International Journal of Advanced Research in Commerce, Management & Social Science (IJARCMSS)
ISSN : 2581-7930, Impact Factor : 5.880 , Volume 05, No. 01(II), January - March, 2022, pp 45-50

ALGORITHMIC BIAS DETECTION AND MITIGATION: ARTIFICIAL INTELLIGENCE

Mrs. Anjali Sandeep Gaikwad[*]

ABSTRACT

The private and community sectors are increasingly turning to artificial intelligence (AI) systems and machine learning algorithms to automate simple and composite decision-making processes. The mass-scale digitization of data and the up-and-coming technologies that use them are upsetting most economic sectors, including transportation, retail, advertising, and energy, and other areas. AI is also having an impact on social equality and governance as computerized systems are being deployed to improve accuracy and drive objectivity in government functions. Algorithms are harness volumes of macro- and micro-data to manipulate decisions affecting people in a range of tasks, from making movie recommendations to helping banks determine the creditworthiness of individuals. The algorithm bias is as an online recruitment tools, online ads, facial recognition technology, and criminal justice algorithms. The bias detection strategy is even when flaws in the training data are corrected; the results may still be problematic because context matters during the bias detection phase.

Keywords: *AI, Facial Recognition, Decision-Making Process, Online Recruitment Tools, Machine Learning.*

Introduction

Algorithmic bias describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others. Bias can emerge from many factors, including but not limited to the design of the algorithm or the unintended or unanticipated use or decisions relating to the way data is coded, collected, selected or used to train the algorithm.

Artificial intelligence (AI) and machine learning (ML) algorithms are generally used all through our economy in making decisions that have extensive impacts on employment, education, access to credit, and other areas. The accessibility of massive data sets has made it easy to derive new insight through computers. As a result, algorithms, which are a set of step-by-step instructions that computers follow to perform a task, have become more complicated and persistent tools for automated decision-making. While algorithms are used in many contexts, we focus on computer models that make inference from data about people, including their identity, their demographic attributes, their preferences, and their likely future behaviors, as well as the objects related to them.

For the study of algorithmic bias detection and improvement algorithms are harness volumes of macro- and micro-data to control decisions moving people in a range of tasks, from making movie recommendations to helping banks determine the creditworthiness of individuals. The common view is that an algorithm itself is usually not biased in any meaningful way (unless it is coded to be biased), but it may pick up and amplify potential biases in the input data. Machine learning algorithms are designed to find statistical corrections in the data; thus, if the input data carries social biases, the trained algorithm is likely to reflect the same biases

COMPAS (e Correctional Offender Management Profiling for Alternative Sanctions) is a rest of sound evidence of: - Predictive utility, Construct Validity, Reliability.

[*] Bharati Vidyapeeth (Deemed to be University), Institute of Management, Kolhapur, Pune, India.

Types of Algorithmic Governance for Evaluation

The key idea behind algorithmic governance is that digital technologies structure our society in multiple ways. Simplicity matters a lot as it is a start pillar of government, but since the algorithms are too difficult in their structure and aren't easy to understand, so it's not the main basic issue. There are many factors that need to be looked upon to comment on whether algorithmic governance is useful. Some of the controversy and concern arising out of algorithmic governance include:

Types of Algorithmic Governance	Description
Datafication and Surveillance	Tons of data are being generate in almost every sector in the present times, and processing this awesomely large set of data is becoming harder for scientists across the world. This situation is known as 'data deluge' and leaves us with no option with the existing technologies to process this entire set of data. Hence the observation of this data is not easily possible, and it is mostly left under-processed.
Agency and Autonomy	Algorithms are all-over today in our day to day lives, and this has led to questions of its probable impact on human autonomy and agency. Humans can act in accordance with objective morality, but machines and programs can't. There is also a fear of AI takeover, with machines and computers taking away control of the planet from humans and becoming the most intelligent species on Earth.
Transparency and Opacity	Data clearness becomes important when automated decision-making systems are questioned about their fairness of the decisions affecting an individual. But too much ease can backfire for the companies, as modern AI is making other factors apart from the source code more vital, which includes their training data sets. These machine learning algorithms - especially deep learning methods, only have code of about a few hundred lines. Even with this simplicity, one would have to look up this massive data and understand its algorithms. This would mean that responsible companies would not pause to have transparency; rather they could educate their users of the key factors driving their algorithm's decision.
De-Politicization and Re-Politicization:	It has always been a common belief that algorithms are de-politicizing, because of their 'objectivity and truth'. But the notion of algorithmic bias defies it, not fully but to a sure point. This is however due to the common social difference and bias that gets reflected in our data sets and hence algorithms. It leaves people suspicious if these algorithms were made biased to differentiate against a set of people and may even be used for repoliticization if there are vested interests.
Bias and Fairness	The biases in algorithms reflect the bias and social inequality already prevalent in society and the algorithm is not to be blamed about it. In a system there are biases in data sets and models. Although, once these biases are reflected in algorithms during operation, it only amplifies the social unfairness and prejudice. Hence, it needs to be removed from the very basic level and we need to look for a way to end these social, gender, caste and racial biases from our society if we want to finally make our algorithms and automated decision-making process fair.

