Journal of Law & Social Studies (JLSS)
Volume 5, Issue 2, pp 388-396
www.advancelrf.org

Man Made Law in the Reign of Biased Machines

Muhammad Sohail Asghar
Assistant Professor,
School of Law, University of Okara, Okara.
Email: muhammad.s.asghar@uo.edu.pk

Syeda Mina Faisal
Assistant Professor,
Faculty of Law, The University of Lahore, Lahore.
Email: mina.faisal@law.uol.edu.pk

Hafsa Naz

(Corresponding Author)
Visiting Lecturer,
University of Okara, Okara.
Email: hnhafsa777@gmail.com

Abstract

Undoubtedly, the automated decision-making systems have substantially transformed our society and the recent decades have witnessed a tremendous expansion in their use. Though these algorithmic systems are getting more popular with every passing day, their intended object is not always accomplished with accuracy due to multiple factors. Consequently, inaccurate i.e., biased algorithms are produced affecting drastically people's real lives by significantly violating their constitutional rights. This paper studies that how algorithms support biased decisions and access such actions from legal, ethical, and technological perspective.

Key Words: Constitutional Rights, Inequality, Algorithmic Bias, Artificial Intelligence (AI)

Introduction

Times have changed, humans are not the only ones now making high stake decisions. The advent and excessive use of AI has opened up a new era in decision making, automated decision-making systems are being deployed in many high-stake domains such as university admissions and employment (Garcia, 2016). Though AI technology is heavily founded on algorithms to make automated decisions but humans still provide the algorithmic foundation (Heilweil, 2020). Consciously or unconsciously our gender and racial biases are trickled into algorithms. Consequently, affecting real lives of a considerate part of the society eventually undermining their constitutional rights by allowing absolute discrimination (Jackson, 2018). This unpleasant practice is defined as algorithmic bias (Heilweil, 2020).

The concept of algorithmic bias is not new, in 1988 St. George's Medical School, United Kingdom was declared guilty of racial and sexual discrimination in its admissions process as the computer program that they used for the initial screening of the admission applications made discriminatory

DOI: 10.52279/jlss.05.02.388396

decisions (Garcia, 2016). In 2012, Xerox contracted with Evolv Incorporated (Walker, 2012) to develop a machine learning system capable to make recommendations on the hiring of its new employees. In order to avoid any kind of racial discrimination, system was not provided information regarding racial identity of the candidates. It was discovered soon that the system was still making biased recommendations by rejecting many applicants belonging to certain ethnicities. Being an Artificial Intelligent (AI) system, it was able to access the ethnicity of the candidates by many proxies of race, such as zip codes and the employees' distance from his home to the office etc. Therefore, such information was also withdrawn from the system. In 2015, photos of two African Americans were tagged as 'gorillas' by Google Photos (Barr, 2015). Afterwards, Google censored this word in their photo tagging tool as a solution for this racist issue (Simonite, 2018). In 2017, a no-touch soap dispenser was introduced which only responded to white hands, it failed to recognize the hands with darker skin tone (Hale, 2017). It all happened as the algorithms used to create the dispenser were poorly trained (Noble, 2016). In 2018, it was discovered that Amazon's recruitment engine did not "like" women. After analyzing the sample data, the engine trained itself that preference should always be given to male candidates, consequently, Amazon had to shut down their recruitment engine during the recruitment procedure (Dastin, 2018). These are only a few examples of algorithmic bias, AI and machine learning abuses pertaining to human rights are much serious than our comprehension at the moment.

As the machine learning systems are trained on human produced data. We, humans as individuals and as society as a whole, are biased. Putting it in another way, there is nothing as such 'unbiased' data, or at least it is very rare to find. With reference to machine learning, huge literature is available demonstrating that algorithms trained on biased data will definitely make biased decisions (Bolukbasi, 2016). If due diligence is not exercised while developing or deploying advanced AI systems, chances are they will not only perpetuate and entrench existing biases rather will create new kinds of biases as well. Though substantial consideration given to people's repulsion regarding machines, people usually appear to prefer uncritically machine-made decisions over human made pronouncements (Skitka, 1999). Two common ways in which algorithmic bias can occur are human algorithmic bias and through over or under representation of data (Heilweil, 2020). Human algorithmic bias focuses on the collection of sample data by individuals and later on use of the same data sets for machine learning. This kind of bias relies on the idea that the data engineers or scientists responsible for collection and training of data have their own biases. These inherent biases of the individuals are reflected in the decisions of an automated decision-making mechanism. The biases originate from the over or underrepresentation of data have the same consequences but occur in a different way. These kinds of biases do not rely on the inherent biases of the individual responsible for data collection, rather on the qualitative and quantitative qualities of the data concerned. If the data scientist fails to properly represent some specific gender or race in the collected data, the chances are that the system shall mistreat or disregard that gender or race in its performance (Alake, 2020).

