Assignment 2: Critical Analysis of a Chosen ML Ethics Scenario
Student Name: Matthew Elliott
Student Number: 40153557

Part 1: Choices

My choice of scenario is that of "**Face Recognition-based Entry into University Buildings**", and the use of using ML models to recognise the face of a person that has access to a building based on existing images of their face. My choice of ethical consideration is **fairness**.

Part 2: Task characterisation and Issue Identification

Task characterisation: Biometric recognition is becoming more a part of our everyday life where people swap credit cards for fingerprints and ID for facial recognition. I envisage this ML task as making use of multiple photographs taken of students and staff which is used to train a model that is able to correctly identify a student or stall member on a previously unseen picture of their face in real time.

At the man entrance to each building, there will be a camera with a monitor displaying the camera view. The staff member or student will have to walk up to the camera and the system should recognise their face. The system then checks if the person is allowed access to that building and open the doors accordingly. If the doors do not open a reason will be displayed for the denied entry (didn't recognise face, not allowed access etc…).

The accuracy and reliability of the facial recognition system is paramount. A false positive could mean someone who should not be authorised to enter the building can. If they had malicious intent, they could end up stealing equipment, data or disrupting the daily activities of the building. This means the ML model will have to have a very high level of confidence in the identification to ensure that the face is correctly analysed and allowed or denied accordingly. A false negative would mean denying a member of staff or a student access to the building. If this happened sparsely to the university population, it could cause frustration. However, if it was found out that the facial recognition cameras were consistently denying entry to a specific subgroup of the population, it could lead to very serious discrimination cases and a public relations nightmare.

Identified Issues: There are several issues in relation to fairness of the model. These issues of fairness will affect one subgroup more than another, simply based on characteristics that the individual in question is not responsible for. Say for example race, sex, age and time of day (in relation to lighting conditions) that they are able to visit the buildings.

Race bias: Every four years, the National Institute of Standards and Technology (NIST) voluntarily evaluates current facial recognition models. As a testament as to how increasingly sophisticated the systems are becoming, they saw a 10-fold increase of accuracy of facial recognition. Worryingly, it was found that the false positive rate of West and East African and East Asian people can be up too 100 times greater than white people. Asian and American Indian individuals were found to have a higher false negative rate compared to individuals with Caucasian and African faces [1]. Without due caution, it is possible that this bias will appear in the QUB entrance system ML model.

Age bias: Due to the way our faces age as we get older, it is possible that there can exist an age bias on facial accuracy. To investigate possible age bias that exists in facial recognition technologies, [2] analysed the recognition rates from three age groups: 18-30, 30-50 and 50-70. It was found that of the three age groups, the youngest age group was discriminated against, consistently having the lowest recognition rate on all the various commercial models tested. There was no clear privileged group between the other two

age groups. The discrimination against the younger population is likely due to less differentiating features in an individual's face. People age differently and gain a more unique look to their face which can be taken advantage of.

Gender bias: The NIST found that on analysing gender on facial recognition rates, there was a bias against women as they had a higher false negative recognition rate. Fortunately, this difference in false negative rates was smaller than the observed difference in race and ethnicity as talked about above. [2] agreed with these findings.

ML models: It is possible that the underlying architecture of a ML model will lead to different bias towards different groups in the testing data. When investigating the ethical concerns of facial recognition auditing, Inioluwa Deborah Raji et al. used and analised the results from commercial ML models from Microsoft, Amazon and Clarifai [3].

	Gender	Age	Name	Smile	Detection(AP_{50})
Microsoft	0.25% **(DM/LM/LF - DF)**	29.47% (LF-DF)	3.90% (LF-DF)	8.02% (LF-LM)	4.25% (LM-DM)
Amazon	0.50% (LF-DF)	29.10%(LM-DF)	6.71% (DM-DF)	9.75% (DF-LM)	0.75% **(LM-DF/LF)**
Clarifai	19.10% (LM-DF)	11.21% (LM-DF)	10.50% (LM-DF)	3.00% (LF-LM)	0.50% **(LM/LF-DF)**

Table 1: Difference in accuracy between the best and worst performing intersectional subgroups by prediction task. The subgroups are darker females (DF), darker males (DM), lighter females (LF) and lighter males (LM) [3].

As seen from table 1, from all machine learning models across all categories exhibited varying levels of bias, confirming that different architectures produced different levels of bias. Looking at the Name column used for facial identification, all models favoured a person that was male or had light skin, or both. Dark-skinned females were consistently discriminated against in all models. This is consistent with the age, race and gender findings from above.

Data set bias: In a university, most of the population are students. Of these students, the vast majority fall between the 18-25 year old age category. While it is expected that the demographics of the staff and student population is much more diverse in terms of race and ethnicity largely due to the international student population compared to the native population, it is evident from personal observations that there is a skew towards the population being white. There is no evident skew based on gender. Having less training data on a sub-group of one of the mentioned groups above (age, race…) could potentially introduce a bias against that sub-group due to the fact that there will be less of that sub-group for the ML model to train on. This effect can be seen when we consider that even though is it common for Asian faces to be falsely identified up to 100 times more often than white [4], it is common for this trend to be reversed in models that originate from China [5].

Time of day bias: While this form of bias may not be initially obvious, consider the person that has a full-time job, or the single parent that only has time to access the campus after dark when the lighting conditions are drastically different from that of the day. Eric P. Kukula & Stephen J. Elliott from Purdue University aimed to evaluate the performance of commercially available facial recognition models under three different lighting conditions: low, medium and high. The study found that the more the light conditions vary between training and testing on a face, the lower the recognition rate. In particular, training the model with low levels of light was found to be the most detrimental on race recognition accuracy [6]. Thus, if a person was to approach the building at night, they may experience a lower recognition rate, assuming the training images of their face were in a brightly lit environment. However, even if an illuminating light was installed next the facial recognition cameras for use at night, there still may be a bias present if the face is illuminated differently due to the colour or direction that light is coming from.

