Kalyani Kawale Follow
Jan 8 · 6 min read · Listen

Are your algorithms biased?

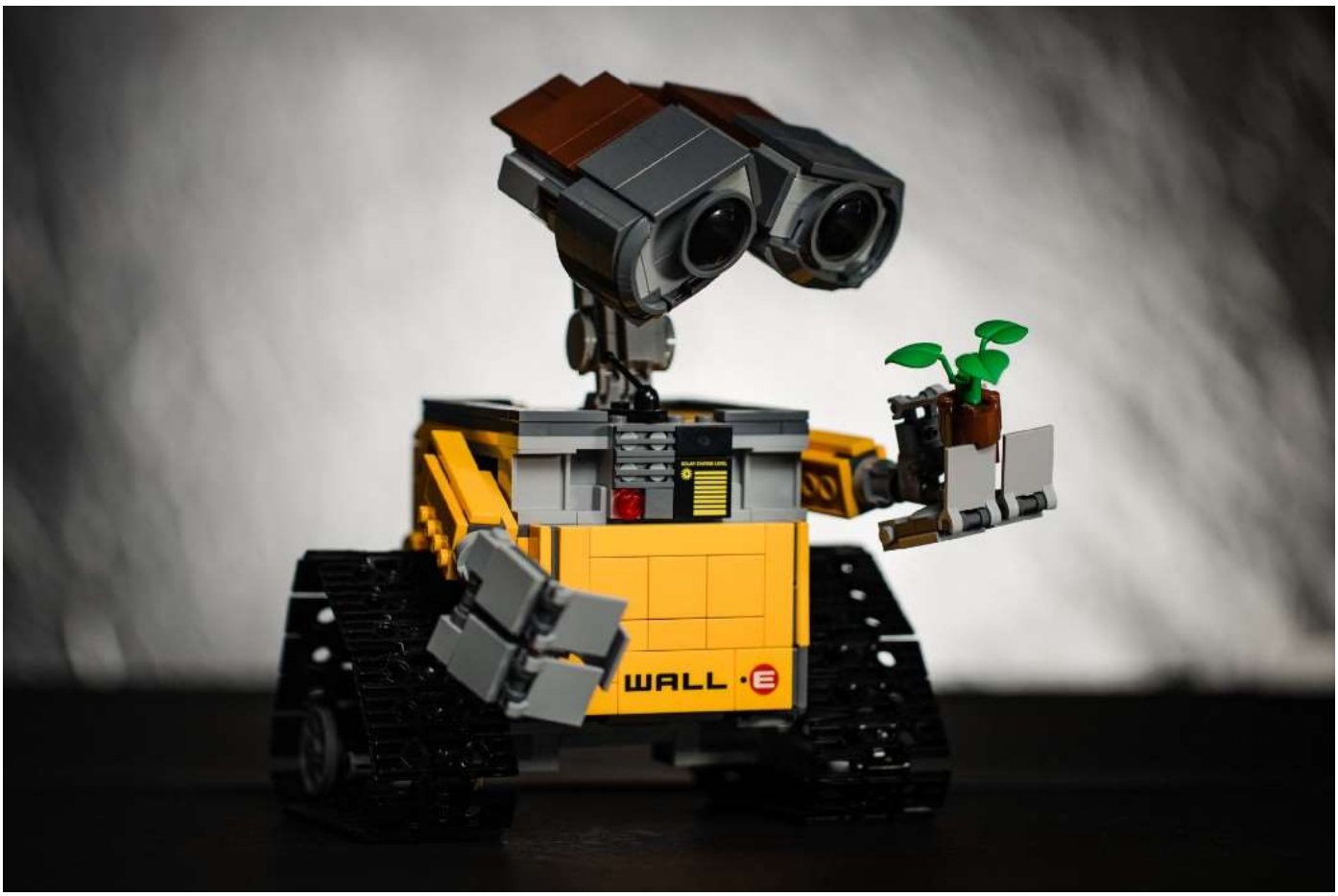

Photo by Jason Leung on Unsplash

Imagine that you are working for a huge multinational corporation and you get an opportunity to work on a Machine Learning project. You are good at Python and have taken this highly rated course on Machine Learning, so, you jump at this opportunity to create an in-house ML application for your company. You end up creating an

you. But, even after months of editing and reuploading your resume every day, you just aren't seeing the results you were hoping for. Eventually, you start losing hope, start questioning your capabilities, start cursing your luck, start thinking of changing your career path, and down the rabbit hole you go, never realizing that it was in fact an unfair AI that was filtering your resume out of the race. Now you must be wondering wouldn't the AI be trained to pick the resume with the most similar features as the job description and you are not wrong. But on the other side was another engineer, just like you, who was given the responsibility of creating this resume filtering AI, who took the dataset given to them, found an algorithm that best fit the dataset and made remarkable predictions for some new, but similar test dataset, while being unaware of the fact that the developed model only worked perfectly for a particular section of the population and not for others. A deeper look into the data and the model predictions would have brought to light that there was one simple feature *distance from the workplace,* which for some reason had been weighted very highly, and denied a huge population of very skilled engineers the opportunities that were clearly a good fit for them.