Right Based Issues Developed from the Emergence of AI

Historically, laws have been designed with reference to human decision makers. However, digital decision-making mechanisms are quite different from the process by which human make decisions. This difference creates incompatibility between the law and the decision makers leading to counterproductive results. This mismatch is exceptionally striking now as the intelligent algorithms have caused some serious legal ramifications. Hereunder, we are going to discuss technology and the issues in order to highlight the gap subsisting between the two. Positive efforts must be made to bridge the gap so that the automated decision-making mechanisms could be developed in a rights-respecting way.

Technology Related Issues

A few fundamental issues caused by the deployment of automated decision-making systems are transparency, track ability and interpretability. Though human decision-making process carry the same limitations, the auxiliary vulnerability is that no one exactly knows that how the decisions are being made. The same is discussed as "transparency box problem" by multiple researchers (Barocas & Nissenbaum, 2014).

While dealing with the AI, only inputs and outputs are known, what AI exactly does at algorithmic level remains a mystery (Pasquale, 2016) which raises serious concerns as to transparency. As the automated decision-making systems rely on large number of datasets representing hundreds of different factors, even the creators of the intelligent algorithms are unaware that what data did it use to reach a decision (Andersen, 2018). All these facts make it almost impossible to explain that why an algorithm resulted in a discriminatory way. Dissecting the algorithms is not the solution to this problem as it would ignite issues of intellectual property (Stevenson, 2018). Consequently, the precise algorithmic mechanics remain secret making mitigation of unfair bias extremely problematic.

Nevertheless, if we continue with the biased algorithm which is extremely discriminatory in nature, the discrimination shall be much caustic as compared to human decision making. Though it is a common believe that human decision making is more elusive than automated decision making (Miller, 2018), algorithms have the potential to discriminate more methodically, constantly and at much greater level than conventional discriminatory practices (Abungu, 2022). AI is not only capable to reproduce prevalent biases rather can drastically discriminate in unexpected ways (Fjeld & Achten, 2020). Human decision making involves only a few factors and decisions but digital decision making is the embodiment of countless different factors derived from huge datasets, making it more complex and systematic leading towards such situation where discrimination could potentially be more widespread (Abungu, 2022). Hence, considering the sophistication of AI, the threats that this technology pose are quite different from the challenges that we confront with other modern technologies. The fundamental difference is of autonomy, we are compromising our independence for something that we don't even completely comprehend. We were able to understand our previous technological inventions and were sufficiently aware of their negative effects but with the intelligent algorithms, the creators of this technology are not even close to fully understand it, what to say about its legal and socio legal implications.

The Legal Issues

As evident from the technology-based issues, intelligent algorithms are unpredictable. These are unexplainable by their creators and are hidden from those about whom they make decisions. It's an alarming situation, if these issues are left unattended, soon enough we will find ourselves in a situation where the citizens will find it almost impossible to question or complaint regarding something that can't be seen and the developers of this technology shall be unable to defend accessing people by using some algorithm that they even themselves cannot give any concrete explanation for (O'Neil, 2016). Such situation shall create a distinctive hurdle regarding the rule of law and the application of constitutional guarantees regarding equality of citizens. The anti-discrimination law, as it is applicable in the most of the countries worldwide, puts the burden of proof on the plaintiff/claimant to prove in the court of law the interrelation subsisting between the rights protected under the law of the land and the alleged biased behaviour. The combined issues regarding lack of transparency, track ability, interpretability and intellectual property make the situation worse for the litigants to perceive that they have been discriminated by the system, to prove this allegation in the court of law is even a more difficult task(Abungu, 2022). The problem related to the detection of bias is further augmented due to the reason that the discipline of Data Science and Artificial Intelligence are not about the individual rather about the datasets representing many individuals.

