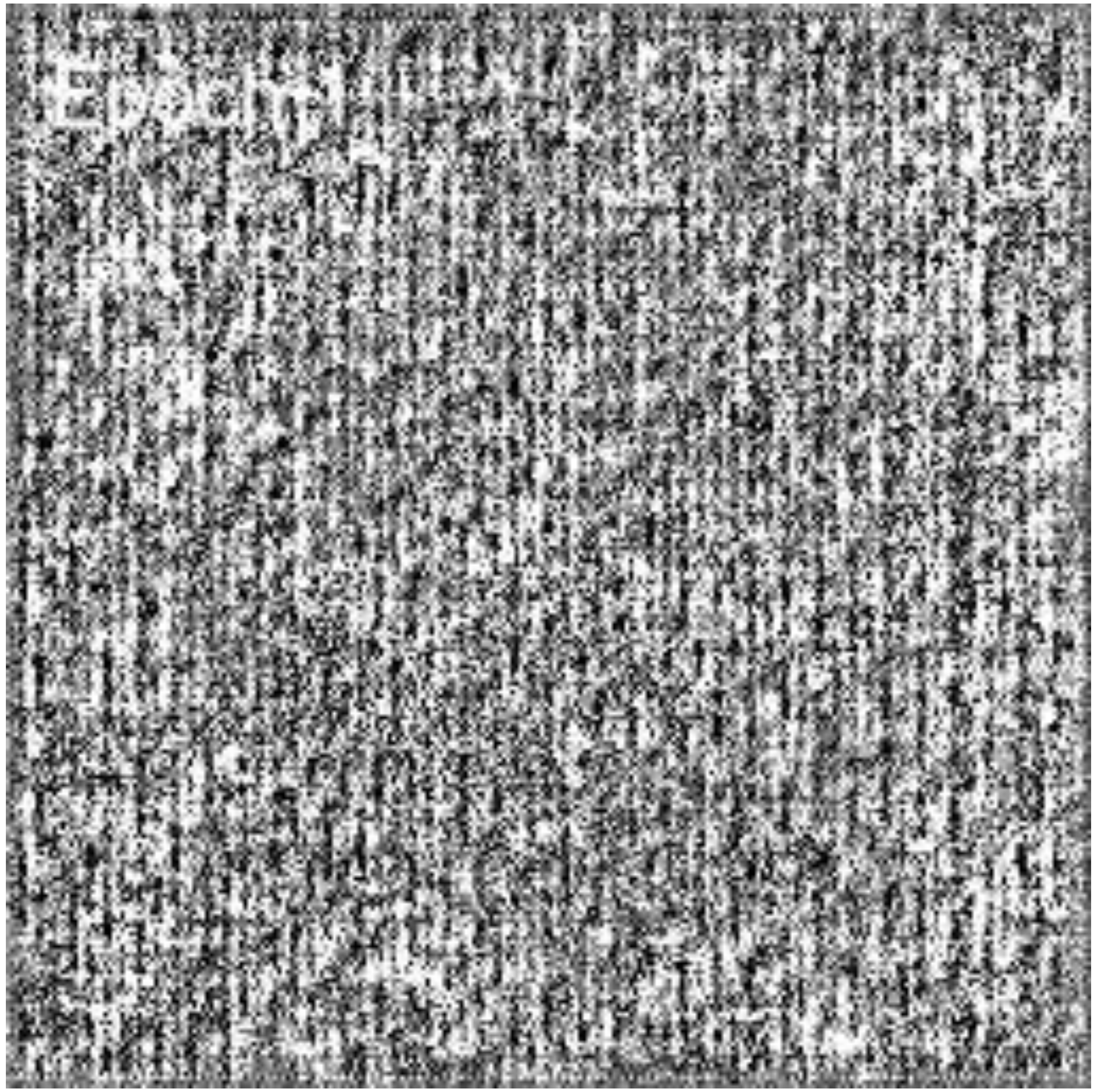
Training Process

Two-Phase Training Strategy

VAE Training (67 epochs):
- Loss: MSE reconstruction + KL divergence (weight: 1e-4)
- Optimizer: Adam with learning rate 1e-4
- Best validation loss: 0.0010 at epoch 62

