Infosys Responsible AI Office

# Infosys Responsible AI Toolkit

# Image Explainability

# API usage Instructions

## Contents



# Introduction

**Image Explain** offers detailed explanations for images generated by Large Language Models (LLMs). It provides an in-depth analysis of the image, highlighting key insights such as the presence of **watermarks**, detection of **potential biases**, and identification of the **visual style** used. In addition, the system evaluates the image using key metrics like **creativity** and **certainty**.

When applicable, Image Explain also performs **object detection**, identifying and labeling objects within the image. It then provides an explanation based on the detected objects, This enhances the interpretability and trustworthiness of AI-generated visuals by combining both image-level and object-level understanding.

Once API swagger page is populated as per instructions given in the github repository Readme file, click on ‘try it out’ to use required endpoints. Details of endpoints associated with ImageExplainability tenet are outlined below.

# Image Analysis

**Endpoint –** /rai/v1/image-explainability/analyze
Using this API, we can view the explanation of the uploaded image along with additional evaluations. The system also performs the following checks:

- **Watermark Detection**: Determines whether the image was AI-generated by detecting embedded watermarks.
- **Bias Assessment**: Identifies any unfair or skewed representations present in the image.
- **Style Classification**: Identifies the artistic or visual style (e.g., Digital Art) and provides a rationale through style analysis
- **Creativity Score**: Measures the uniqueness and originality of the image generated.
- **Certainty Score**: is a measure of how confident the system is in its analysis of the image

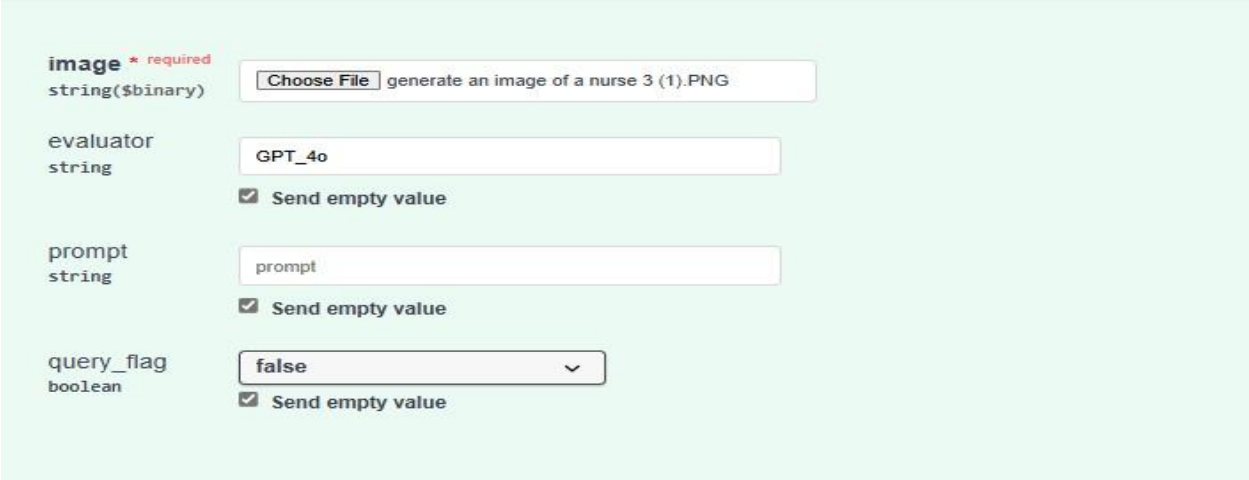
**Input:**

**image**: Upload the image you want to analyze. **evaluator**: Use GPT-4o as the

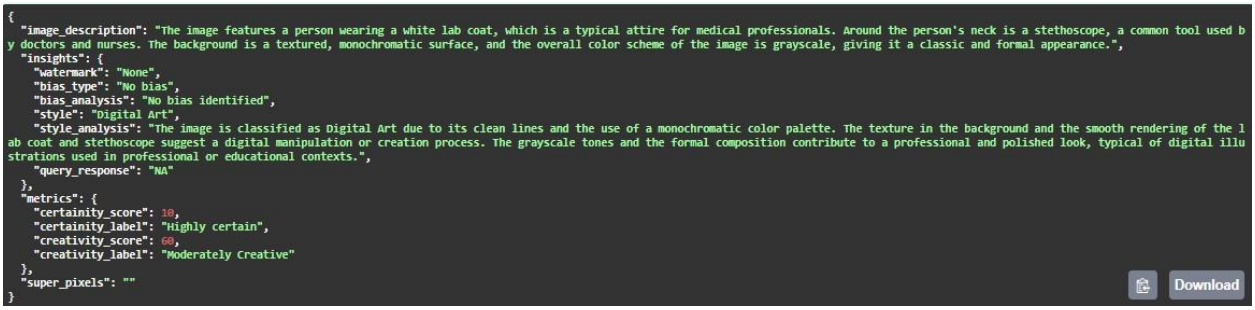
evaluator. **prompt**: Provide the prompt you want to use for analysis or querying the

image.

**query_flag**: By default, query_flag is set to **false**. If you're asking any query related to the image, set query_flag to **true** and provide the query in the prompt field.




**Response**:



# Explanation of Image with Detected Objects

**Endpoint** - /rai/v1/image-explainability/object-detection

Using this API, we can get detailed explanations of images based on detected objects.

**Request Payload:**


**image**: Upload an adversarial or non-adversarial image with bounding boxes or object detection annotations. **evaluator**: Use GPT-4o as the evaluator.

**Response:**