# Project Proposal: Multi-Class Classifier for Flower Images

## Problem statement

The goal of this project is to build a classifier that can correctly identify a picture of a flower. This could be a fun app to use. If you are outside walking and see a flower you could take a picture and find out which species it is.

For example, given this picture our model could tell us that this is a picture of a Balloon flower.

But why is this interesting? Well, there are a lot of different kinds of flowers. The dataset we have picked has three hundred flowers. Accurately guessing the right flower is very hard when there are this many classes.

## Data

For this project we will use the Flowers299 dataset from Kaggle.

Size:

- 299 classes of flowers
- ~200 - 400 images per flower, total: 116,000 images

## Approach

This is a supervised problem since we have a labeled data set. We will likely need to use convolutional neural networks (CNNs) to extract features here, since we have image data and it would be too time consuming to find the features in other ways. Other techniques that can be used in this area would be transfer learning, that is finding an existing model and tuning it, and also data augmentation to reduce bias in our model.

## Final Deliverable

The final deliverable will be a web app that takes a picture of a flower as input and then predicts the class it belongs to. The app may also return the confidence of the answer.

## Resources Required

Here are some rough estimates of what we will need to train our model.

- Memory: About 1GB of ram is enough. We can load the images in batches to reduce this cost.
- GPU/CPU: Maybe 50-100 GPU hours? I don't have any good reference points here. Running just one epoch on the full training set takes about an hour, and we probably need to run at least 50 epochs.