**CS512 - Project proposal**
**- Spring 2023 -**
# Real-time Hand Gesture Recognition

**Names of team members:**

-   Bastien Leduc - A20520860
-   Paul Legout - A20522029

## Description of the problem addressed:

The purpose of this project is to implement an AI model for live hand gestures detection, using MediaPipe.

Hand gestures detection is one of the most common applications of computer vision. These types of live models can be used for several purposes:
-   video conferencing services,
-   home automation systems,
-   the automotive sector,
-   services for people with speech and hearing impairments, etc.
-   Human-computer interaction

There are going to be two major parts in this project.
The first part consists of implementing a real-time hand detection system using MediaPipe from which we will be able to collect and extract information about the different hand poses, such as hand landmarks and bounding box coordinates.
The second part consists of using the extracted image information to train a hand gesture recognition classifier, using a CNN.
Once the model is trained, the last part is to incorporate the model in a global application: live translation from the video stream of our webcam

## Methods used: (description of the papers)

**Link of the first study paper :**
**On-device Real-time Hand Gesture Recognition**
https://arxiv.org/pdf/2111.00038.pdf

This paper uses MediaPipe to detect landmarks. Then, a classifier is trained (fully connected neural network), and it returns the associated hand gesture emoji.
From this paper, we will use the MediaPipe part, and implement our own CNN classifier.

**Link of the second study paper :**
**Real-time Dynamic Sign Recognition using MediaPipe**
**https://ieeexplore.ieee.org/document/9982822**.

This paper shows the use of CNN and LSTM models to recognize the hand gesture and motion proper to the America Sign Language. It shows how the model was trained using hand landmarks, as well as body posture (holistic) to recognize the words of the language itself. We will take inspiration from this paper to build our hand gesture recognition model.

# Data source:

**Link of the third study paper :**
**HaGRID — Hand Gesture Recognition Image Dataset**
**https://arxiv.org/pdf/2206.08219.pdf**

We are going to use the third paper as our main data source. This paper introduces an enormous dataset Ha-GRID (HAnd Gesture Recognition Image Dataset) for hand gesture recognition (HGR) systems. This dataset contains 552,992 samples divided into 18 classes of gestures. The annotations consist of bounding boxes of hands with gesture labels and markups of leading hands. As a matter of simplicity, we will only keep 100 images of each 18 class for training purposes, and we will either collect our own annotations, or use the suggested ones for this dataset.

# Responsibilities:

We are going to divide the workload the following way:

- Bastien: Data preparation and preprocessing. I will be in charge of the MediaPipe hand recognition interface implementation. For all the images of the dataset, I will extract the hand information (landmarks and bounding boxes coordinates) and assign each tuple to the correct sign label. We may perform data augmentation to increase the size of our dataset.
- Paul: Model construction and training. I will use the preprocessed dataset for training our CNN model using the hand landmarks and bounding box coordinates.
- Both: Once we get a sufficiently performant model, we will save it and merge the two parts, that is giving to the network the frames from the camera feed and display the recognized hand gesture label on the video.

**References:**
**Video implementation of hand-tracking module with mediapipe.**
**https://www.youtube.com/watch?v=NZde8Xt78Iw**