## 1. Dataset

For my deepfake detection project, I will be using the dataset provided by the Deepfake Detection Challenge at https://www.kaggle.com/c/deepfake-detection-challenge/data. This Dataset provides more than 470 GB of videos of people, either deepfaked or real. I will further separate them into training, validation, and test sets.

## 2. Methodology

i. Data Preprocessing
This dataset is great because it is the one provided to complete the official Deepfake Detection Challenge, from which my project takes inspiration. To preprocess the data, I will iteratively apply convolutions to denoise the images, and then pooling to enhance vital information and reduce image size.

ii. Machine learning model
I want to make a machine learning model for image classification and identify if an image is either a deepfake of authentic. I will use a Convolutional Neural Network to analyze the images, since it needs less pre-processing than other image classification algorithms.

iii. Final conceptualization
I plan on making a webapp, where the user can upload or select an image containing people on which my model will be applied. It will say if the image is a deepfake or not. My goal is to make the model return binary values, either deepfake or not, but ideally, I want it to return the probability of the image being a deepfake, for more in-depth results.