**Overview:**

The objective of this project is to create an intelligent bot that will play the game of rock, paper and scissors with the user using audio or image input. I will be recording audio and taking pictures of myself, friends and family to create a dataset to train models on. A neural network will be used to find patterns in the data for recognising the users input alongside a convolutional layer for better handling of large amounts of data in the form of an image or audio.

**Aims:**

Must:

* Make use of a neural network to allow the user to play an intelligent game of rock, paper and scissors
* Use a 2D Convolution layer in the network to allow for image classification
* Have a good degree of accuracy and useability for the general user, allowing some different hand types to be recognised
* Allow the user to play the game of rock, paper and scissors via image input

Should:

* Allow the user to play the game of rock, paper and scissors via audio input
* Use a 1D Convolution layer in the network to allow for audio classification
* Have a high degree of accuracy and useability for the general user, allowing many different voice and/or hand types to be recognised
* Allow the user to easily and intuitively ask to play a game via audio or image input as well as switch between searching for tweets and playing anytime

Could:

* Use object detection to allow the user to play the game of rock, paper and scissors through a video input
* Have a near perfect degree of accuracy and useability for the general user, allowing most different voice and/or hand types to be recognised

**System Requirements:**

System:

All computers running any major non mobile OS should be able to run the program with no additional hardware or peripherals required except a standard keyboard and webcam. Python is required to be installed as well as several Python libraries:

* All Previous Libraries from part 1
* NumPy
* TensorFlow
* Keras
* Sounddevice
* PIL

User:

* Users can ask to play a game with the chat bot and can compete in a game of rock, paper and scissors (Commands: “Lets play a game”)
* Users can take a photo (with the program) to input their choice for the game
* Users can record audio (with the program) to input their choice for the game
* Users can record video (with the program) to input their choice for the game

**Techniques to use:**

Neural Network:

A neural network consists of 3 main sections, the input layer, output layer and hidden layer.

The input layer consists of all the data points we intended to put into the network for picking up on patterns and trends, if we had a 20x20 image we could take the data from each pixel of the image and feed this into the input layer meaning we would need 400 input neurons.

The hidden layer is where all the pattern finding is done, each neuron in this layer is a function of all the neurons in the previous layer that will be looking for some specific pattern in the data once it is trained, each neuron we add can look for a different pattern and each layer we add allows the network to pick up on more complex patterns, this layer is trained via back propagation which adjusts the weightings in each function (neuron) depending on what outcome we got and what outcome we actually want in the output layer.

The output layer consists of the classes we would like the model to choose between, in this case 3 neurons, one for rock one for paper and one for scissors. The number outputted to each neuron is how likely the network thinks the input is that class, the neuron with the highest number is the guess the network would make. The network will have to be shown many labelled examples of the classes so that neurons can be adjusted to pick up on the patterns present in each class.

1D/2D Convolution:

A convolutional layer will help reduce noise and the amount of data points inputted from an image. It will do this by creating a matrix to pass over the image and combining the pixels underneath the matrix via the weight in each position of the matrix, the matrix will be moved along the image until each pixel has been seen by a position in the matrix, these weights will be adjusted through back propagation just like our neurons in the neural network.

**Data pre-processing:**

For images I would like just about anyone to use the program which would be very hard to do as people have many different cameras and environments/backgrounds. To counter this, I intended to do 2 different things. Images in my training data will be taken with many different lighting conditions to allow the model to generalize to more hands and a cropping and resizing feature will be implemented so that the image is focused on just the hand and nothing more.

For audio, all clips trained and recorded will be exactly 2 seconds so that the network receives the same size input without data loss. To help the model generalize I will instruct participants I am recording for my training data to say the lines in multiple voices and pitches so more voices can be picked up on in future.