# Project Overview - The Faux Fighters

Github: <https://github.com/annabchox/deepfake-detector>

## Nutshell

Deepfake image detector

## Dataset

Data: [source](https://zenodo.org/record/5528418#.ZFl1m3bMKHu), [kaggle mirror](https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images)

## Problem Statement

The rapid evolution of generative artificial intelligence (GPAI, LLMs) social media has rapidly increased the public’s access to powerful, deceptive tools. One such concern is the increasing prevalence of deepfake images, which pose a significant threat to public trust and undermines the epistemic integrity of visual media. (Source).

These manipulated images can be utilized to spread false information, manipulate public opinion, and polarize communities, which can have serious consequences for both social and political discourse. In this project, we aim to develop a machine learning model that can detect differences between deepfakes and real images to combat the spread of manipulated visual media and protect the integrity of social discourse.

## Metrics

metrics we will use to evaluate our model?

* Classification metrics
  + (Balanced) Accuracy, F1 scores, classification report, confusion matrix
  + Recall likely more important!

## Schedule

### Monday 5/22

* final presentation

### Sunday 5/21

* aim to finish project
* Powerpoint finalized by Sun
* have emergency time

### Saturday 5/20

* Powerpoint draft done by Sat
* All models + powerpoint graphs done
* Then decide on whether to pursue stacked model

### Friday 5/19 (Flex Time)

* likely work time here from class
* Have a functional min code of models (possibly use gridsearch cv)

### Weekdays: Mon to Thurs 5/15-18

* ~~Thurs: EDA done~~
* Wed: Cleaning done/ EDA started
* Tues: Cleaning more
* Mon: Cleaning more

### Saturday (4p-)

* ~~Cleaning started~~
  + Reid: Saturday is good, Sunday not sure(mother’s day)
  + Chris: Saturday afternoon onwards
  + Anna: Saturday free after 3 pm
* have shared stuff active
  + ~~Github~~
  + ~~Google Drive~~
  + ~~Google Slides~~
* ~~have dataset downloaded~~
* ~~set guidelines on how to style code~~
  + ~~cleaning + precode a sample model together to develop a skeleton and then base our parallel models on that skeleton~~
* ~~draft out workflow/project timeline + people’s individual contributions~~
  + ~~potential~~ [~~example~~](https://towardsdatascience.com/end-to-end-pipeline-for-setting-up-multiclass-image-classification-for-data-scientists-2e051081d41c)~~,~~ [~~example~~](https://www.youtube.com/watch?v=dw96S_iFFbI)
  + ~~potential steps,~~ [~~sample~~](https://docs.google.com/presentation/d/1BMky-MH5rmhEaPY-F8qM8sa8ABOaDd0knDvAZhiyNTc/edit#slide=id.p)
    - ~~Data Cleaning~~
    - ~~EDA + Graphs~~
    - ~~Modeling~~
    - ~~Evaluation + Graphs~~
    - ~~Deployment (?)~~

### Friday 5/12

* None

### Thursday 5/11

* ~~reconvene to confirm topic~~
* ~~fill in Deliverables section~~
* ~~outline loose schedule~~
* ~~deliver our project idea to instructional team~~
* ~~double check weekend coworking?~~

# Meeting Notes

## Meeting 1

* Deepfake Image Detection
  + Data: [source](https://zenodo.org/record/5528418#.ZFl1m3bMKHu), [kaggle mirror](https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images)
  + Thurs after convo NNs, discuss if want to pursue deepfake
  + Backup: classifier on misinformation