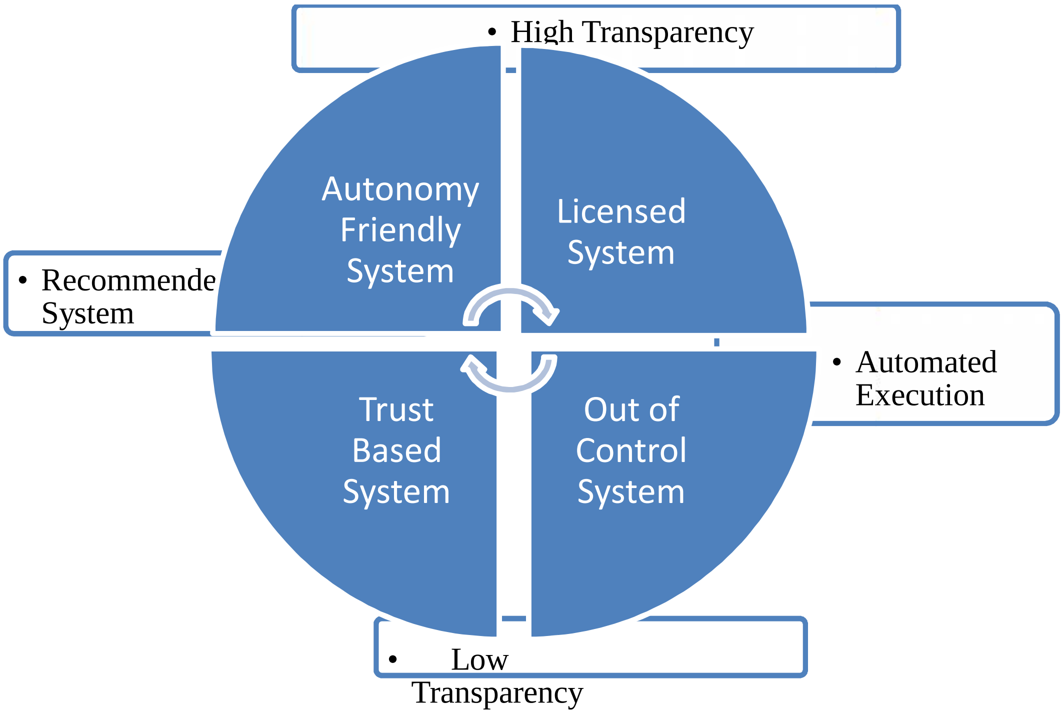

Figure: Types of Algorithmic Governance Systems

Algorithmic Impact Assessment

The automated algorithmic decision making, making use of certain parameters of evaluation structure the format of how various public systems work, like in case of the criminal justice system, criminal activity prediction, predictive police, energy usage optimization, adapted education, show evaluation systems, profile matching algorithms, etc. Often such systems operate as black boxes with little or no transparency since it generally stands out of the perspective of external scrutiny, monitoring and accountability. Some of the key elements to consider while considering the inadequacies of algorithms include:

- Assessing the impact of the automated decision systems on the factors of biasing within the community by the organizations implementing them.
- Periodic external technological audits of the processes to evaluate the methods and consider the diverse impacts.
- Transparent disclosure of the algorithmic flow and input data for the data-based decisions.
- It should also mandate a due process for beneficiaries to question an architecture based on inefficient, biased, or insufficient assessments of the technology being used.
- The direct or indirect tradeoff between productivity and decisions of the concerned entity.
- An open documentation to learn about the algorithmic data flow within a digital process.
- Study of long sets of input and output data or information to conclude an error or ineffectiveness of the core process.