In supplement to the above, certain other legal issues exist from Pakistan's perspective. Anti-discrimination law is not yet introduced in Pakistan, though "equality" is protected under Art. 25 of the Constitution of Islamic Republic of Pakistan, 1973. This Article provides the protection against direct discrimination and is not so much effective against indirect discrimination. The idea of indirect discrimination is widely developed through international human rights regime. Pakistan, being signatory of many international human rights instruments, recognizes the concept of indirect discrimination to some extent, though there is no specific legal provision in the civil law of Pakistan that recognizes it or which could be used to validate a claim when an individual is victimized by this kind of biasness. This under-development of the legal concepts in the national legal framework of Pakistan is quite troubling as the deployment of intelligent algorithms for decision making comes with the high probability of indirect discrimination by way of proxy. Thus, amendment in the prevailing interpretation of discrimination is a dire need of the hour.

Strategies to Mitigate Algorithmic Bias

Mitigation of algorithmic bias requires a blend of technical, procedural, and policy driven measures at various stages of algorithms development, deployment, and regulation. After reviewing the extensive literature available on the issue, a few strategies are proposed hereunder to mitigate algorithmic bias:

Fairness and Equality

This is the fundamental principle to ensure algorithmic fairness and eliminate discrimination. This could be achieved by ensuring that the intelligent algorithms do not strengthen the existing social biases that could result into discrimination. This principle puts AI developers under a soft obligation to find out ways in which automated decision-making systems could be controlled from making the biased decisions (OECD, 2019). A number of strategies are available in governmental and non-governmental documents that could be adopted to achieve fairness and non-discrimination while using AI, such as, putting legal safeguard on the use and collection of sensitive data through data privacy and protection legislation (AccessNow & Amnesty, 2018). This shall not only lead to the development of more trustworthy intelligent algorithms, shall also strengthen the right to privacy (EuropianCommission, 2019). As we already have discussed that sample data used for the training of algorithms could not only be biased at the data collection phase but during the programming phase as well when the people responsible for the development of the algorithms bring in their own biases while developing the algorithms (AccessNow & Amnesty, 2018). This bias in the data modeling phase could be minimized by hiring people from more diverse backgrounds. Placement of oversight mechanisms is another solution to this problem. Such mechanism should be capable to assess the purpose of AI, its requirements, boundaries and decisions in a transparent manner.

Transparency and Traceability

Automated decision-making system has the potential to affect a person's life and reputation. There is a fair possibility that he could have been treated unfairly by the system. Such decision making must be transparent, meaning thereby, the decision making must be understandable to the creators of the system as well as to those who have been affected by such decisions. Efforts should be made to foster traceability of decision-making system which is essential in case the decision is opposed or challenged. AI need to be developed in a way that, if needed, its developers could explain not only the inputs and outputs but also the parameters & factors that formed the final decision. The one way to achieve transparency is that every step that the AI has taken in its decision making should be documented to efficiently pinpoint any flaws in the decision-making process. Moreover, in order to enable people to contest AI made decisions, it is of paramount importance that they must be informed beforehand that they are interacting with an intelligent algorithmic system rather than a human being. Appropriate tools should also be provided to humans for the effective interaction with AI enabling them to access the outcomes on their own and challenge the same more effectively. The creation of such mechanism shall not only enable the right as to information but also the right to appeal and redressal. Therefore, transparency is the pre-condition for the efficacious attainment of accountability.

Accountability

Accountability is important while developing responsible AI technology which does not adversely affect human beings. To this end, we have to held accountable only human beings for the decisions made by the intelligent algorithms. The threat of liability would put the developers of AI under burden to develop such intelligent algorithms that do not treat human beings in a discriminatory way. Though, the existing literature did not discuss this accountability process through judicial mechanism but non judicial means are discussed which include continuous human supervision for identification, assessment, documentation and reduction of AI related negative impacts. It carries importance specially to those who have been adversely affected by the AI. Where any adverse effects are detected, the potential victims must be reported. The current debate on the reliable intelligent algorithms has exhibited the importance of human rights impact assessment.

Human Centric

Current discourse on responsible AI has ignited the need that AI must respect and endorse human values and freedoms, rule of law, social justice and democracy. This theme is central while developing socially viable AI technology. In order to create more human centric AI, increased human involvement is required throughout its lifecycle to ensure absolute human control over it. Governance approaches such as human on the loop, human in the loop, and human in command could be implemented in order to ensure human control. How and to what extent these approaches need to be implemented is a task for the AI developers or experts, but one thing that is central to all these approaches and which cannot be compromised on is that the human beings shall always be the final decision makers (EuropianCommission, 2019).