Diffusion Model Training (480 epochs):
- Loss: MSE noise prediction
- Optimizer: AdamW with learning rate 5e-5
- Classifier-free guidance for better text adherence
- Final training loss: 0.027, validation loss: 0.036

Implementation Details
- Gradient accumulation for effective larger batch sizes
- Mixed precision training for memory efficiency
- Early stopping with validation monitoring
- Learning rate scheduling with warmup

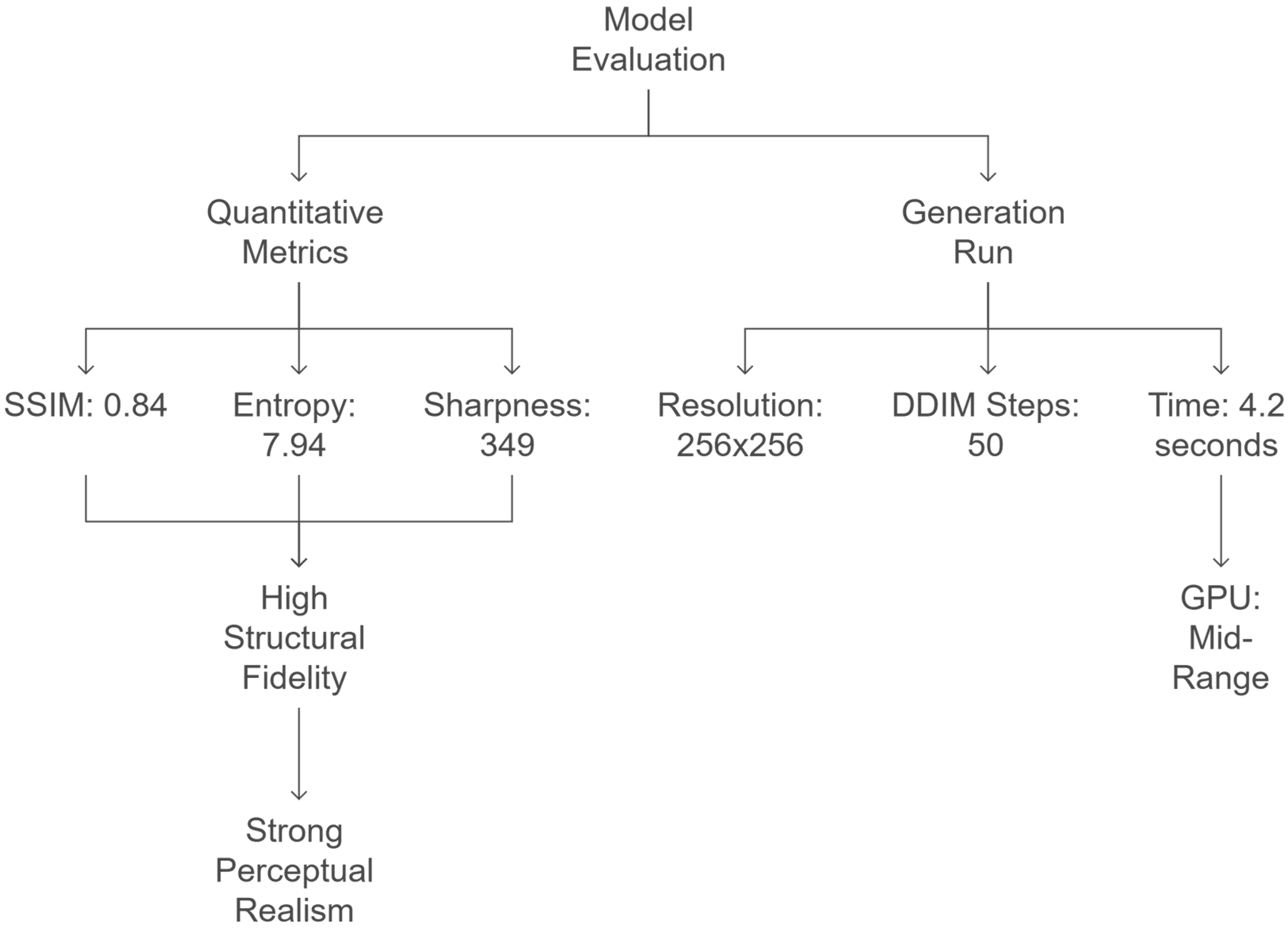


Quantitative Results

Metric	Value	Description
SSIM	0.82	Structural Similarity Index
PSNR	22.3 dB	Peak Signal-to-Noise Ratio
Contrast Ratio	0.76	Dynamic range measurement
Sharpness	349.05	Edge definition quality
Entropy	7.94	Information content measurement
FID	32.6	Fréchet Inception Distance