Estimating the Impact of Bias

For any estimation process to conclude the impact of possible bias or error, it is important to ponder upon the following question to deeply understand and estimate the impact of underlined bias:

- Who are the beneficiary of the final decisions and how can they be precious by any biased result?
- How are the training data being chosen and what were the factors of variety being taken into consideration while collecting the data?

- Is the input data consistent and include the variety factors?
- How can an algorithm be tested?
- What are the methods to make sure the inclusion of variety in the design and implementation phase?
- What is the threshold for bias?
- What are the ways in which bias may offer any form gain to the worried stakeholders?
- What interventions may be necessary to restrict the impact of bias in any system?
- What can be the level of honesty that the design for an algorithm may have?

Framework for Evaluating Algorithmic Policy Instruments

The existence of algorithmic biases in technical products has converted cultural authorities, such as the United Nations, to establish proper legislation on algorithmic biases. Although efforts made to determine how algorithmic policies should be made have been meager, significant qualitative frameworks have been established to evaluate algorithmic policy instruments.

The first bias includes bias in input, which arises when the algorithm is fed with huge datasets. These datasets may be culturally biased themselves and hence reflect cultural inequity in the algorithm.

The second bias includes bias in the code, which perpetuates when a certain technology is constructed in such a way that it works better for a certain group of people as compared to another group. An example of this is the facial recognition technology, which is wired to identify faces of color and hence, creates a racial bias.

The third bias includes bias in the context where the input data and code which is logic to meet the purpose of the algorithm. The purpose is the real-world application which comprises a business-model and biases catering that business model pave way for bias to creep into the algorithm.

The above **3 biases in an algorithm** are the primary barriers to algorithmic policy instruments. The following includes a basic framework to remove these barriers and evaluate algorithmic policy instruments:

Bias in an Algorithm	**Description**
Barriers to unbiased inputs:	• Proprietary data: The datasets that are input in algorithms consist of information, generally hard to find to the general public and are hence confidential. This data is hence an asset for companies and can be exchange at high prices. Technical product firms don't have direct access to unbiased data and hence, the lack of data easy to get to to these firms tends to initiate a bias. • Prejudice in data: Data sets fed into the algorithm might be knowingly or unknowingly biased and providing value to certain parameters over the other. For example, in Boston, persons use the Street Bump app which uses Smartphone data to examine and mark road conditions. The improper diffusion of this app in different locations has caused the inbuilt algorithm to discriminately allocate repair resources to different locations. • Personalization: Algorithms personalize processes and results as per the users. Hence, the previous information of the user can be, sometimes, used to decide what should be shown and hence presents a single perspective to the user. • Localization: Algorithms tend to customize processes and results based on location. For example, a news algorithm may suggest trending news as per the location of an individual
Barriers to unbiased code:	• Existence of black boxes: Most software doesn't reveal its underlying logic due to security reasons and to maintain a competitive high ground. This causes consumers to be uninformed of the methodology of processes and results being offered. With the advent of sophisticated techniques like machine learning, it can be difficult for both developers and users to identify the underlying logic. • Lack of traceability: An existing bias can be eliminating if it may be traced. However, due to the presence of a plethora of sources of bias, it becomes

	extremely difficult to trace it. For example, children's YouTube recommendation system works on machine learning, the bias which can be attributed to either of the developer, the training data or malicious users. • Unpredictability of code: The presence of millions of lines of code for industry-ready software makes it prone to errors and more difficult to find bugs. Complex systems can be hard to manage, and biases may creep-in while fixing bugs. • Instability: Software firms nowadays release multiple updates in a short duration of time, owing to cloud computing and other such sophisticated technologies. The algorithm of the software is hence sensitized to a constantly changing operation of a dataset. • Lack of diversity in development teams: While drafting the code for software, the lack of perspectives among developers makes them miss out on important aspects in avoiding bias. The lack of perspective might be due to a lack of variety in the development teams. For example, the tech industry still has more male than female members
Barriers to unbiased context:	• Anticipated use and goals: While establishing the optimization technique, firms tend to anticipate the behavior of the users and hence introduce biasing to establish a perfect business model and maximize their profits • Changing media habits: Data has integrated into the media industry and is providing players in the media industry with information such as the demographics of people living in a location and their usual times of browsing. Content optimized as per this information may be biased towards a group of people. • Vertical integration: When on a platform, a firm is involved in the production and sales of its own content, the inbuilt algorithm might be conferred in such a way that it provides a bias towards the firm's own product over another firm's product. This is precisely common in E-Commerce firms selling their own products, such that pricing policy is biased towards the contents produced by the firm. • Manipulation: The most apparent way of creating bias is data manipulation. Inert users to fudge numbers and create a upsetting dataset may create a forced bias in the algorithm.

