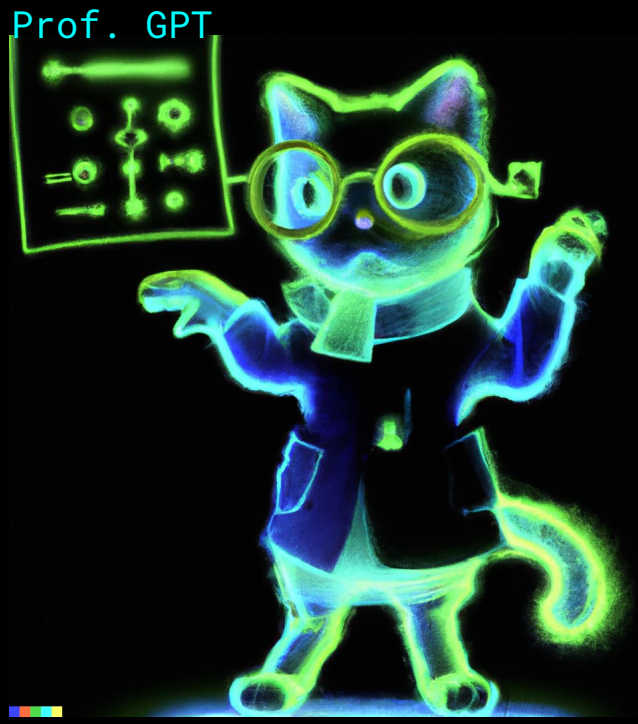
Computer Vision Tasks

Nadia Ahmed

Computer Vision Tasks

We would like to enable computers to interpret, analyze and understand visual data…

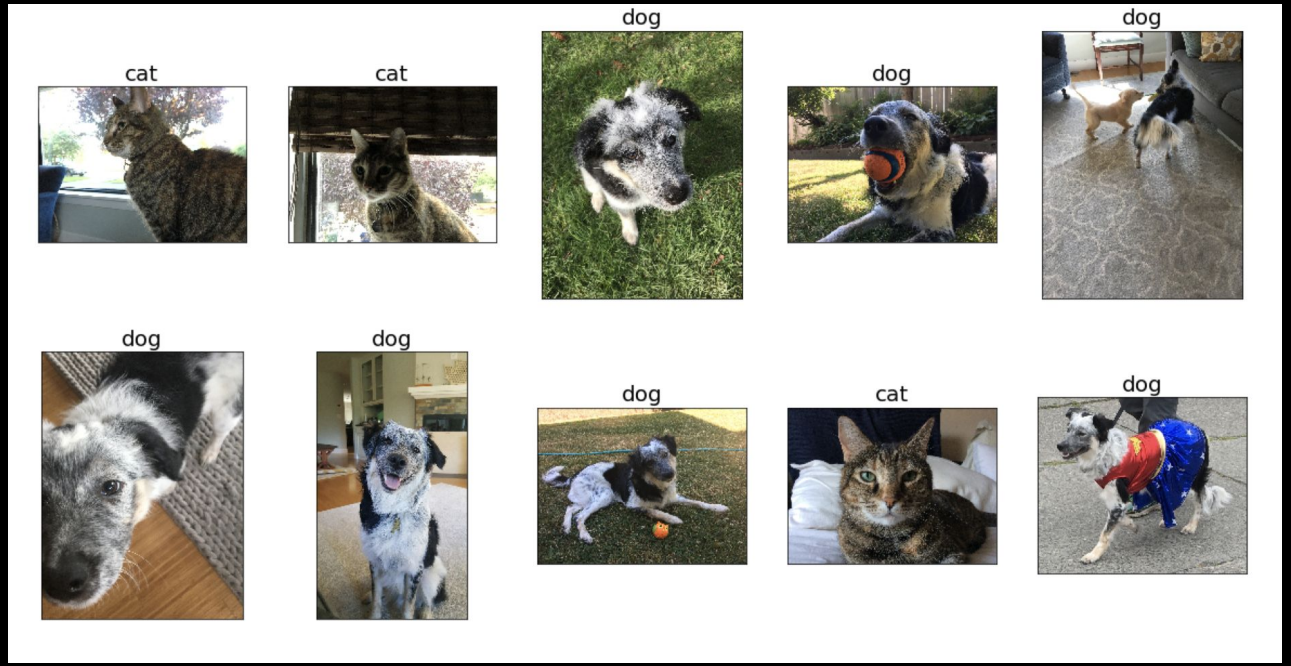
- Image classification
- Object detection
- Semantic segmentation
- Instance segmentation
- Pose Estimation
- Optical Character Recognition
- Face Recognition