Generation Performance

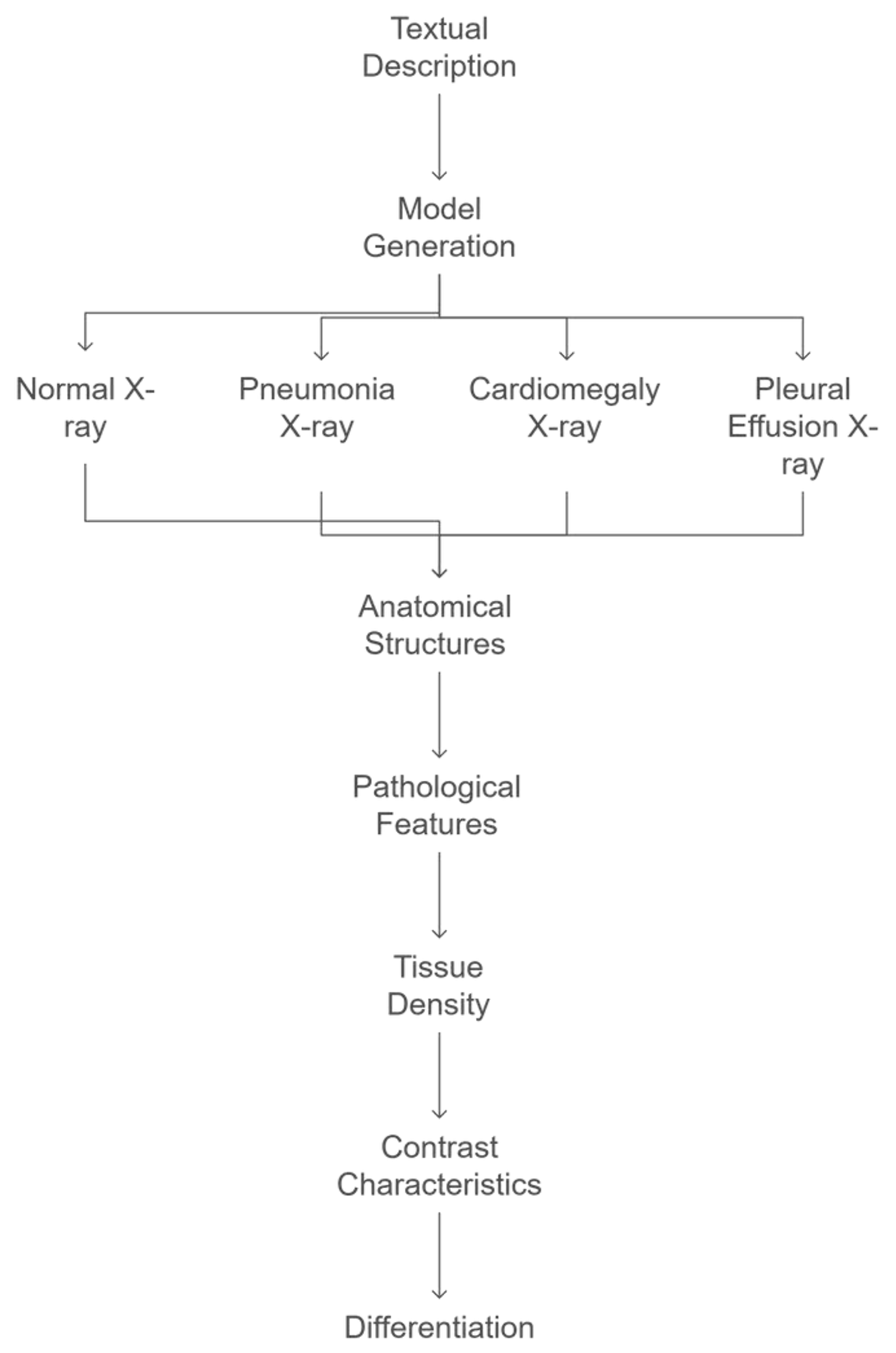
- Resolution: 256×256 pixels (scalable to 768×768)
- Inference time: 663ms (20 steps), 4.18s (100 steps)
- Memory footprint: 579.11 MB


