1. What is your name?

Tyler Supersad

2. What is your Student ID?

SUP20486969

3. What programme are you on?

BSc Computer Science

4. What is the working title of your project (this can be changed at a later date)?

Deepfake Detective

5. What is the theme of your project?

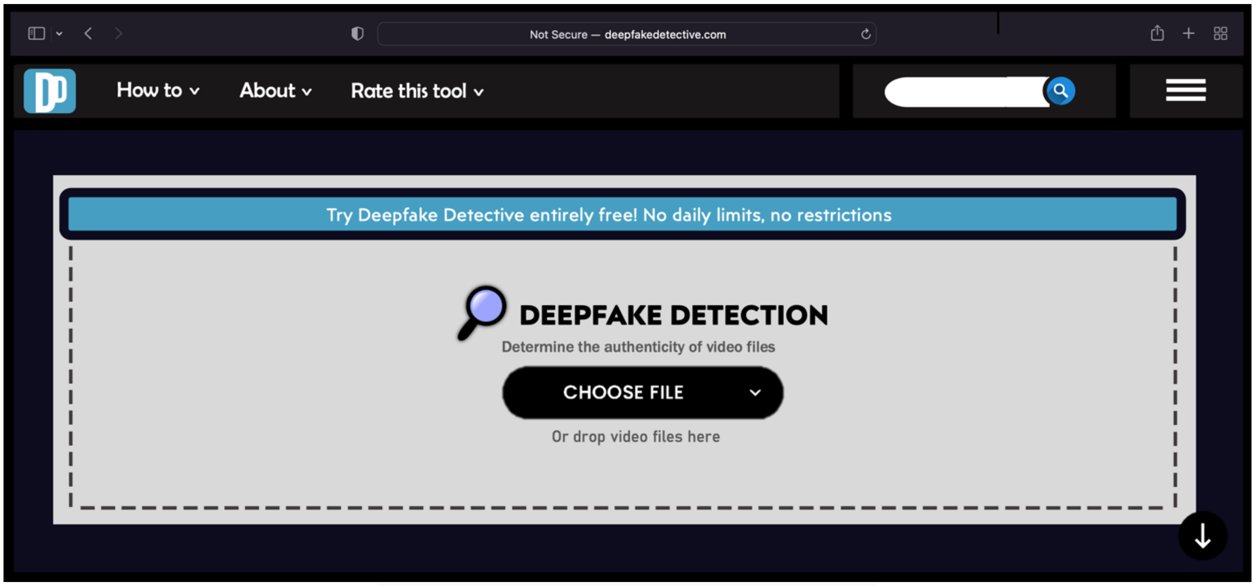
Artificial Intelligence & Machine Learning

6. Describe your project in 500 words or less.

Deepfake Detective is an application that discriminates between real and deepfake from a particular video uploaded by the user. *It is important to note that the term 'deepfake' extends beyond visual manipulation to the alteration of audio in video, however for the scope of my project, I will solely focus on the visual aspect of deepfakes.* To continue, such uploads will undergo a series of evaluations from a detection model that'll eventually predict the overall probability of a video being a deepfake. Depending on the sway of the probability, the video will be given a binary classification of *real* or *fake* which will prominently be displayed on the webpage for the user to see.

Through efficient training and testing using Python, the deepfake detection model will be enabled by a vast dataset, specifically designed, and shared from *Meta AI* to achieve related objectives. With over 100,000 total original/deepfake clips sourced from 3,426 paid actors, the training data will be composed of their frames (specifically frames with an evident face) annotated as *real* or *fake*. Per their usage, the model will conduct a process called, Object Detection, to (1) learn and identify an object in the frame/image (300x300 pixels) as a discernable human face, and (2) correctly label that object as real or fake. Once model training is completed, testing and validation may occur to discover the most optimal model to feed the user video where it too will be fragmentized frame by frame for the deepfake evaluation.

With Python serving as the backend to deliver the deepfake classification, JavaScript will be utilized for building the front-end of the web app. As presented in Figure 1, the Deepfake Detective concept is very simple by design, but it fulfills its intended purpose of allowing users to select or drop their video file for an evaluation. (303 words)



Try Deepfake Detective entirely free! No daily limits, no restrictions

Determine the authenticity of video files

Or drop video files here

Figure 1: Deepfake Detective Concept

7. Please list up to three aims of your project. An aim is an expected outcome of your project (e.g., issues it will address, how it might improve or enhance a situation for stakeholders, etc.).