*source: Le Chat GPT

Image Classification

Goal: Assign a single class label (image category) to the image.



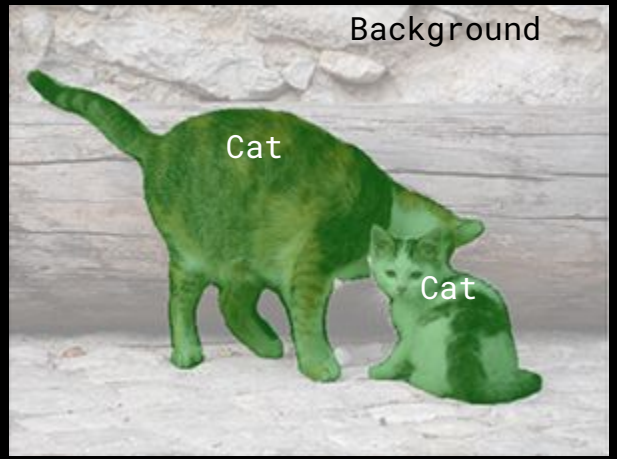
*source: apple.github.io

Semantic Segmentation

Goal: Assign a semantic category label to every pixel in an image.

- “stuff”: background
- “objects”: foreground

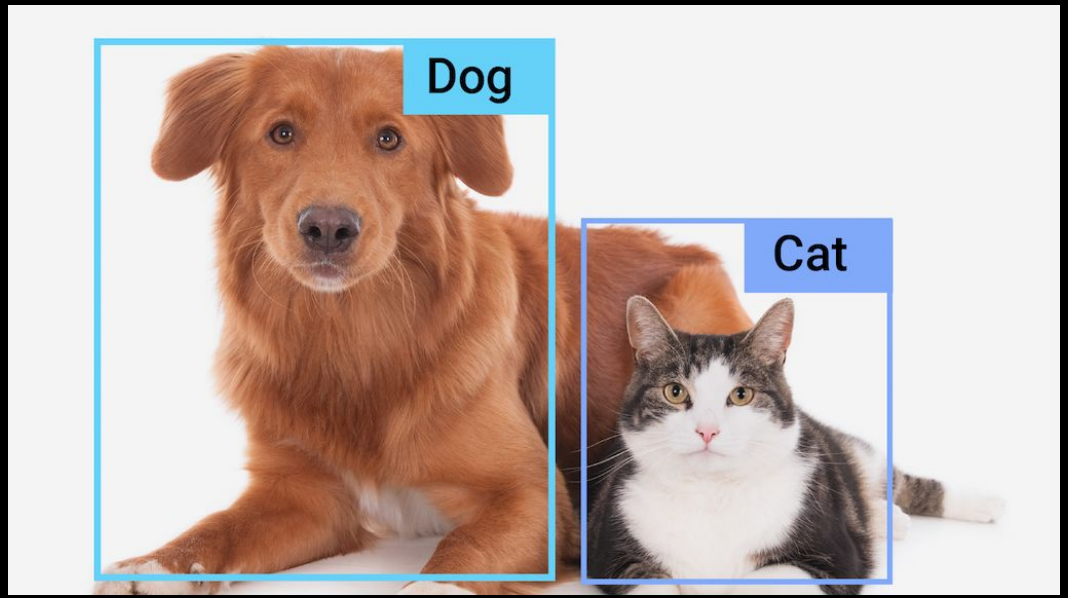
Every pixel gets a label.