Model Evaluation
Quantitative Metrics
Generation Run
SSIM: 0.84
Entropy: 7.94
Sharpness: 349
Resolution: 256x256
DDIM Steps: 50
Time: 4.2 seconds
High Structural Fidelity
GPU: Mid-Range
Strong Perceptual Realism

Qualitative Results

Generated Examples:
- "Normal chest X-ray, clear lungs."
- "Right lower lobe pneumonia with consolidation."
- "Mild cardiomegaly with pulmonary edema."
- "Left pleural effusion with atelectasis."

Observations:
- Model successfully captures key anatomical structures
- Pathological features match textual descriptions
- Realistic tissue density and contrast characteristics
- Good differentiation between pathological conditions



Example Generations

Prompt 1

Normal chest X-ray with clear lungs and no abnormalities.

Prompt 2

Right lower lobe pneumonia with focal consolidation.

Example Generations

Prompt 3

Bilateral pleural effusions, greater on the right.

Prompt 4

Cardiomegaly with pulmonary vascular congestion.

Interactive Application

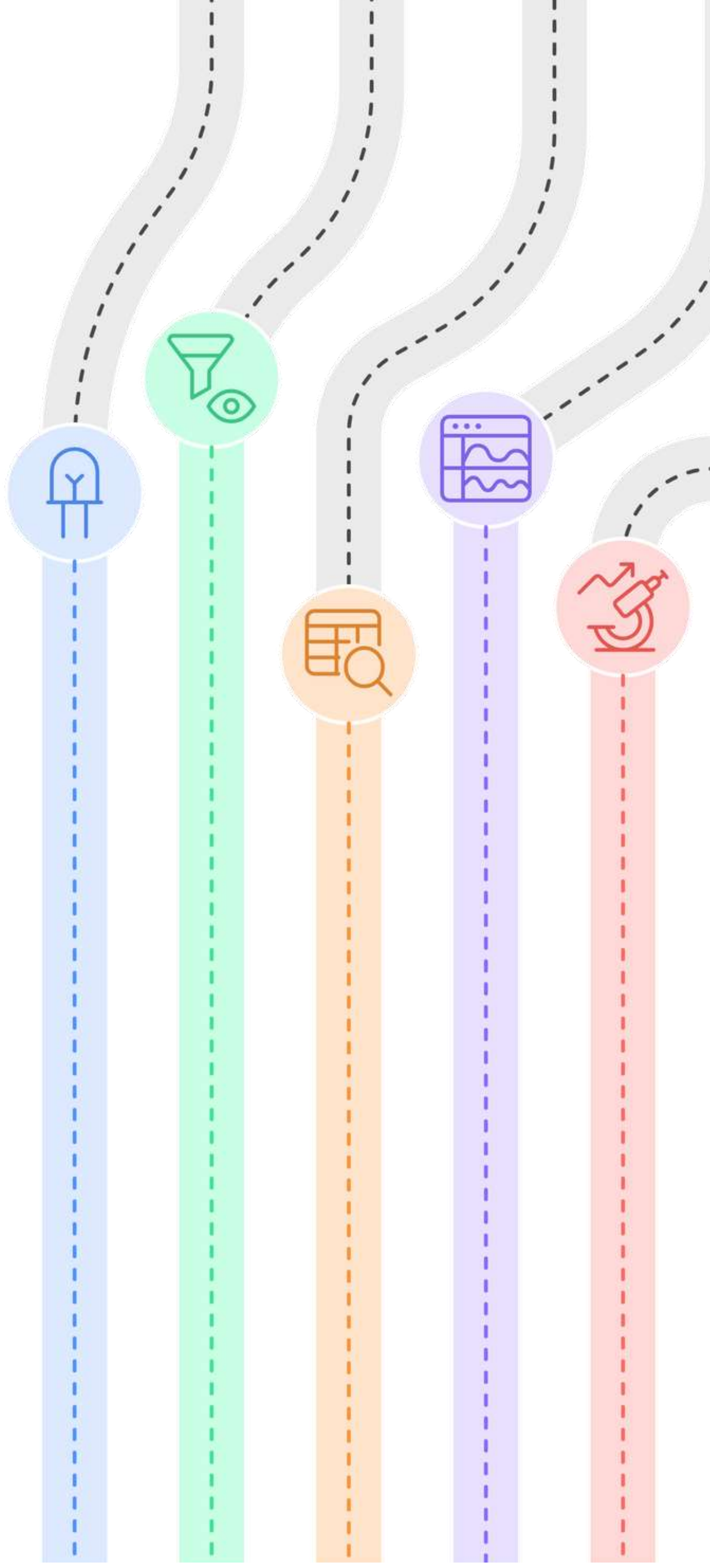
X-Ray Generator:
- Text-prompt based generation interface
- Adjustable quality and guidance parameters
- Real-time metrics calculation

