**Université Nice Côte d'Azur**
Enrico FORMENTI

**Année 2025/26**
**M1 Info**
**AIMA**

# Project
# Celeb face recognition

## Contents



## 1 Project statement

Your are supposed to develop a Python project which can recognize people in a picture on the basis of their faces. The project is structured into three parts, each one with its own purpose and its own dataset. These parts are described in the following sections.

### 1.1 Extracting bounding boxes

The purpose of this first part is to extract from a generic picture bounding boxes containing the image of a human and then save the sub-images obtained into a working directory (please call it `working`) for further manipulations. The filenames of the sub-image should be called with a precise name as follows:

`original-filename-label-nn-bb-x1-y1-x2y2.jpg`

where `original-filename` is the filename of the current image from which we are extracting the bounding box, `label` is the category of the object that has been recognized (for example `person`) and `x1-y1` (resp., `x2-y2`) are the coordinates of the top left corner (resp., bottom right) of the bounding box expressed in pixels.

For performing this task, we are going to use a (quite famous and quite good) pre-trained network `yolo11n` by `Ultralytics`. The network is specialized in object detection in images and can work also in real-time (video). It can recognize almost 90K of different objects and for each of them provide interesting metadata such as bounding boxes, varius kind of metrics, *etc.* All the details and necessary information about the `Ultralytics` library can be found here.

### 1.2 Extracting faces

At the beginning of this stage, you should have a list of images containing persons in the `working` directory. This list should be provided to the new stage that we are going to describe in this section. The idea is to use the Python package `retinaface` to extract pixel perfect bounding boxes of the faces in the pictures produced at the previous stage. The package can be installed using the `pip` and a tutorial and more informations can be found on Github here.

### 1.3 The final stage

At this stage we have a sequence of images containing faces (and only that). The final goal is to use a pre-trained network to recognize the celebrities in your image list (we want only to provide the name and surname and not their physical characteristics). The idea is to use a pretained model from the Visual Geometry Group of Oxford University (VGG) which can be found here. More details on the library can be found here.

### 1.4 Evaluating your model

You are expected to provide arguments to evaluate each stage of the project using the methods seen in our classes.