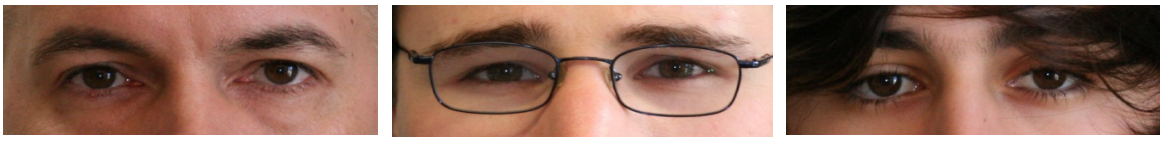
hugomcp@di.ubi.pt, 2024/25

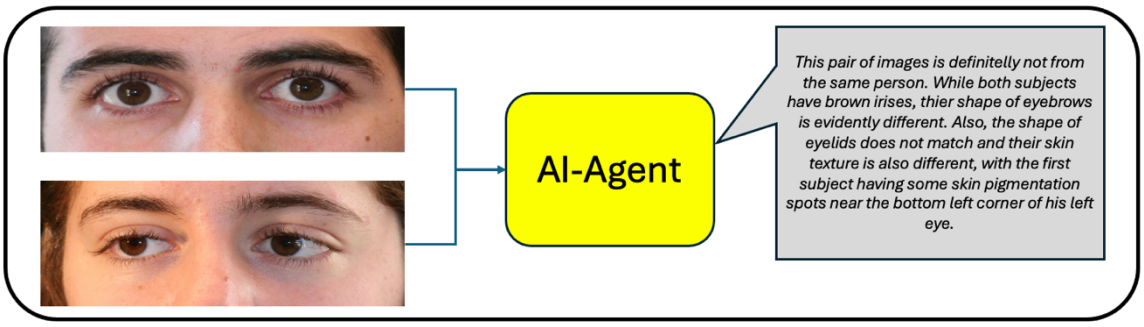
Practical Project - Interpretability/Explainability and LLVMs

Consider the "**UBIPr**" dataset, which is a version of the **UBIRIS.v2** set, with images cropped in a way that they cover wider parts of the ocular region than the original **UBIRIS.v2** data. It is particularly suited for experiments related with **Periocular Recognition**. Some examples are shown below:



The dataset can be obtained from: http://iris.di.ubi.pt/ubipr.html

The main goal in this project is to develop an automated agent that uses/interacts with a Large Language Vision Model, in an "interpretable/explainable pairwise biometric verification" problem. As illustrated in the schema below, the problem involves to receive a pair of periocular images and: 1. Determine if both images are from the same subject (1="yes"; 0="no"); and 2. Explain the decision, in a text/visual format.



In order to accomplish the final goal, there are three main phases:

1. Write a summary report and a presentation about the most relevant techniques for machine learning interpretability/explainability
    a. Due date: 07/03/2025, groups of 2 students

2. Design and implement a pairwise biometric verification CNN, using Python + Keras, that receives a pair of images and returns "1" if both images regard the same subject, or "0" in the opposite case;
    a. Due date: 28/03/2025, individual

3. Design an interaction plan between the agent developed before and a LLVM of your choice, to come out with the final interpretable/explainable system
    a. Due date: 30/05/2025, individual

PS: Note that all models used in this work should be "open source" or "free" versions of commercial LLVMs.

Biometric Verification System-Specific Features

Feature Representation Analysis

Evaluate the feature embeddings of the biometric data to understand:

- How similar or distinct embeddings are for **positive** and **negative pairs**.
- Dimensionality reduction (e.g., t-SNE or PCA) to visualize the separation of positive vs. negative pairs.

Threshold Sensitivity

Test different decision thresholds for similarity scores and analyze:

- The **False Acceptance Rate (FAR)** vs. **False Rejection Rate (FRR)**.
- Trade-offs captured in the **Receiver Operating Characteristic (ROC)** curve.

Interaction Between LLMs and Vision Models

Textual Explanation of Visual Similarity

Use a VL model (e.g., CLIP or Flamingo) to generate natural language explanations for the similarity between two biometric samples, such as:

"These two faces are similar due to matching eye shapes and iris color."

Prompt Tuning for Pairwise Analysis

Evaluate the effect of **prompt engineering** in LLMs for tasks such as:
Generating descriptive explanations of why two biometric features are similar or different.
Rephrasing biometric outputs into user-friendly narratives (e.g., "Match confidence is 80%, which falls within the threshold").

Embedding Consistency Between Models

Compare embeddings from the biometric verification model and vision-language models:

Are their representations consistent in terms of **similarity rankings**?
Use similarity scores (e.g., cosine similarity) to cross-validate outputs.

Error Case Analysis with LLMs

Use an LLM to analyze and explain edge cases where the verification system failed:
Generate hypotheses about **why certain pairs were mismatched** or incorrectly matched.