Facial Recognition Software and Racial Bias - 800 words

Name : Kweku E. Acquaye

Q: For your chosen topic,
Discuss:
- *The issue*
- *The reason or source of the issue*
- *The effect(s) or consequence(s) of the issue*
- *Possible solutions to the issue*

A:
The Issue
The use of facial recognition algorithms and software (FRAS) is a benefitial, indispensable, and inevitable part of today's and future societies. It provides convenience, speed, cost-effectiveness, and security in the process of identification in public buildings, restricted buildings/places of work, airports, national borders, and so on.

To date, FRAS has depended on graphs of facial markers, i.e. basically sets of distances and ratios of facial features, to return a positive identification and thus identify individuals. FRAS trained in predominantly white caucasian countries have ubiquitously been assessed to be biased against some racial minority groups[1,2,3,4,5]. The rate of false positive identification is particularly worrying among ethnic Black populations, and also among South Asians and Far Earstern Asians. Considering that people of Black ethnicity are already disproportionately, adversely, and unfairly represented in arrest and incarceration statistics for all manner of reasons, societies that rely on FRAS AI systems with their current false positive rates and inherent bias against minorities stand in danger of the accusation of entrenching and multiplying injustice.

The occurrence of this issue is not in doubt – several published cases have mainly been by fair-minded white caucasian researchers who have nothing to gain in putting forth such alerts except the desire to see a just and equitable society.

Reason for/Source of Issue
It would be fair to say, in my opinion, that FRAS bias is not a result or manifestation of deliberate malice, but an example of early unconscious bias in the data used to train the AI models. Considering that identification is obtained by matching graphs of facial features, it is easy to see why models trained on predominantly white caucasian samples, when confronted with subjects of darker skin colour, would not be able to determine the correct distances between say the corners of the mouth, the centre of eye pupils to the tip of the nose and/or tip of the chin, e.t.c. and their fractional relationship to other features and distances, i.e. ratios that define the graph.

There is currently sufficient awareness of this issue due to several publications in academic journals[4,6], and also in popular magazines[2,3] and newspapers. It remains to be seen what remedial action builders of FRAS would take to address the problem. In the meantime, those in the AI Ethics community have already taken steps to advocate and publish guidelines on how to address this[7,8,9].

Effect of Issue
The harm to society of AI systems mis-identifying citizens is enormous, and affects all regardless of race. Individuals who fall foul of such mis-identification can be arrested, questioned, have their homes searched and turned upside down, sectioned under the mental health act, face trial and be wrongfully imprisoned. As mentioned above, considering that people of Black ethnicity are already disproportionately represented at the wrong end of the justice system for all manner of reasons – some historic e.g. deprived backgrounds, and some plainly wrong e.g. mistrust and irrational fear, it is particularly grievous to add another tool and layer of such importance. In 1999 The Metropolitan Police was found to be "institutionally racist" in a Judge-led public enquiry[10]. In recent times there have been other similar high profile cases, e.g. Bibaa Henry and Nicole Smallman pictures scandal[11].

Adding another bone of contention to communities who already mistrust the Police and the Justice system is counterproductive for all. Since western politicians, newspapers, and intelligentsia tend to be quick to point fingers at other nations and accuse them of trampling human rights, it may be benefitial if western democracies are not seen to be presiding over a different kinds of, but equally unpalatable, human rights abuses in their midst.

Possible Solutions
FRAS should be trained on more representative data, i.e. a lot more data from ethnic minorities should be included in the training, validation, and test data to address class imbalance.

Positive identifications from FRAS should be confirmed by other traditional means before individuals are accosted or arrested, until a time in the future when false positive rates fall below 0.00001% (i.e. 1 in 100,000).

References

1. FRVT Pt3. Verification Project, NIST.
    [online]: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

2. Garvie, C. and Frankle, (2016), *Facial-Recognition Software Might Have a Racial Bias Problem,* The Atlantic Technology Archive, pp. 1-6. [online] : https://www.theatlantic.com/technology/archive/2016/04/the-underl/ , [link]: https://drive.google.com/file/d/1tl1KoXk2Py5BqzGpI8EF3cpRy8zJiGNM/view?usp=sharing

3. Rauenzahn, B., Chung, J., and Kaufman, A. (2021), *Facing Bias in Facial Recognition Technology*, The Review Publication, Mar 20, [online]: https://www.theregreview.org/2021/03/20/saturday-seminar-facing-bias-in-facial-recognition-technology/

4. Leslie, D. (2020), *Understanding bias in facial recognition technologies: an explainer*, The Alan Turing Institute.
    [online]: https://doi.org/10.5281/zenodo.4050457

5. Hardesty, L. (2018), *Study finds gender and skin-type bias in commercial artificial-intelligence systems*, MIT News, Feb 11, [online]: https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

6. Castelvecchi, D. (2020), *Is facial recognition too biased to be let loose?*, Nature **587**, pp. 347-349, [online]; https://www.nature.com/articles/d41586-020-03186-4

7. FRVT 1:1 Verification Project, NIST. [online]: https://pages.nist.gov/frvt/html/frvt11.html#_frvt_participation_statistics_

8. Romine, C. (2020), Facial Recognition Technology (FRT), NIST. [online]: https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0

9. Lunter, J. (2020), *Beating the bias in facial recognition technology*, Biometric Technology Today, 2020 (**9**), 5-7. [online]: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7575263/

10. The Sir William Macpherson Report (1999), [online]: https://www.gov.uk/government/publications/the-stephen-lawrence-inquiry
    (to be formatted to acceptable referencing standard)

11. Bibaa Henry and Nicole Smallman Police Scandal (2021), [online]: https://www.bbc.co.uk/news/uk-england-london-59389906
    (to be formatted to acceptable referencing standard)