Enhancement Pipeline:
- Multiple presets optimized for different viewing needs
- CLAHE, windowing, edge enhancement, vignetting
- Side-by-side comparison tools

Dataset Explorer:
- Browse real X-rays from the training dataset
- View paired radiological reports
- Compare real vs. generated images

Model Information:
- Architecture visualization
- Performance metrics
- Training process details

Research Dashboard:
- Advanced analytics
- Multiple condition comparison
- Custom enhancement experimentation



What I Learned

Technical Insights:

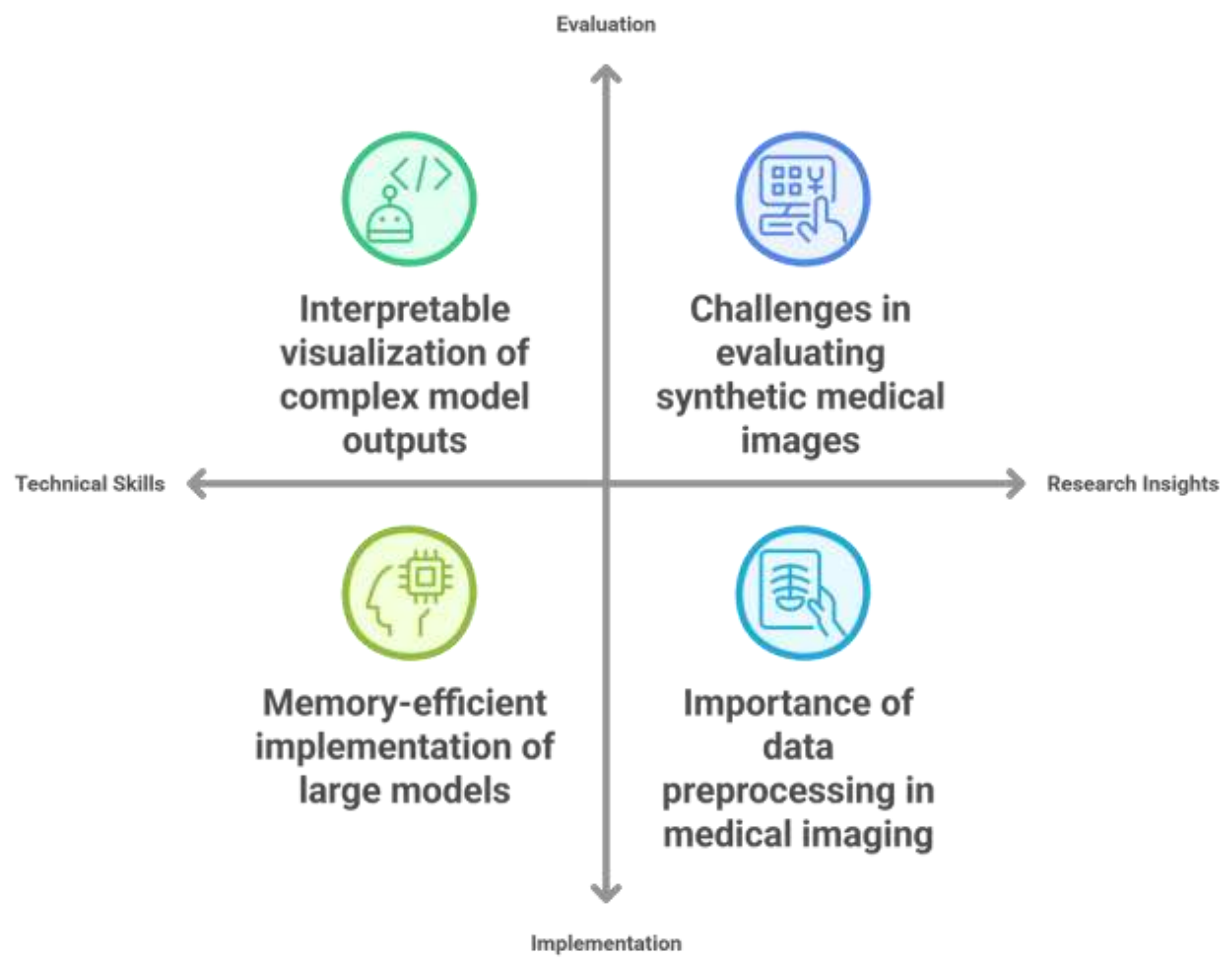
- Effectiveness and challenges of training latent diffusion models.
- Importance of domain-specific fine-tuning (BioBERT).