Twin / look-alike bias: If two people have extremely similar facial features, this pair of people may be discriminated against, by the system recognising one person as another. This could negatively affect their attendance record or in-person services that requires them to be first registered as being present inside the building.

Part 3: Addressing identified issues

Addressing dataset bias:

There are many studies to suggest that the bias as seen from the examples of the identified issues above is due to the training data not having enough samples of a particular demographic. As identified above, [5] highlighted that models from China reverses the discrimination against Asian people commonly seen on ML models from the West.

Evenly distributed training dataset: One solution is to use a database that is more balanced across races, sex, gender and age. This can be achieved by constructing a dataset with both students that are in attendance of the university as well as dummy data of specific sub-categories where required to ensure fair distribution of the mentioned categories. This same principal can be applied to lighting conditions. Multiple pictures of the students and staff members should be taken under different lighting conditions and all entered into the training data of the model. The same applies to the dummy data, ensuring there is an even distribution of lighting conditions for the model to train on.

ML model construction:

Per-demographic tuning: Another solution to this problem of facial recognition bias on ethnicity and gender was proposed in the paper "Face Recognition: Too Bias, or Not Too Bias?" [7]. The authors proposed first identifying any skew in performance of model for any demographic. On these findings, a per-demographic identification confidence threshold was calculated. This allowed the model to level out it's false positive rate and achieve a higher true positive rate. However, initially classifying the test subjects' race and gender in real-time brings about its own ML fairness considerations. It is noteworthy that Cynthia M. Cook at. Al showed that race played larger discriminatory factor for fairness than gender in facial recognition [8]. Now consider the previous statement with the hypothesis that classifying race will be a much simpler task than classifying gender in terms of complexity and additional fairness considerations. I propose QUB should explore per-demographic tuning based on the race of the test subject.

Improved hardware

*High quality cameras to detect subtle facial features***:** Ensuring the quality of the camera that is taking the picture of the person in question is essential. Using the skin texture features extracted from the area adjacent to the nose and eyebrows, [9] is accurate enough distinguish between a set of identical twins. However, this solution is likely not compatible when you consider the cost of the high-quality camera equipment needed along with how precise the image of the forehead would need to be. There has been no work assessing the accuracy of this under different environmental and lighting settings.

Considering other forms of biometric identification

*Use of iris data***:** Y Sirotin [10] showed that iris scans showed no higher rates of false positive identification across race, gender or age demographics. While the use of an iris scanner may sound compelling on paper, and could theoretically solve almost all of the fairness problems described above, in the real world this may not be ideal. A detailed cost analysis and well as the convenience of iris scanning would have to be taken into consideration. Ideally, the iris scanner would not be necessary if all the solutions above are explored and implemented.

Bibliography

[1] NIST, "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects," 2019. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf.

[2] M. I. M. J. B. S. M. I. J. C. K. R. W. V. B. M. I. a. A. K. J. F. I. Brendan F. Klare, "Face Recognition Performance: Role of Demographic Information," December 2012. [Online]. Available: https://ieeexplore.ieee.org/document/6327355?arnumber=6327355&tag=1.

[3] I. D. R. e. al., "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing," 2020. [Online]. Available: https://arxiv.org/pdf/2001.00964.pdf.

[4] BBC, "Facial recognition fails on race, government study says," 20 December 2020. [Online]. Available: https://www.bbc.com/news/technology-50865437.

[5] J. A. F. J. A. N. A. J. O. P. Jonathon Phillips, An Other Race Effect for Face Recognition Algorithms, 2010.

[6] E. P. K. &. S. J. Elliott, "Evaluation of a Facial Recognition Algorithm Across Three Illumination Conditions," September 2004. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1346921.

[7] J. P. Robinson, "Face Recognition: Too Bias, or Not Too Bias?," 2020. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2020/papers/w1/Robinson_Face_Recognition_Too_Bias_or_Not_Too_Bias_CVPRW_2020_paper.pdf.

[8] J. J. H. ,. M. I. Y. B. S. ,. M. I. J. L. T. a. A. R. V. Cynthia M. Cook, "Demographic Effects in Facial Recognition and Their Dependence on Image Acquisition: An Evaluation of Eleven Commercial Systems," January 2019. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8636231.

[9] B. C. DeLean, "Vision-based operating method and system," 2002. [Online]. Available: https://patents.google.com/patent/US7369685B2/en.

[10] Y. Sirotin, "A comparison of demographic effects in face and iris recognition.," 2019. [Online].

[11] G. L. Y. H. C. Q. ,. Y. F. a. S. T. Joseph P Robinson, "BFW dataset," 13 December 2020. [Online]. Available: https://github.com/visionjo/facerec-bias-bfw.

[12] G. L. Y. H. C. Q. Y. F. a. S. T. Joseph P Robinson, "https://openaccess.thecvf.com/content_CVPRW_2020/papers/w1/Robinson_Face_Recognition_Too_Bias_or_Not_Too_Bias_CVPRW_2020_paper.pdf," 2020. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2020/papers/w1/Robinson_Face_Recognition_Too_Bias_or_Not_Too_Bias_CVPRW_2020_paper.pdf.

[13] Jones Day, "Proposed Algorithmic Accountability Act Targets Bias in Artificial Intelligence," June 2019. [Online]. Available: https://www.jonesday.com/en/insights/2019/06/proposed-algorithmic-accountability-act.