**Notes:**

**Can algorithms truly be fair:**

Algorithms have many biases that arise unbeknownst to the user or even the developer due to a multitude of reasons. One of these reasons is due to the use of biased training data. AI is trained using millions of historical data points which will not be representative of the world as it is right now. Also, A lot of historical data is outdated due to stereotypes. There are also many types of bias such as historical bias – as spoken about already –, automation bias, selection bias, etc. These biases skew the fairness of an AI algorithm, as it does not allow for the results to be representative of the population. An example of this is the COMPAS algorithm by ProPublica which mislabelled white defendants as low risk as well as being twice more likely to label black defendants as future criminals in comparison to white defendants. This may arise due to unconscious bias in the AI developers but also may arise from historical bias. Black people and people of colour have historically been marginalised and criminalised for centuries, so in the past, many incarcerated people would have been these people. If this data was used to train the AI model, then based on the law of probability, then the results of the system will clearly be biased as was shown from the case study. In order to reduce these biases, the AI developers could use more representative data sets using strategic sampling strategies for example, stratified sampling which could be fed to the AI so the results are more accurate and inclusive while reducing bias.