Practical Insights:

- Significant influence of data preprocessing on generative quality.
- Trade-off between image resolution and computational resources.

Personal Growth:

- Improved deep learning pipeline development and debugging skills.
- Enhanced problem-solving abilities with limited resources and datasets.



Discussion (Strengths & Limitations)

Strengths

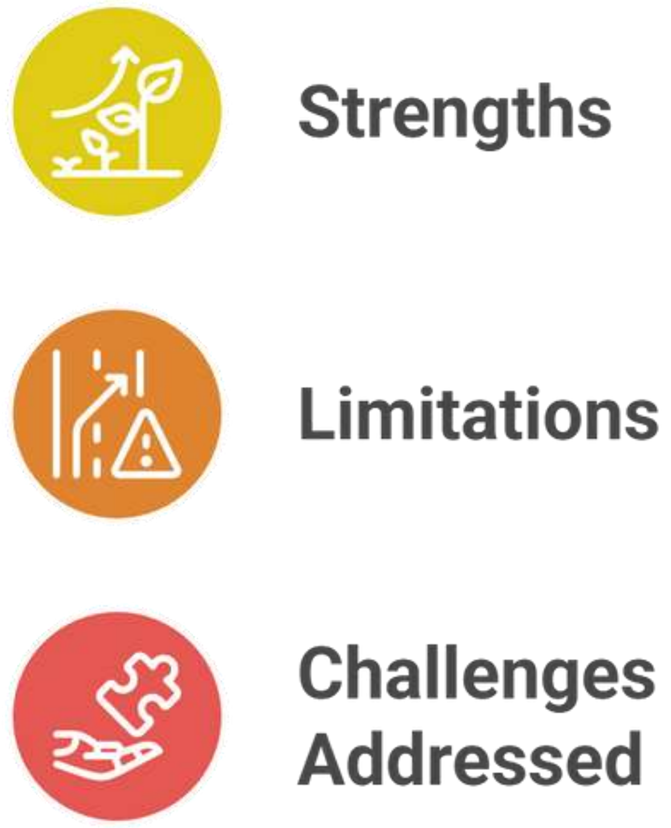
- Successfully generates anatomically plausible chest X-rays
- High image quality metrics (SSIM: 0.82, PSNR: 22.3 dB)
- Effective text conditioning for various pathological features
- Complete pipeline from generation to enhancement

Limitations

- Resolution constraints due to computational resources
- Text-dependency affects generation consistency
- Occasional anatomical inconsistencies in complex scenarios
- Output quality variations between pathological conditions

Challenges Addressed

- Memory optimization for model deployment
- Balancing latent compression vs. detail preservation
- Implementing enhancement pipeline for diagnostic quality
- Creating an intuitive interface for non-technical users



