
DATA SCIENCE ACADEMY

# 1 Introduction

Developments in media creation technologies have been progressing over the years. In this digital era, with the advancements in Artificial Intelligence, and Machine Learning (ML) in particular, we have models capable of performing tasks that surpass the human performance, and generative ML models are no exception. We have models able to create realistic content starting from pure noise. In the right hands and with the right intent, these developments are a powerful tool to generate engagement and explore different communication paradigms. The topic is getting explored, and we now have ML models that can create images, text, audio and videos from scratch with the proper data in place (figure 1).

In the top right of figure 1 we have an approach to generate videos, commonly referred as Deep fake, that makes use of two ML models and with little effort can create realistic outputs of a person impersonating other. This content may confuse many sources of content and information. Several of these deepfakes have proliferated through many social networks platforms, and they are reaching millions of people every day. This false content can be harmful and provoke catastrophic consequences if not detected as such. On a more personal level, imagine that someone creates a deep fake of your face, and uses it to access confidential information, such as bank accounts, or your mobile phone. Thus, we are on an era that what we see, hear or read requires ways for fact verification. Moreover, research efforts have been created to study and develop models that could detect if the content is a fabrication, or it is true if it is fake or real.

In this work, we are going to embrace this theme by creating models that are able to detect if a given content is real or fake. For this, we will focus on

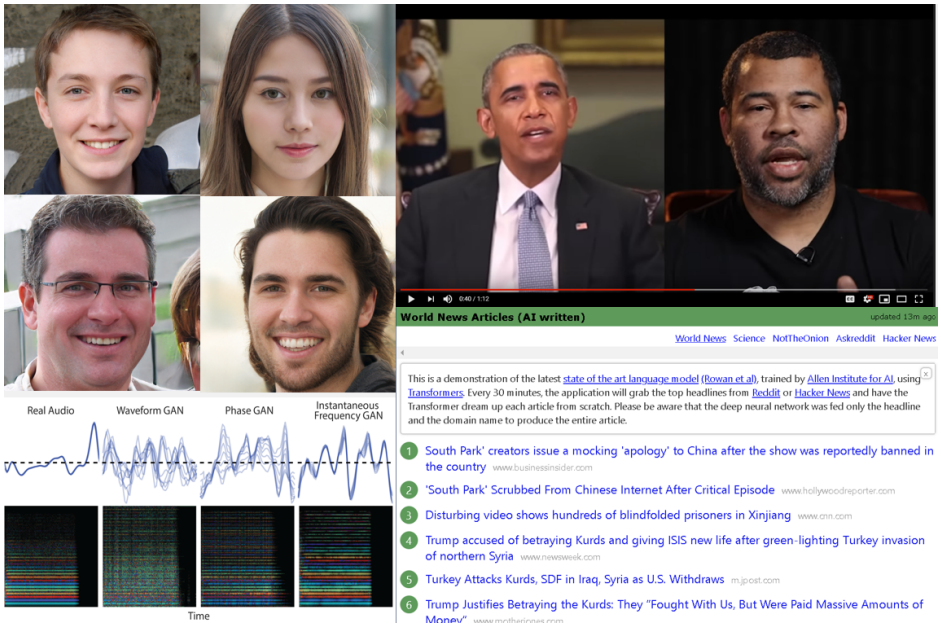


Figure 1: Deep fake example, from: https://www.youtube.com/watch?v=cQ54GDm1eL0

the image domain, and a dataset will be provided for you to train and test your solutions.

## 2 Real or Fake

In general terms, the main objective is to analyse and explore the image dataset provided and create an approach that can distinguish between real and fake images. In figure 2 we have images from both classes.

To fulfil this task, you should tend to the following objectives:

- Prepare the pipeline necessary to process the data and create a model to classify between real and fake faces.
- Use model selection and parameterization techniques to improve your model.

Based on the knowledge gathered from previous modules, you should apply what you have learned to solve this particular problem.

The dataset is divided into a training dataset and validation dataset. Thus, the training dataset should be used to prepare your approach and models and then you should test with the validation dataset. Moreover, the

Figure 2: Sample of images from the dataset. Can you tell what is real and what is fake?

dataset is organised by folders of each class, real and fake. You are in charge of preparing the data, analyse it and pre-process it has you see fit. Based on this step of data preparation and analysis, you should then train your models. The link to the dataset is the following:

- http://shorturl.at/ovxGI

We will resort to the validation dataset to see how fit is the models that we are training/creating. Thus, the validation part of this work is to be measured in terms of the following performance metrics:

- **Accuracy**
- **Precision**
- **Recall**
- **AUC** - Area under the Receiver operating characteristic (ROC).

You should prepare your code to report the aforementioned metrics. Moreover, you should provide, at least, the **confusion matrix** for the validation dataset. As a challenge for this part, you can add additional information about the performance by visualising the results and extract further analysis and metrics that are suitable, i.e. export the ROC curve, precision recall curves or a plot of wrongly classified examples.