*source: aws.amazon.com

Object Detection

Goal: Identify and localize objects within an image and drawing 2D bounding boxes around them and retrieve the correct category label.



Note: “stuff” or background does not get labeled. Only “object” or foreground items do.

*source: developers.google.com

Instance Segmentation

Goal: Identify and assign a semantic and an instance label to every pixel of an object and distinguishing it from other objects.

** We would like to find all of objects detected and assign a mask and a class label.

*source: celestai.com

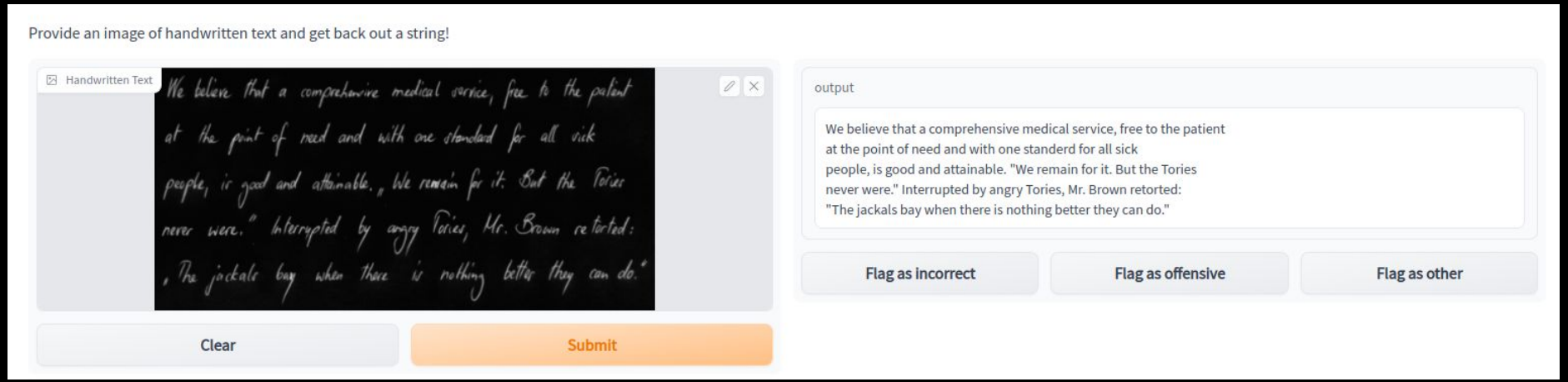
Pose Estimation

Goal: Determine the position and orientation of objects in 3D space based on 2D input.

*source: cao et al 2019

Optical Character Recognition

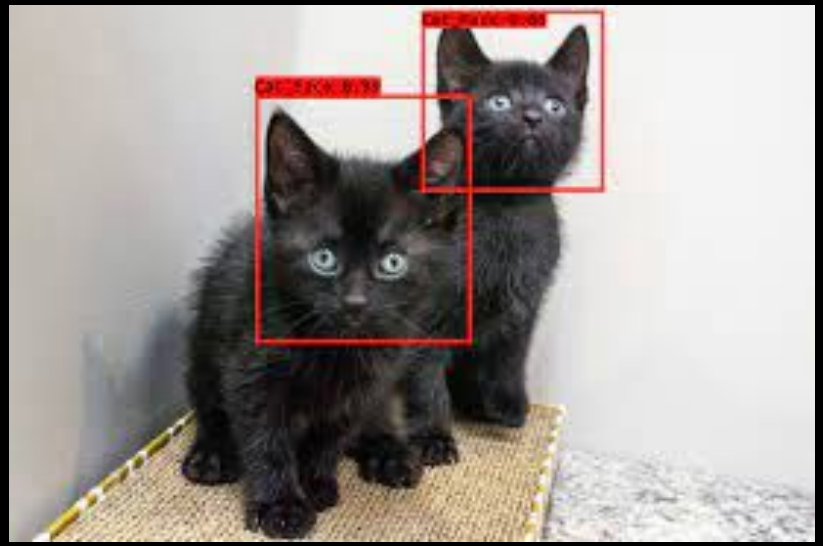
Goal: Recognize and interpret text characters within images or videos.



