**Algorithmic Bias**

When we talk about AI taking over the world, we generally think of a Terminator-like robot (Or Ultron- if you're a Marvel fan) killing us all. But a more accurate and scarier version of AI are the algorithms all around us, invisible and pulling our string, making us dance to their tunes- quite literally in some cases.

Since the dawn of the digital age, decision-making in finance, employment, politics, health, and human services has undergone revolutionary change. Today, algorithms rather than humans control which neighborhoods get policed, which families attain needed resources, and who is investigated for fraud. (Eubanks, 2018) Hence, these automated systems play an increasingly significant role in determining the 'winners' and 'losers' in our society. And with the tremendous increase in productivity, accuracy, and insights offered by the algorithms, it is, no doubt, here to stay. But more and more evidence has been put forth regarding how these algorithms also risk magnifying human bias and errors on an unparalleled level. They reflect the values, preferences, and perspectives of the humans who designed them and the data fed to them.

Graphical user interface, application

Description automatically generated

Text

Description automatically generated with medium confidence

The above article (Ahmed, 2020) is a prime example of racial bias, similar to the racial soap dispenser or Google's classification of black faces as 'Gorilla's. Joy Buolamwini does some exciting work to address the issues with facial recognition at MIT media lab through her organization, The Algorithmic Justice League. (Mentioned in the week 6 materials) According to her, this Good Fight isn't just about being recognized. (BUOLAMWINI, n.d.)

Text, website

Description automatically generated

Another avenue for algorithmic bias practice is the newly automated hiring process that perpetuates harmful racially motivated and gendered stereotypes. In the case of the Amazon hiring algorithm, it learnt from previous hiring decisions and marked down applications from women. This is just another in the long list of biases in algorithms. We, developers, are both the perpetrators and the victims in this scenario, and hence we need to understand why this is happening.

**Algorithms VS Data**

While we blame the algorithms, some of the blame is rightfully placed on the data. If we use flawed data to train the algorithm, we must expect bad results. The algorithms try to find the pattern hidden in the data and, through it, classify the data. It's not that the algorithm is biased, but that the data is, and hence the errors are inherited. Primarily if you're using reinforced learning where a step in the right direction is rewarded and one in the wrong direction is penalized so, if the data we're using is biased, we're essentially reinforcing this bias. Of course, this doesn't mean that the algorithm is blameless and just reflects the dataset. Some models are better than the others, and understanding how this works, allows us to pick better models rather than simply depending on the data. We can't always 'correct' the data as the origin of the bias might not be apparent, especially in the case of a high dimensionality problem.

Model design choices to maximize test-set accuracy do not hold static other properties we care about, such as robustness and fairness. (Hooker, 2021) As such, effort should be made to pinpoint the cause of the bias and teak the design accordingly. Just like 'Privacy by Design,' 'Equal Opportunity by Design' or Fairness and Robustness should also be added by design. We can't just absolve the algorithms of all the blame and point the finger at the data; we need to look deeper at the origin of bias and counter it. (Hooker, 2021)

Books like 'Weapons of Math Destruction and 'Automating Inequality: How high-tech tools profile, police, and punish the poor' talk of a Digital Dystopia where the poor and vulnerable are targeted even more and a chilling Kafkaesque quality (Pilkington, 2019) that threatens to envelop us. Another interesting case is the ProPublica article about Machine Bias in the Criminal Justice System.

Graphical user interface, text, website

Description automatically generated

The article is given as a reference. (Julia Angwin, 2016) The mathematical models used for these are often proprietary and have the effect of being a black box. What's worse, there's no recourse when the algorithm makes a mistake!

**What can be done?**

The important thing is that the people who use machine learning-based algorithms in their decision-making process should ensure that historical forms of discriminations present in the data fed to the algorithms are not perpetuated or introduce new previously unknown classes. The UK governments work to include anti-discrimination through measures like Equity, Diversity, and Inclusion at the Centre for Data Ethics and Innovation is a fine example of things to be done.

Every Revolution was looked upon with fear until we learned to regulate it. Datasheets for Datasets is another exciting idea introduced by Gebru. She proposes a datasheet to accompany every dataset that documents its motivation, composition, collection process, recommended uses, and so on. The hope is that Datasheets for datasets will facilitate better communication between dataset creators and dataset consumers and encourage the machine learning community to prioritize transparency and accountability. (Gebru, et al., 2018). The idea does have it’s merit and is getting popular and along with it is another popular idea- AI fairness checklist. (Madaio, et al., 2020) The idea is to adopt and integrate the checklist into AI design and development lifecycles- Fairness by Design.

**Conclusion**:

Instead of fearing for our lives with AI, let’s focus on creating computational techniques that are both innovative and responsible.

# References

Ahmed, M., 2020. UK passport photo checker shows bias against dark-skinned women. *BBC News.*

BUOLAMWINI, J., n.d. *ALGORITHMIC JUSTICE LEAGUE.* [Online]   
Available at: https://www.ajl.org/

Eubanks, V., 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor..* USA: St. Martin's Press, Inc..

Gebru, T., Morgenstern, J. H., Vecchione, B. & Vaughan, W. J., 2018. Datasheets for Datasets. *ArXiv,* Volume abs/1803.09010.

Hooker, S., 2021. Moving beyond “algorithmic bias is a data problem”. pp. Volume 2, Issue 4.

Julia Angwin, J. L. S. M. a. L. K., 2016. Machine Bias. *ProPublica.*

Madaio, M., Stark,, L., Vaughan, J. W. & Wallach, H., 2020. *CoDesigning Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI.* s.l.:s.n.

Pilkington, E., 2019. Digital dystopia: how algorithms punish the poor. *the Guardian.*