This is a classic example of what is called Algorithmic Bias. Simply put, an AI algorithm is considered to be biased when it makes decisions that are unfair to certain groups of people and not to others, over and over again [1, 2, 3]. It is important to understand that, this unfairness towards a particular group is not by chance, but by bias. The bias may be racial, gender, disability, or something else entirely which we may not even conceive as a bias because algorithmic bias is not just the reflection of biases that the AI creators hold [4]. An algorithm may behave unfairly due to a number of reasons, ranging from problems in the dataset (such as using historical data with human biases integrated into it), the way an ML model is designed (maybe declaring and removing a feature because you feel it is redundant), to even the way the algorithm is used (maybe it is being used by the bad guys who only intend to make personal profits) [5].

Now that you know this, what do you do? If you are an AI/ML engineer or an aspiring one, are you thinking, what can *I* do about this. *I* don't have the time to think about

of a space telescope like the James Webb Telescope and prevents it from making breakthrough discoveries about our universe because some aspect of your model design makes it biased towards only sending some information back and withholding some other. This may not happen to you in a billion years, but then again it might. An unfair job placement AI system may not seem like a huge problem, but the pace at which the use of AI is growing in our communities may bring about unwanted discrimination if left unchecked. There are dozens of examples out there of biased AI taking decisions in even the matters of life and death such as in the fields of criminal justice or access to loans [6].

The question remains, what can you do about it? For starters, take it seriously. Be aware that the algorithm you are developing may become biased and treat people unfairly. Take a moment and think about the ways biases may get introduced or created in your system, and discuss it with your peers. Try to come up with solutions or at least explore automated options to make your system fair. Know that there are efforts being taken for years now to help address this problem, companies like IBM and PwC have toolkits to analyze biases and maybe you could benefit from those [3, 7].

Algorithmic bias is a term that has vast implications. It is an entire field in itself, which has given rise to a new requirement for fairness debugging tools [8], which would automate the process of finding or predicting whether your application is prone to any biases. But before these tools hit the mainstream AI market, the least we can do is educate ourselves and take measures that are within our reach to make our AI fair.

Sure, it all sounds very idealistic, and granted this is not an easy task, as algorithmic bias comes with a plethora of challenges, especially in the commercial domain, but do you ever fear that AI is going to take over the world and dominate humans? You know, it might, but it doesn't have to. We may end up living in a dystopia, under the constant fear of our robot overlords, or in a utopia, in everlasting peace with AI, which helps humans live better than ever before. And it depends on whether we choose to educate ourselves and take action on apparent problems like algorithmic bias or we let the algorithms become racial and discriminatory. And now, if you are thinking, this is so

the information age is moving way faster than that and we may be approaching the Singularity, not the Physics one but the Technological one [9], within decades and not centuries, beyond which the changes to human civilization itself are thought to be unfathomable. So, think again and next time you are developing an ML application, take a minute to consider the impact it may have on society.

References

[1] Algorithmic bias (2021) Wikipedia. Available at: https://en.wikipedia.org/wiki/Algorithmic_bias (Accessed on: 4 January 2022)

[2] Algorithms and Society. (2020) Algorithmic Bias, Discover How It's Affecting You! 16 August. Available at: https://www.youtube.com/watch?v=SA-Lu_mv5RQ (Accessed: 4 January 2022)

[3] Best, M., Rao, A., and Lendler, J. (2021) Understanding algorithmic bias and how to build trust in AI. Available at: https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html (Accessed: 7 January 2022)

[4] Hooker, S. (2021). Moving beyond "algorithmic bias is a data problem". Patterns. 2. 100241. 10.1016/j.patter.2021.100241.

[5] Algorithms and Society. (2020) Algorithmic Bias, Discover How It's Affecting You! 16 August. Available at: https://www.youtube.com/watch?v=SA-Lu_mv5RQ (Accessed: 4 January 2022)

[6] Heilweil, R. (2020) Why algorithms can be racist and sexist. Available at: https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency (Accessed: 4 January 2022)

[7] Samuel, S. (2019) This AI makes you look like a masterpiece — while teaching you about its own bias. Available at: https://www.vox.com/future-perfect/2019/7/25/20708589/ai-portraits-art-bias-classical-painting (Accessed: 4 January

Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Paper 600, 1–16. DOI: https://doi.org/10.1145/3290605.3300830

[9] Reedy, C. (2017) Kurzweil Claims That the Singularity Will Happen by 2045. Available at: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045 (Accessed: 7 January)