Chair: Kevin Naidoo, Minutes: Johann

Apologies: Professor Scott Hazlehurst, Ahmed Ibrahim

Minutes:

Ishamel and Nkosingiphile: Managed to locate the given data on the cluster. The data is sorted into different intensities (red and green) across multiple folders. The main investigation has begun to convert data into vectors using a system that doesn't need to know the size of the data.

Raphi and Ben: Facial recognition – existing facial recognition models with will used and tweaked to fit the scope of the project. Creating a new model would require too many resources to be viable. The prebuilt model should be able to successfully identify gender, race, eye colour, and hair. To identify features on a face 68 points are used for detection, which unfortunately does not include hair. Further investigation is being done to improve racial and hairstyle recognition. The error rate is still relatively high and is caused by additional features on the photos that aren't specifically trained for, e.g. beards or very diverse skin tones. Investigating RGB – HEX conversion.

Prof. Bekker: Remember to document adversarial events that go wrong. Maybe consider focusing mainly of eyes, e.g., compensate for mask wearers. Consider Dall-E, Mid Journey and Stable Diffusion as interesting ML resources.

Rael and Gia: The air quality sensor device cannot be powered by the main power supply. Currently testing how long the device can run using powerbanks. Ideally the device should be powered for 2 weeks to a month on a single powerbank. Unfortunately, an Arduino cannot be used as power supply. IOT: A suggestion was made to use Adafruit and a webhook instead of polling. The data captured by the sensor is also stored on an SD card. The device will measure the humidity in an area. The humidity can influence the air quality readings.

Joe and Hraklas: The ML model to track protests is progressing. Currently the main sources of data are the ACBD(?) protest websites and twitter. Twitter is used to identify date and location of tweets about protests in SA. Managed to pinpoint IP addresses within a bounding box that could localize an area with a protest. IF the location is too accurate, the ethical consideration should be considered.

Tristan: Using some of Prof Hazlehurst's developed workflows with additional tweaking and testing identifying the allocated nodes on the cluster that receive jobs can be identified. For the moment, the process is static, and the data created for testing. The following week will be used to make the process dynamic.

Devlan and Chavi: The firewall visualization is currently explored using FireViz. Blocked requests are shown in red, passing requests are red. For now, it works on a single machine/IP and created a graph to visualize the firewall. The aim is to implement on a complete network. The ruleset for the cluster is needed, but Prof Hazlehurst is withholding the information for the time being.

James and Johann: Capturing the data from twitter is proving a lot more complex than initially anticipated. Only 1% of tweets contain the geographical data. However, the user's bio might contain a

location tag and will be explored. The visualization of the sociogram is complex, because titter users have large amount followers and an when going even as deep as 3 layers, the dataset becomes large and complex. A step-by-step approach is implemented.

Mohamed and Thabo: Finding datasets for pool water test strips is proving trickier than anticipated. For the MVP, we are building a test strip reader that identifies the colors on a plain background. Android Studio is used to create the basic application. Currently the linking of the app to an image processing technique is being explored. The position of the different color tests, e.g., chlorine level, for now will be static.

Kevin: Exploring pattern recognition using TensorFlow. Managed to create a CNN that can identify images of cats and dogs. Further patterns will be used for identification to steer the CNN in the direction of archeological pattern recognition.

General:

Reminder that the weekly cohort meetings are compulsory and that both members of each group must attend. Additionally, the meeting starts at 11:00 exactly and late-comers will be reprimanded.