Due to the power of Artificial Intelligence, deepfake technology has enhanced multimedia data manipulation by making them able to be executed with ease. As this becomes an imminent threat instigating malevolence towards privacy, democracy and national security, the planned implementation of this project would (1) provide an extensive overview of deepfake methods and propositions of probable forces to combat them, (2) pose thorough discussions on challenges and research analysis related to deepfake technologies, and (3) detect and determine the integrity of digital visual media to help guard against the aforementioned dangers of their manipulation.

8. Please list up to five artefact objectives of your project. Objectives are tangible tasks that you will complete. They are typically steps/activities that you must complete in order to deliver your project successfully.

Objectives that my project necessitates to perform adequately include:
1). Load and handle dataset with over 100,000 original & deepfake clips.
    1a). Use face recognition package to locate face landmarks in a clip.
    1b). Add padding to zoom out of face when a face is located.
    1c). Annotate each frame of the video as *real* or *fake* depending on the clip classification.
2). Collect and store all frames of the input video.
3). Utilize face detection to pull all discernable faces from each frame.
4). Plot facial landmarks for each frame and measure frame-by-frame facial disparities.
5). Obtain a deepfake probability of the available frames and determine the overall classification of the video.
6). Integrate JavaScript front-end with the Pythonic back-end detection model.

9. Please list background/literature/technology review sources that you have used to inform your project (e.g., evidence of research, links, and sources).

[1] S. Lyu, “Deepfake Detection: Current Challenges and Next Steps,” *IEEE Xplore*, 2020. https://ieeexplore.ieee.org/document/9105991 (accessed Apr. 22, 2021).

[2] T. T. Nguyen *et al.*, “Deep learning for deepfakes creation and detection: A survey,” *Computer Vision and Image Understanding*, p. 103525, Jul. 2022, doi: 10.1016/j.cviu.2022.103525.

[3] L. Bondi, E. Daniele Cannas, P. Bestagini, and S. Tubaro, “Training Strategies and Data Augmentations in CNN-based DeepFake Video Detection,” *IEEE Xplore*, Dec. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9360901 (accessed Aug. 20, 2021).

[4] L. Guarnera, O. Giudice, and S. Battiato, “Fighting Deepfake by Exposing the Convolutional Traces on Images,” *IEEE Access*, vol. 8, pp. 165085–165098, 2020, doi: 10.1109/access.2020.3023037.

[5] P. Korshunov and S. Marcel, “Subjective and Objective Evaluation of Deepfake Videos,” *IEEE Xplore*, Jun. 01, 2021. https://ieeexplore.ieee.org/abstract/document/9414258?casa_token=vqZKsOBPqw0AAAAA:hxBDs-qdZ9LXykwnjC6pi3i9eFcZ3deSk5UTqiDGOTV8gV0nCbn7rZOYcoRjahZulWAWcA5FJ4jm (accessed Oct. 22, 2022).

10. Please describe any risks, ethical issues, or other factors that your project may have to consider.

As a project manager, I have identified a few risks that I may encounter while developing Deepfake Detective. Some risks include low performance of my detection model, and time. With regard to delivering a low performance detection model, I view a possibility where I am unable to consider and solve proper variables that are required to construct a trained model with an outstanding capability of identifying faces that are real and faces that are fake. Although my goal isn’t to deliver a model with 99.99% accuracy, I do aim to deliver one that gets relatively close. The following are a few queries that can hinder my model performance, if not solved:

- How to handle videos with blatant cuts/edits?
- How to handle frames with multiple objects (human faces)?
- How to not confuse motion blur as a disparity related to deepfakes?

Another fear I possess is relevant towards my time management for this project. Generally, I schedule deadlines for specific tasks that need to be completed in a set amount of time; however as seen in the last half of the Spring Semester (2nd Year – Software Engineering), I struggle moving past a certain deadline due to my desire to achieve a great result, especially on a stylistic standpoint. To mitigate this issue and ensure I do not impact my delivery time negatively, I will instruct and follow a regimented schedule for myself. It is also worth noting that I will be making the front-end ancillary compared to the construction of my detection model as a restrictive method to develop the back-end to the best of my ability.

As far as I am aware, I will not come upon any ethical issues as the dataset I’ll be utilizing has been designed expressly for research on deepfakes and features more than 3,500 different paid actors, each of whom consented to participate (according to Meta AI).