# How does an image classifier work?

An image classifier is an AI system that, given an image, automatically assigns it one of the predefined labels (such as "dog", "car", "house", etc.). Our classifier is made up of hundreds of neurons, each of which detects a specific pattern in an image ("round things", "eyes", "white surface", ...), which helps the classifier to make a final decision about what is in the image.

# Your task

You are given the output of a system which, given a list of 10 most relevant neurons, writes a description in plain English of how the classifier has detected a given class. Your task is to evaluate the quality of the text produced by the system based on a list of neurons - which we call the "meaning representation".

**Meaning representation** contains the final decision of the model and the list of neurons. For each neuron, you are given:

- description - description of the pattern which is detected by the neuron
- positions - in which parts of the image the given pattern was detected.

For instance:
`{'description': 'Objects with led, text, and circular objects', 'positions': ['top', 'top-right corner', 'left', 'center', 'right']}`
means that the neuron detected 'Objects with led, text, and circular objects' on several parts of the image (top, top-right corner, ...).

The text produced by the system should be human-readable and summarize the list of neurons.

# Survey instructions

After reading the meaning representation and the output text, you will be asked:

- *Does the text contain information that was not present in the meaning representation?*
  In the text, we **can omit** some information or put it in different words, but we should **not add new** information which is not grounded in the meaning representation.
- *Is there any important information from meaning representation omitted in the text?*
  Note, that the text summarises data from the meaning representation, so omitting not-critical information **is expected**. **IMPORTANT NOTE**: neurons (instances in curly brackets) are sorted by their relevance to the classification in the descending order (the neurons that are higher/enlisted first in the meaning representation are more important). Omitting some spatial information, especially compressing it or making a valid, conglomerate spatial positions does not qualify as omitting important information.
- *Is the text linguistically correct?*
  Whether the text is fluent, without grammatical errors, etc.
- *If any spatial compression occurred between explained neurons, the said compression is correct?*
  In the output text, instead of enumerating ['top', 'top-right corner', 'left', 'centre', 'right'], the system can say "in the entire top and entire centre of the image".

- *Overall, do you find this meaning representation to text transformation acceptable i.e. sufficiently good for explanation purposes?*
  Your overall assessment of whether this is a good textualisation of the meaning representation. It is important to note, that you are asked in this question to evaluate the specific transformation from a given meaning representation to a given textualisation of it. Abstract this answer from the standalone quality of the explanation without pairing it with a given meaning representation.