*source: fsdl

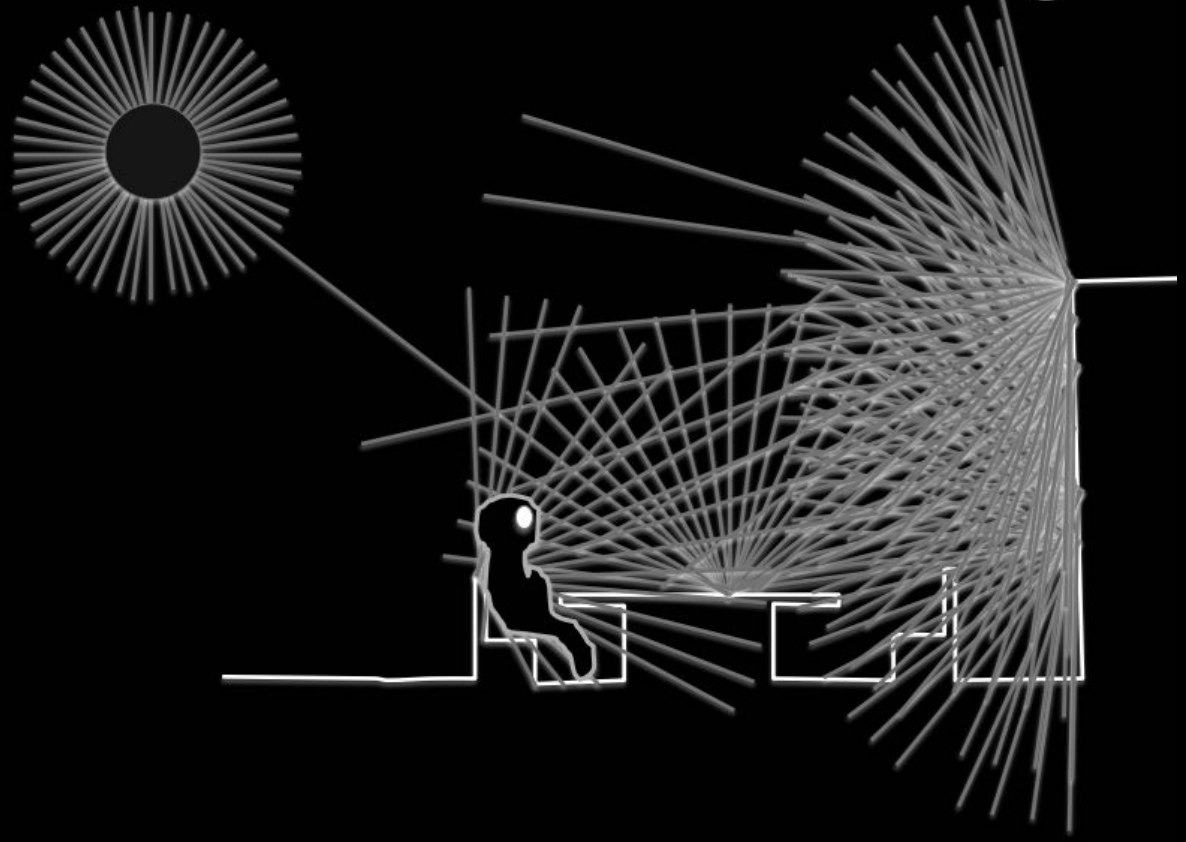
Facial Recognition

Goal: Identify and verify the identity of individuals based on facial features.



