Research Proposal: Latent Reasoning AI Agent for Chest X-Ray Diagnosis Assistance

Megha Kataki

Nov. 5 2025

Abstract

This proposal outlines the development of an AI agent that leverages latent reasoning to assist clinicians in interpreting chest X-rays and related medical information. The proposed system combines visual reasoning in latent space with multimodal understanding of clinical text to identify abnormalities, generate localized attention maps, and provide concise, interpretable rationales for its findings. Through this project I would first like to design and implement a latent reasoning agent that is able to understand image and text data. Second, use benchmark datasets to evaluate accuracy of model. Third, build a simple user interface to carry out reasoning that relies on image processing as well as patient information/symptoms. Anticipated outcomes include a reproducible prototype, evaluation results showing improved interpretability and calibrated uncertainty while also giving a confidence interval for given diagnosis. Along with a final report highlighting safety and ethical considerations in AI-assisted medical diagnostics.

1 Motivation

1.1 Problem Statement

Wait times in emergency rooms and urgent care are notoriously long. Creating an AI agent that helps speed up the diagnosis process may elevate some of this wait time. The goal of this project is to develop an AI agent that supports radiologists in chest X-ray diagnosis by combining visual understanding and medical reasoning. The agent takes an image and optional clinical notes as input, and outputs. In the diagnosis we hope to provide potential diagnosis, confidence intervals, and a summary for reasoning. This agent operates as ab aid, ensuring human oversight remains central in diagnostic decisions.

1.2 Background and Context

Recent advances in multimodal AI and latent visual modeling have enabled systems that reason over compact image embeddings rather than raw pixels. Datasets such as MIMIC-CXR and CheXpert provide labeled chest X-rays and reports, enabling. Models like Vision Transformers (ViT) and multimodal encoders (e.g., CLIP) can learn strong cross-modal representations. However, diagnostic tasks require interpretability and robustness beyond what general-purpose vision-language models currently achieve. By introducing latent reasoning we can address the robustness diagnosis need.

1.3 Limitations of Existing Approaches

Most current medical image AI systems act as static classifiers without interactive reasoning or uncertainty handling. They often lack:

- Transparent reasoning for each prediction (limited interpretability);
- Need for deep understanding of complex findings;
- Integration of non-image data such as medical notes or prior exams;
- Mechanisms to defer uncertain cases to human experts.
- Unseen diagnosis or unfamiliarity.

Our approach aims to fill these gaps with a reasoning-based, explainable agent that aligns with human diagnostic workflows, and addresses uncertainty.

1.4 Objectives

The objectives of this proposal are:

1. Develop a latent reasoning module that produces interpretable visual–textual representations of X-rays.
2. Train and evaluate the agent on public, de-identified chest X-ray datasets (MIMIC-CXR, CheXpert).
3. Incorporate uncertainty calibration and region-level attention to enhance interpretability and safety.
4. Build a user interface.

2 Approach

2.1 Overview

Our proposed frameworks:

1. **Perception Encoder:** A pretrained vision to extract latent reasoning from chest X-rays.
2. **Latent Reasoning Module:** Reasons over visual embeddings to identify potential findings and providing latent explanations.
3. **Text Fusion Module:** Integrates optional clinical text or radiology reports to aid predictions and reasoning.
4. **Decision and Explanation Layer:** Produces probabilistic output and short textual rationales rather than one clear diagnosis.
5. **Agent Controller:** Uncertainty thresholds and ambiguous cases will be sent to clinicians.

2.2 Design Principle

The design prioritizes:

- **Transparency:** The agent provides interpretable outputs rather than x,y predictions.
- **Safety:** It operates under a human-in-the-loop framework and flags uncertain results.
- **Efficiency:** Using latent reasoning reduces computational overhead compared to full generative methods.
- **Modularity:** Each component (perception, reasoning, explanation) can be improved or replaced independently.

3 Experiments

3.1 Experimental Setup

We will evaluate the model on two major datasets:

- **MIMIC-CXR:** 370k chest X-rays with associated radiology reports.
- **CheXpert:** 220k chest X-rays with 14 labeled pathologies and uncertainty labels.

Baselines:

- Conventional CNN-based classifiers.
- CLIP-style multimodal models without explicit latent reasoning.

Metrics:

- **Accuracy and AUC:** Evaluate detection of key conditions.

- **Calibration:** Expected Calibration Error and Brier score.
- **Localization:** Intersection-over-Union for heatmap-region accuracy.

Ablation Studies:

- Compare models with and without latent reasoning.
- Test different uncertainty thresholds for clinician referral.

Timeline

- **Week 1–2:** Conduct literature review and finalize problem formulation. Collect and preprocess the medical imaging and text datasets.
- **Week 3–4:** Implement baseline CNN and CLIP-style multimodal encoder for image–text alignment.
- **Week 5–6:** Develop latent reasoning module and integrate it with the multimodal diagnostic agent framework.
- **Week 7–8:** Train the full model, perform hyperparameter tuning, and run ablation studies.
- **Week 9–10:** Evaluate performance, conduct interpretability and robustness tests, and build a lightweight web-based demo interface.
- **Week 11:** Finalize written report and deliver presentation of project findings and demo results.

Expected Contributions

- A prototype AI agent capable of interpretable latent reasoning on chest X-rays.
- Evaluation showing improvements in interpretability and uncertainty calibration.
- An open-source research framework for responsible medical AI.
- Presentation reporting all findings.
- A report discussing ethical, safety, and regulatory considerations for AI in healthcare.