# Prototyping a Multimodal XAI Toolbox to Enhance Transparency of Black-box Systems

Hung Nguyen[1], Hung Cao[1]

[1]Analytics Everywhere Lab, Faculty of Computer Science, University of New Brunswick

## Introduction

Enhancing AI decision transparency is crucial in sectors like healthcare and banking, where explainability is highly sought after. Current Explainable AI (XAI) methods require significant expertise, limiting their accessibility to non-experts.

By integrating Large Vision Models (LVMs) — which interpret visual data in human-like ways — with XAI, we prototype a multimodal XAI toolbox to provide text-based explanations of AI decisions, making complex AI systems more transparent and understandable to all users without the need for deep technical knowledge.

## Related Work

**Explainable AI (XAI)** Within the context of working mechanisms, all XAI methods fall into two classes: Gradient-based and Perturbation-based.

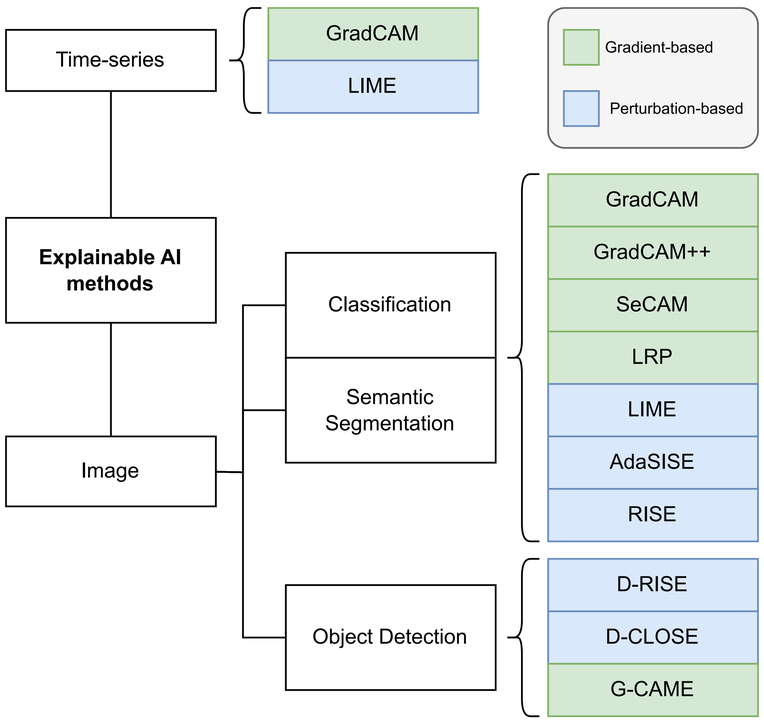


Figure 1. Classification of XAI methods by their mechanisms.

**Large Vision Models (LVMs)** The rise of LVMs stems from advancements in LLMs. These LVMs combine language understanding, reasoning, and visual perception. In our research, we explored a new addition to the LVM family: GPT-4 Vision.

