# Flower Image Classification using Convolutional Neural Networks (CNNs)

## 1. Research Question

Is a Convolutional Neural Network (CNN) capable of accurately classifying flower images into their correct categories [daisy, dandelion, rose, sunflower, and tulip]?

## 2. Dataset Description

- **Source:**
  The dataset appears to be custom-compiled and has been provided by the Module Leader (6CS012, AI and ML). It has been uploaded to a directory on Google Drive and is used for the subsequent coding tasks.
- **Classes Included and distribution per class**
  The dataset includes five flower categories, the table below shows the division of number of images in each classes.

| FLOWER CATEGORIES | NO. OF IMAGES |
|---|---|
| **Daisy** | 763 |
| **Dandelion** | 1051 |
| **Rose** | 783 |
| **Sunflower** | 732 |
| **Tulip** | 983 |

- **Total Images:**

  The dataset comprises **4312 images** across all classes.

- **Image Resolution:**

The images in the dataset have varying shapes. After performing a brief analysis, it was determined that all images need to be resized to a consistent size for further processing and model development. Based on observations, resizing all images to 256x256 pixels would be optimal before feeding them into the training model.

- **Data Split:**
    - Training Set: 3450 images
    - Validation Set: 862 images
    - Test Set: 5 images
- **Collection & Labeling Method:**

  The images were manually labeled by sorting them into folders corresponding to their flower class. Each folder is named according to the flower type. The labeling process was visually verified to ensure the integrity of the dataset. However, there are some irrelevant images present that could negatively impact the model's performance and should be addressed.

- **Challenges:**
    - Ensuring image quality and relevance (removal of noisy/blurry images).
    - Preprocessing and resizing all images uniformly.
    - Checking and discarding corrupt images.

## 3. Conclusion

The dataset is well-suited for training and testing CNN-based image classification models. Its diversity in visual features, combined with a balanced class distribution, forms a strong foundation for effective learning and building a high-performing model.