Diversity and Inclusion

One important way to overcome algorithmic bias is to increase diversity and inclusion in the AI system. Societal diversity must be mirrored in the decision of AI. It is an attainable object by signing the people from different social backgrounds and ethnical origins. Inclusion and active contribution of the marginalized groups throughout design, development and deployment phase could ensure a rights-respecting AI system (EuropianCommission, 2019). As the algorithmic bias is an essentially socio technological problem (Microsoft, 2023), we need to deal this issue from its very root. It is a fundamental factor in order to avoid inequality and discrimination (AccessNow & Amnesty, 2018).

Private Technology Companies

Due to the non-existence of AI framework at national as well as international level, the large technology companies have formulated their own regulations and code of ethics regarding intelligent algorithms. These are essentially self-governing policies, for the time being supported by the European Commission as well (COM, 2018).Though themes of all these companies are almost similar but what a specific theme exactly means depends on the values of the company. However, broadly speaking, all working for the attainment of the same goal i.e., responsible AI. The comprehensive review of AI policies of tech giants such as Google (Google, 2023), Microsoft (Microsoft, 2023), IBM (IBM, 2023) and Ten cent (Tencent, 2023) regarding algorithmic bias and discrimination were vitally found similar.

Legal Efforts to Mitigate Algorithmic Bias and Regulate Artificial Intelligence

European Union's GDPR

General Data Protection Regulation (GDPR) is the European Union's most decisive legislative effort to regulate automated decision making and to protect user's personal data across EU. It is very detailed and formal legislation which treats data privacy and protection as human rights. Article 22 of the Regulation specifically deals with the automated decision making. The core underlying principle of this Article is that no one could be forced to be the subject to a decision which depends solely on the intelligent algorithms for decision making and such decisions have the potential to affect him legally. Stating differently, Article 22 calls for the human oversight which now has also become a crucial ingredient for the development of reliable AI.

Apart from Article 22, there are some other provisions as well in GDPR that are applicable to automated decision making, for example lawfulness, fairness and transparency [Art. 5(1) (a)], accuracy [Art. 5(1) (d)], and right to object [Art.21]. Though, there is some criticism as well regarding GDPR generally and Art. 22 more specifically (EuropeanParliament, 2020) (the analysis of the criticism is outside the purview of this paper).

Artificial Intelligence Act (AIA)

This Act is the EU's recent draft legislation to regulate Artificial Intelligence (Com, 2021). This draft legislation provides a regulatory framework for the development of reliable AI. It addresses the apprehensions of European stakeholders that certain legislative gaps exist in the European Union to regulate Artificial Intelligence and it is mandatory either to amend the current laws or to legislate new ones. Different AI posed challenges such as algorithmic bias, opacity, unpredictability etc. are addressed by this legislative development. On demand of the stakeholders, the future-proof definition of AI is provided under this law and an impartial challenge-based approach has been adopted to classify the AI systems depending on the type of threat they pose. The risk classifications include:

Unacceptable Risk

Such artificial intelligence systems which pose direct threat to the livelihood of people have been recommended to declare illegal. It includes the systems that have the capacity to infringe human free will or can influence human behaviour (EuropeanCommission, 2021).

High Risk

Intelligent algorithmic systems that have been classified in this category, have to comply with the strict regulatory obligations prior to putting them in the market. The obligations include risk assessment, traceability, human oversight and risk mitigation systems etc. A few examples of high-risk AI systems include credit scoring for bank loans, CV screening software for employment, educational software used for student assessment, among others.

Limited Risk (Generative AI)

Because of their limited risk, the draft legislation proposed that such AI systems shall be subject to certain transparency requirements. For example, a human must be informed that he is interacting with a chat boot, rather than an actual human being, so that he may have the free will either to continue or step back (EuropeanCommission, 2021).

Minimal Risk

AI systems such as spam filter do not pose any threat to human right protections or safety of citizens, they are proposed not to be regulated.

On June, 2023 MEPs approved Parliament's negotiating position on the Artificial Intelligence Act. The discussion on the same will now start with European Union countries in the Council regarding final shape of the law. The goal is to reach an agreement by the end of 2023 (EuroPar, 2023)

Conclusion

To sum up, technological advancements such as automated decision making are always best known for the positive influences that they cast upon society. But at the same time, it is important to be mindful that these innovations carry their own imperfections. Though, algorithmic decision making has increased and enriched with the passage of time, the technology may only be as good as the individuals or data create it. Therefore, it is crucial to appropriately collect, train, and oversee the data in order to avoid algorithmic bias. As, if left unchecked, algorithmic bias can drastically affect people's lives, ultimately breaching their constitutional rights and freedoms by allowing utter discriminations.