## Proposed Prototype

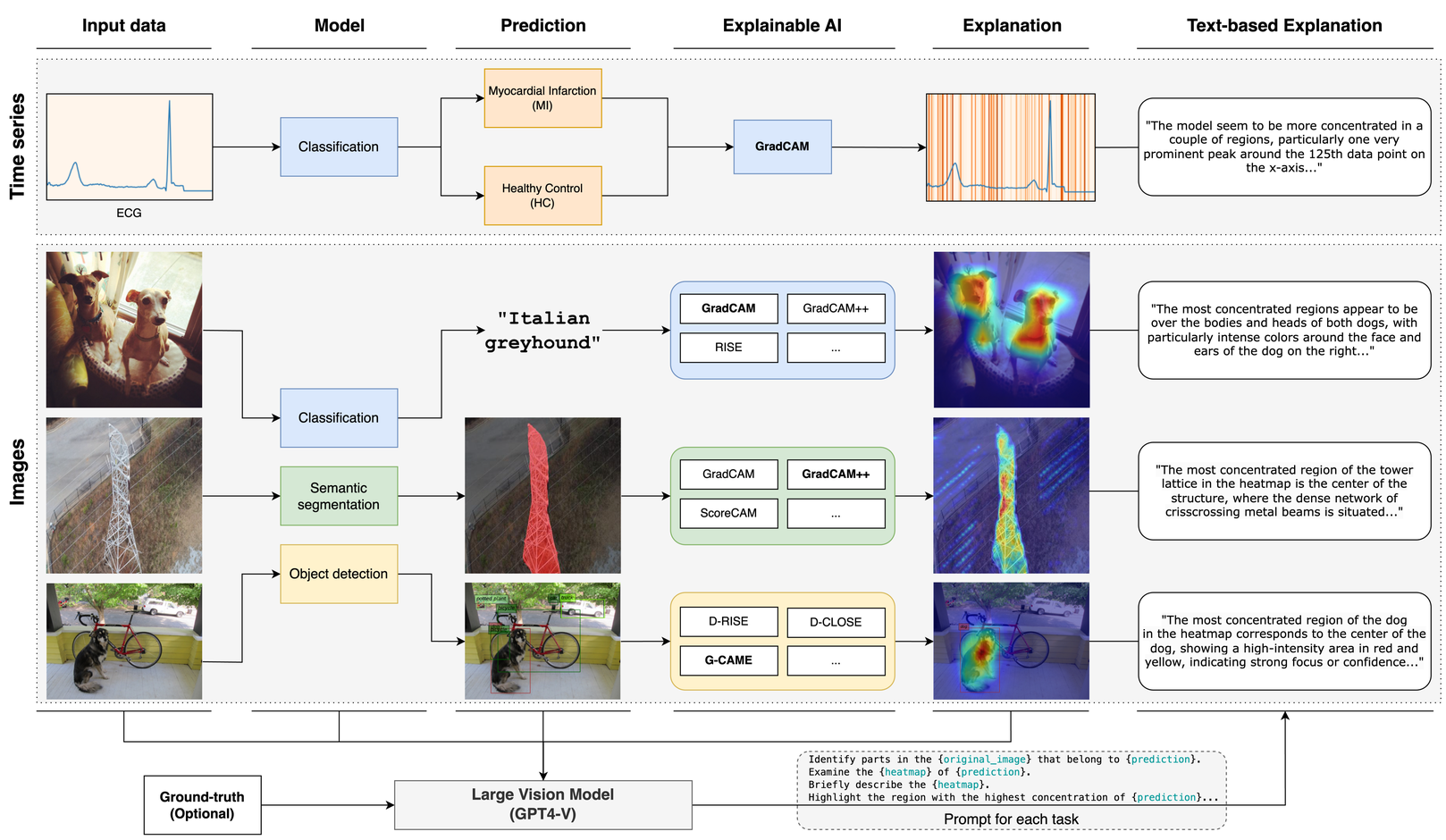


Figure 2. The architecture of our multimodal XAI toolbox for time-series and images input.

The architecture of our multimodal XAI toolbox is divided into two main blocks, aiming to provide text-based explanations for two types of data:

- **Time-series:** We demonstrate the ECG data with a classification model to detect whether a patient has myocardial Infarction (MI) or Healthy Control (HC).
- **Images:** We demonstrate three visual perception tasks: Classification, Semantic Segmentation, and Object Detection.

**Block 1: Explanation Map Extraction with XAI** The process involves uploading the data and selecting the desired task, with specific models assigned accordingly. Following the image analysis, users can specify the predicted class and choose the XAI method to generate saliency maps, which highlights the areas of interest for the model's decision-making process.

**Block 2: Text-based Explanation with LVM** We integrate different information to aid the LVM in generating text-based explanations, such as the input data, ground truth, model's top-1 prediction, and explanation map.

1. We employ a structured prompt for each task, starting with presenting the input data and explanation map to help the LVM identify areas of interest.
2. We combine the saliency map with the model's prediction to verify the accuracy.
3. We compare the model's prediction with the ground truth to determine the reliability.

This comprehensive process ensures explanations are coherent and faithfully based on the model's analysis of the input data.

Acknowledgement: This work is funded by NBIF Fund No TRF2003-001

hung.ntt@unb.ca