Critical Questions for Algorithmic Accountability

The above barrier to unbiasedness in algorithms can be abridged into a set of 3 question sets that provide the tool to believe an algorithmic policy as efficient.

The following are the questions:

- Is the data entered in an algorithm culturally exhaustive? The source of this data is reliable, but do the data differ significantly from other similar data sets of other cultures? Is the dataset free from prejudices?
- How effectively is the algorithm regulating the data and deriving bias-free results? Does the algorithm's logic make sense to users and developers? How effective is automation while maintaining unbiasedness? Are the individuals who established the logic culturally exhaustive?
- Are algorithms efficient in any given cultural situation? Does the algorithm behave unclearly when users change their presentation? Can the algorithm account for conflict in user expectations? Do the makers of the algorithm consider cultural unbiasedness as an important factor while developing the algorithm?
- Do teams developing machine learning datasets assess the quality and quantity of data generated and gathered to ensure populations is sufficiently and accurately represented?
- Do teams developing machine learning datasets ensure that existing datasets are not being appropriated for uses they may not be built / suited for?
- Do teams developing machine learning datasets document their provenance, creation, and use?

Recommendations

Recommendations for having a check on the algorithmic bias include the practice of up progression of numerous laws to include the digital process and practices, including the ones that derive data from the data-based processing and analysis. With the improvement of digital service delivery modules and practices by the state and the organizations, it is important to have a narrow check in place, in the form of an audit infrastructure for digital solutions.

Conclusion

This research paper focuses on equity from design to execution and incorporate technical carefulness, fairness this type algorithm should be reflected. That is, when algorithms are correctly designed, they may avoid the unlucky penalty of improved systemic bias and wrong applications.

Some decisions will be best serve by algorithms and other AI tools, while others may need thoughtful kindness before computer models are designed. Further, testing and review of certain algorithms will also identify, and, at best, mitigate biased outcomes. For operators of algorithms seeking to reduce the risk and complications of bad outcomes for consumers, the promotion and use of the mitigation proposals can create a pathway toward algorithmic fairness, even if equity is never fully realized.

References

1. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Retrieved from https://www.reuters.com/article/us-amazoncom-jobs-automation-insight/amazon-scraps-secretai-recruiting-tool-that-showed-bias-against-womenidUSKCN1MK08G.
2. Nazanin Alipourfard, Peter G Fennell, and Kristina Lerman. 2018. Can you Trust the Trend?: Discovering Simpson's Paradoxes in Social Data. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. ACM, 19–27. [4]
3. Nazanin Alipourfard, Peter G Fennell, and Kristina Lerman. 2018. Using Simpson's Paradox to Discover Interesting Patterns in Behavioral Data. In Twelfth International AAAI Conference on Web and Social Media.
4. Alexander Amini, Ava Soleimany, Wilko Schwarting, Sangeeta Bhatia, and Daniela Rus. 2019. Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure. (2019).
5. Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica 2016.

A. Asuncion and D.J. Newman. 2007. UCI Machine Learning Repository. http://www.ics.uci.edu/$\sim$mlearn/ {MLR}epository.html

6. Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. 2019. Scalable Fair Clustering. In Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97), Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, Long Beach, California, USA, 405–413. http://proceedings.mlr.press/v97/backurs19a.html
7. Alekh Agarwal, Miroslav Dudik, and Zhiwei Steven Wu. 2019. Fair regression: Quantitative definitions and reduction-based algorithms. In Proceedings of the International Conference on Machine Learning. 120–129
8. Sina Aghaei, Mohammad Javad Azizi, and Phebe Vayanos. 2019. Learning optimal and fair decision trees for non-discriminative decision-making. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 1418–1426
9. Nazanin Alipourfard, Peter G. Fennell, and Kristina Lerman. 2018. Can you trust the trend? Discovering Simpson's paradoxes in social data. In Proceedings of the 11th ACM International Conference on Web Search and Data Mining. ACM, 19–27.

○○○