Conclusion & Future Directions

Achievements

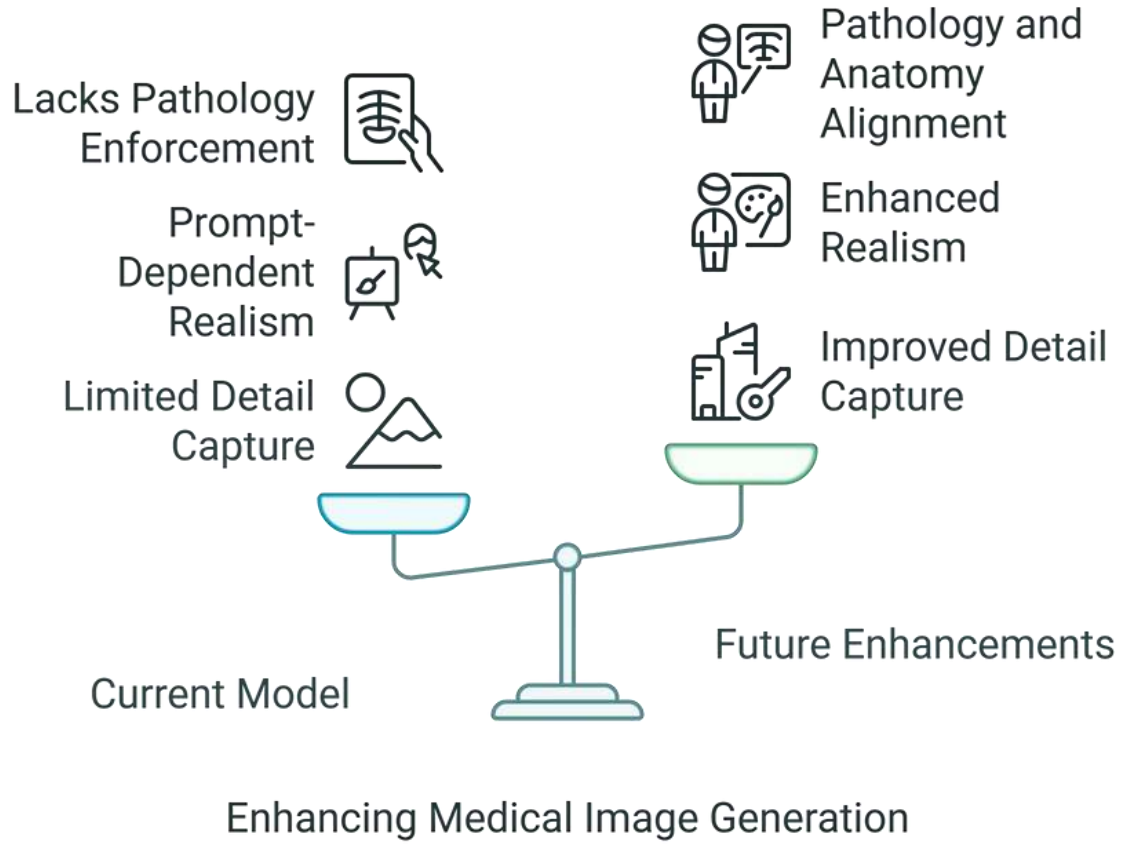
- Developed an end-to-end text-to-X-ray generation system
- Created a complete interactive application with multiple modes
- Achieved high-quality synthetic images with strong metrics
- Implemented clinically-inspired enhancement techniques