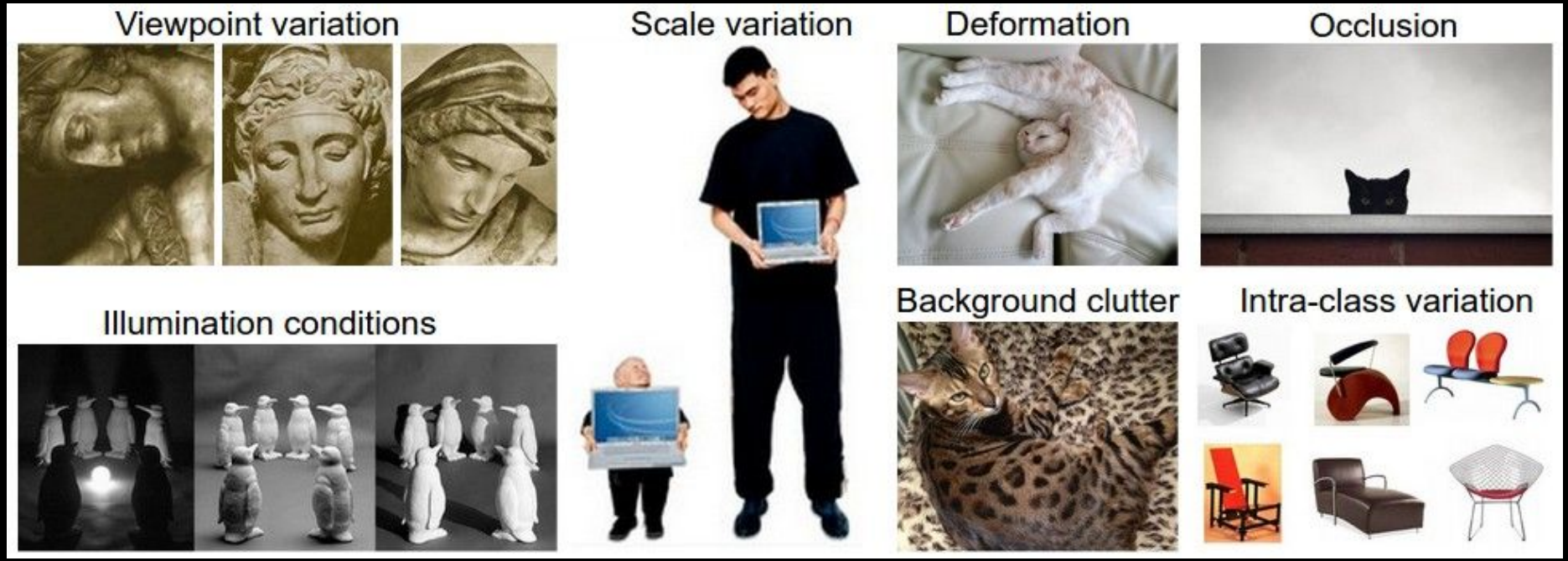
*source: A. Klein, 2019

The structure of ambient light



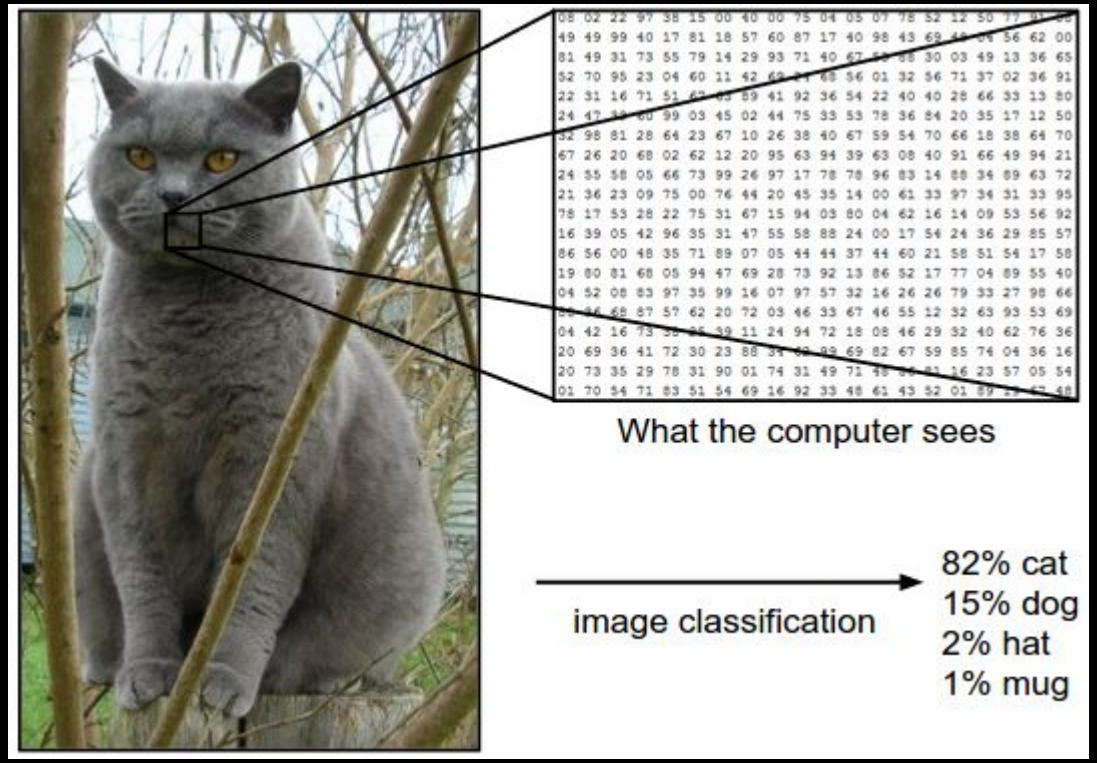
*source: A.Torralba SANE 2018

Challenges



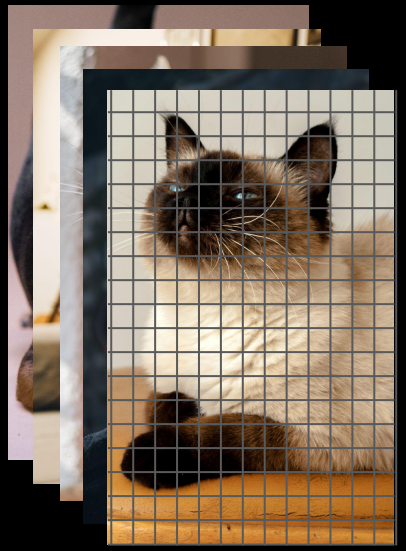
*source: Stanford 231n

What a Computer Sees is Not the Same as What We See…



*source: Stanford 231n

Parameterized Mapping From Image to Label



How do we assess the performance?

- **score** function (map raw data to class score)
- **loss** function agreement between prediction and ground truth

ML Magic or Math???

$R^D \rightarrow R^k$

"cat"

"dog"

k

Which computer vision task is this illustrating?

k: number of labels

y_i: label sample i

$y_i \in R^k$ (1,..k)

N: examples

D: dimensions

x_i: sample i

$x_i \in R^D$

We select the **model architecture** based on the output dimensions of what we would like to accomplish.

*source: Stanford 231n

We select the **model architecture** based on the output dimensions of what we would like to accomplish.

The output layer of a deep learning network is responsible for producing the final output of the model. The output layer typically consists of one or more neurons, where each neuron corresponds to a specific output class or category.

- In a **classification task**, the output layer may consist of **multiple neurons**, **one for each class**, with the output of each neuron representing the **probability** that the input belongs to that class.
- In a **regression task**, the output layer may consist of a **single neuron** that outputs a **continuous** value.
- In **object detection**, the architecture might include additional layers for generating **bounding boxes** and predicting **class labels** for each box.
- In **semantic segmentation**, the architecture may consist of an encoder-decoder structure with skip connections to preserve spatial information resulting in a **pixel-wise label map**.

*source: Le Chat GPT