**AI IMAGE RECOGNITION**

AI Image Recognition teaches computers to identify and understand images through exposing it to a large sample of examples, helping it to learn the differences between various objects.

**Steps**

1. **Training your data** – the AI is trained using a large set of images.
2. **Pattern/ Feature Recognition** – the AI looks for patterns and features in the data provided, like shapes, colours and textures. It learns what makes a cat a cat – its specific characteristics.
3. **Prediction** – Once the AI is trained, the AI can analyse new images. When you show it a picture, it checks for the learned features and makes predictions about the image contents to identify the object.
4. **Feedback** – Sometimes the AI gets it wrong, and it can be retrained with new examples or corrections to improve its accuracy.

normal text transformer

numerical vector (image identifier)

Input image

vision transformer

Vision transformer: *the vision transformer (ViT)* breaks down images into smaller **patches** (small square of the image) instead of looking at the whole image at once. The *ViT* uses ‘’attention’’ to focus on important parts of the images and learn how each patch relates to each other.

*i.e. if there’s a cat in the image, it can pay more attention to the patches that show the cat’s features*

After processing all the patches, the model combines all information to make a final decision about what’s in the image.

Normal Text transformer: designed to process and understand **natural language**. The transformer takes a piece of text as input and broken down into smaller parts (**tokens**) and it gets turned into a **numerical representation**. The transformer then produces a text output based on what it was trained to do (a translation/ summary/ answer to a question)