References

Abungu, C. (2022) Algorithmic Decision-Making and Discrimination in Developing Countries, Case Western Reserve Journal of Law, Technology & the Internet, Vol. 13, No. 1, pp. 41-79

Alake, R. (2020) Algorithm Bias in Artificial Intelligence Needs to Be Discussed (And Addressed), https://towardsdatascience.com/algorithm-bias-in-artificial-intelligence-needs-to-be-discussed-and-addressed-8d369d675a70

Andersen, L. (2018) Human Rights in the Age of Artificial Intelligence https://www.accessnow.org/wp-content/uploads/2018/11/AI-and-Human-Rights.pdf

AccessNow and Amnesty International (2018) 'The Toronto Declaration.https://www.accessnow.org/wp-content/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

DOI: 10.52279/jlss.05.02.388396

Barr, A. (2015) Google Mistakenly Tags Black People as Gorillas, Showing Limits of Algorithms. The Wall Street Journal, Issue 1

Barocas, S. & Nissenbaum, H.(2014) 'Big Data's End Run around Anonymity and Consent', Privacy, Big Data, and the Public Good Frameworks for Engagement, Cambridge University Press.

Bolukbasi, T. et al., (2016) Man Is to Computer Programmer as Woman Is to Home-maker? https://www.researchgate.net/publication/305615978_Man_is_to_Computer_Programmer_as_Woman_is_to_Homemaker_Debiasing_Word_EmbeddingsCOM(2018)237 Communication From the Commission to The European Parliament, The European Council, The Council, The European Economic and Social Committee and The Committee of The Regions, 'Artificial Intelligence for Europe.'https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&rid=1

Dastin, J. (2018) Amazon scraps secret AI recruiting tool that showed bias against women https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

European Commission (2021) New Rules for AI https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683

European Commission (2019), 'Ethics Guidelines for Trustworthy AI' Independent High Level Expert Group on Artificial Intelligence, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Parliament (2020), The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence: Study. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf

EuroPar (2023) EU AI Act: first regulation on artificial intelligence (European Parliament)

https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Fjeld, J., Achten, N., et. Al., (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center Research Publication No. 2020-1, Harvard University.https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482

Garcia, M. (2016) Racist in the Machine: The disturbing Implications ofAlgorithmic Bias. World Policy Journal, Issue 33(4). pp. 111-117

Google, (2023) 'Our Principles' (Google AI) https://ai.google/principles/

Hale, T. (2017) This Viral Video of a Racist Soap Dispenser Reveals A Much, Much Bigger Problemhttps://www.iflscience.com/this-racist-soap-dispenser-reveals-why-diversity-in-tech-is-muchneeded-43318

Heilweil, R. (2020) Why algorithms can be racists and sexists, Vox, https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency

IBM, (2023) 'AI Ethics' https://www.ibm.com/artificial-intelligence/ethics

Jackson, J. R. (2018) Algorithmic Bias. Journal of Leadership, Accountability& Ethics, Issue 15(4)

Miller, A. P. (2018) Want Less-Biased Decisions? Use Algorithms. Harvard Business Review https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms?utm_source=linkedin&utm_medium=social&utm_campaign=hbr

Monahan, J.&Skeem, J. L., (2016) Risk Assessment in Criminal Sentencing,12 Ann. Rev. Clinical Psychol.489

Microsoft (2023) 'Responsible AI Principles from Microsoft'https://www.microsoft.com/en-us/ai/responsible-ai

Noble, S. (2016), Challenging the Algorithms of Oppression, The Florida State University Libraries, https://guides.lib.fsu.edu/algorithm

O'Neil, C. (2016) 'Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy', Penguin Books.

OECD, (2019) 'Recommendation of the Council on Artificial Intelligence', (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Pasquale, F. (2016) 'The BlackBoxSociety', Harvard University Press.

Simonite, T. (2018) When it Comes to Gorillas, Google Photos RemainsBlind. Wired,https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

Skitka, L. J., et al., (1999) Does Automation Bias Decisionmaking?Int. J. Human-Computer Studies, 51, pp. 991-1006

Stevenson, M. (2018)Assessing Risk Assessment in Action, Minnesota Law Review (103:303) pp. 304-384

Tencent, (2023) "ARCC": An Ethical Framework for Artificial Intelligence https://www.tisi.org/13747

Walker, J. (2012) Meet the New Boss: Big Data, Wall St. J. (Sept. 20),https://www.wsj.com/articles/SB10000872396390443890304578006252019616768.