A. OSYPCHUK,
Bachelor student

M. KOSHELEVA,
Senior Lecturer, language advisor
Zhytomyr Polytechnic State Unitversity

BIAS IN AI DECISION-MAKING THE ETHICS OF ALGORITHMIC FAIRNESS

Artificial Intelligence (AI) has impacted many aspects of our daily lives from health care to criminal justice, finance and hiring, and much more. Although AI has helped us a lot it still has flaws. It's all about bias. AI decision-makers can incorporate biases that are present in the algorithms they work with. This thesis looks at the ethical problems that can be caused by AI bias. For AI to assist justice, instead of enhancing inequality, what is required is not only a technical fix but also a thought process fix concerning ethical responsibility at all levels of design [1].

AI systems often inherit biases from their training data. When trained on historical data, these models may learn the biases – relating to race, gender, or socio- economic status – present in the data. Some AI hiring technologies, for instance, have a natural bias for male workers, reflecting the male-dominated hiring patterns in their training data. Such prejudices help to introduce non-discriminatory practices into what are supposed to be neutral algorithms [2].

Poor representation in datasets is another cause of AI bias. AI prediction outcomes are less accurate and result in unjust outcomes if there is an underrepresentation of certain groups. This is particularly troubling in the realms of healthcare and criminal justice. Algorithms can reinforce discrimination against minorities in predictive policing. Due to historical over-policing, the crime data used to train these algorithms places an unfair emphasis on these neighborhoods. As a result of the biased outcomes mentioned above, a feedback loop is created where the police are sent to communities that are already being over-policed [3].

AI biases in finance have seen algorithms that determine whether loans are approved. This credit is denied to people from certain races or poor socio-economic backgrounds at a disproportionately high rate. When people are denied access to financial services because of their identity, it not only impacts the individual but also creates economic inequality. Thus, the developers' ethical responsibility is not only to ensure the technical performance of their models but also to ensure that the AI systems do not perpetuate social injustices.

To tackle bias in AI systems, blending technical solutions and policy intervention along with ethical considerations is required. A good approach to this technically, is the building of fairness-aware algorithms that target the minimization of discriminatory outcomes. These models are specially designed to be sensitive to biases so that predictions do not unfairly penalize a group. For example, algorithms could obtain training using «debiasing» methods for fairer and more equitable results. Methods could include reweighting or better representation of data from under-represented groups, or adversarial training to identify and mitigate unfair patterns.

Being open about AI objects helps to correct AI bias. Algorithmic transparency means making the way AI makes decisions clearer and easier to understand for developers and users. «Explainable AI (XAI)» and similar techniques may reveal how a model arrives at its conclusions and assist in detecting and correcting bias before it causes any major problems. Being open also means getting teams from different fields involved in developing the process which includes international experts, sociologists, and domain experts, who can offer more ideas and anticipate more challenges [4][5].

In addition to technical fixes, we need to think about ethical accountability. Today, bias in outcomes is usually thought to be a technical failure that can be fixed. However, bias in AI ultimately reflects what society thinks. Thus, any ethical solution will also have to tackle the systemic issues underlying how data gets collected and how decisions are made. Developers, companies, and regulators all have a shared responsibility to ensure that AI technologies don't perpetuate social injustices. Regulations that require fairness assessments and ongoing audits of these systems would prevent the use of flawed algorithms.

Developers should train models more effectively and ethically, with inclusion and fairness at the core, rather than as an afterthought. The training dataset should be made as representative as possible and the models should be monitored for unintended discrimination. Besides, getting the communities impacted by AI systems to participate in the design and test processes can provide meaningful input that ensures these technologies serve everyone equally.

AI decision-making systems are increasingly being questioned for their potential biases, which can significantly affect the outcomes of cases such as criminal sentencing and hiring decisions. Training algorithms on historical data often leads to bias, which can, however, be reduced through design, transparency, and ethical responsibility. The answer is to not reiterate technical things. This acknowledges the social and historical context in which we persist. Further, it seeks to invert it by giving recommendations that are fairer. Throughout practice, there is inclusivity in effort.

The only way to ensure that AI helps people everywhere is to ensure that the underlying technology is innovative but ethical, meaning we design it to be fair.

REFERENCES

1. Holding AI Accountable: Who Gets To Tell the Story? [e-resource]: https://pulitzercenter.org/event/holding-ai-accountable-who-gets-tell-story
2. What is machine learning bias (AI bias)? [e-resource]: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-bias- algorithm-bias-or-AI-bias
3. The good, bad and ugly of bias in AI. [e-resource]: https://www.cfainstitute.org/insights/articles/good-bad-and-ugly-of-bias-in-ai
4. What is explainable AI? [e-resource]: https://www.ibm.com/topics/explainable-ai
5. Unveiling the Spectrum of Explainable AI: A Deep Dive into XAI Techniques. [e-resource]: https://medium.com/@shreeraj260405/unveiling-the-spectrum-of-explainable-ai-a-deep-dive-into-xai-techniques-1ccfa856ac96