Future Work

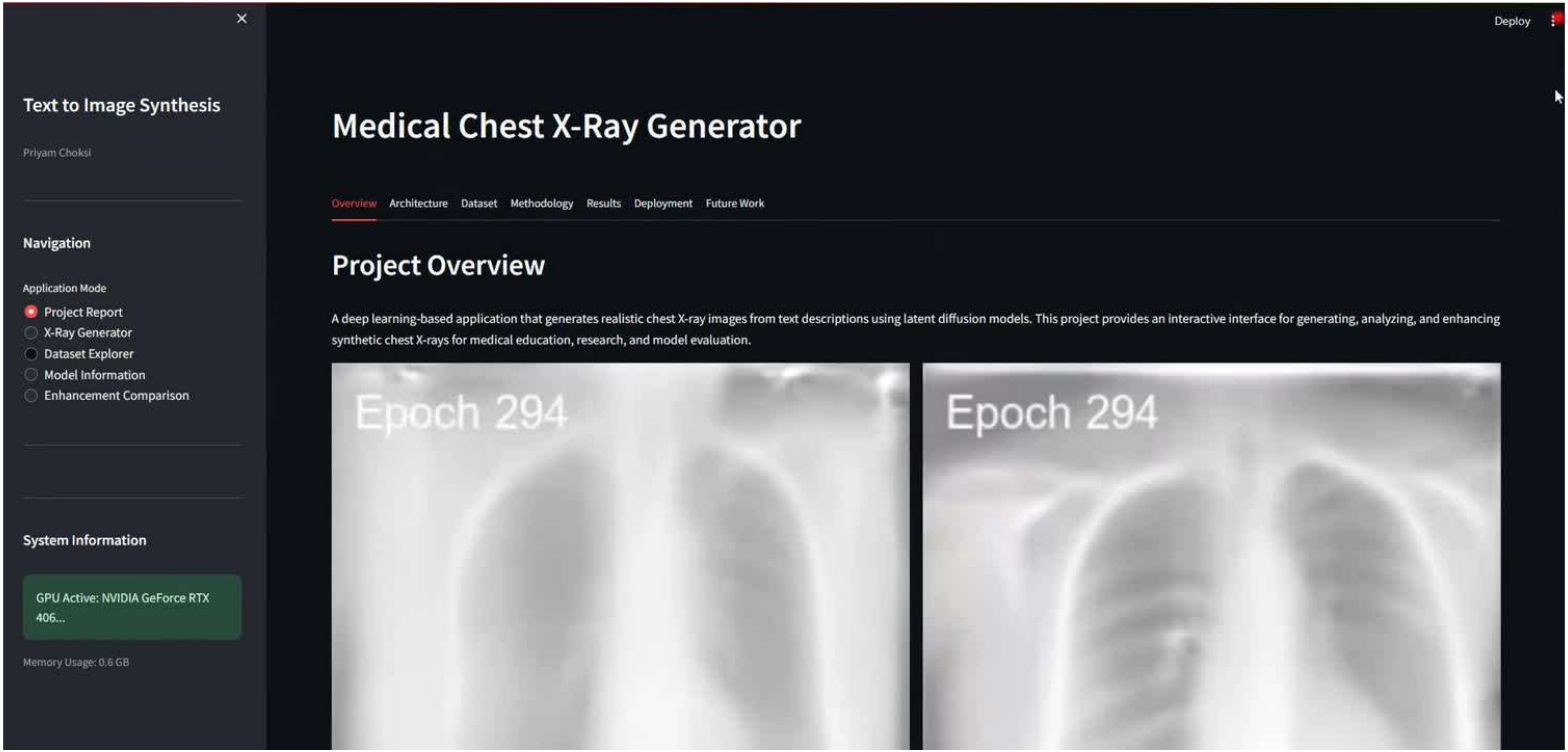
- Clinical Validation: Expert radiologist assessment study
- Higher Resolution: Scaling to 1024×1024 for clinical detail
- Multi-modal Conditioning: Combining text with clinical data
- Performance Optimization: Faster generation through distillation
- Enhanced User Interface: Task-specific clinical workflows

Potential Applications

- Medical education and training
- AI diagnostic model development
- Research into rare pathological conditions
- Teleradiology training and simulation



App Demo


Deploy
Text to Image Synthesis
Priyam Choksi
Navigation
Application Mode
Project Report
X-Ray Generator
Dataset Explorer
Model Information
Enhancement Comparison
System Information
GPU Active: NVIDIA GeForce RTX 406...
Memory Usage: 0.6 GB
Medical Chest X-Ray Generator
Overview
Architecture
Dataset
Methodology
Results
Deployment
Future Work
Project Overview
A deep learning-based application that generates realistic chest X-ray images from text descriptions using latent diffusion models. This project provides an interactive interface for generating, analyzing, and enhancing synthetic chest X-rays for medical education, research, and model evaluation.
Epoch 294
Epoch 294

Project Resources

1. Github (Project Files) : priyam-choksi/cxr_diffusion (https://github.com/priyam-choksi/cxr_diffusion)
2. Google Drive (Dataset & Model checkpoints) : Link (https://drive.google.com/drive/folders/1fNZavpgZ46zEHnimYAHWQ6-mwJhuXYTy?usp=drive_link)
3. Youtube (Streamlit App Demo): Link (https://www.youtube.com/watch?v=mzvOV1ZnXeE&ab_channel=Priyam)

Additional Training & VAE Gif and Model Outputs