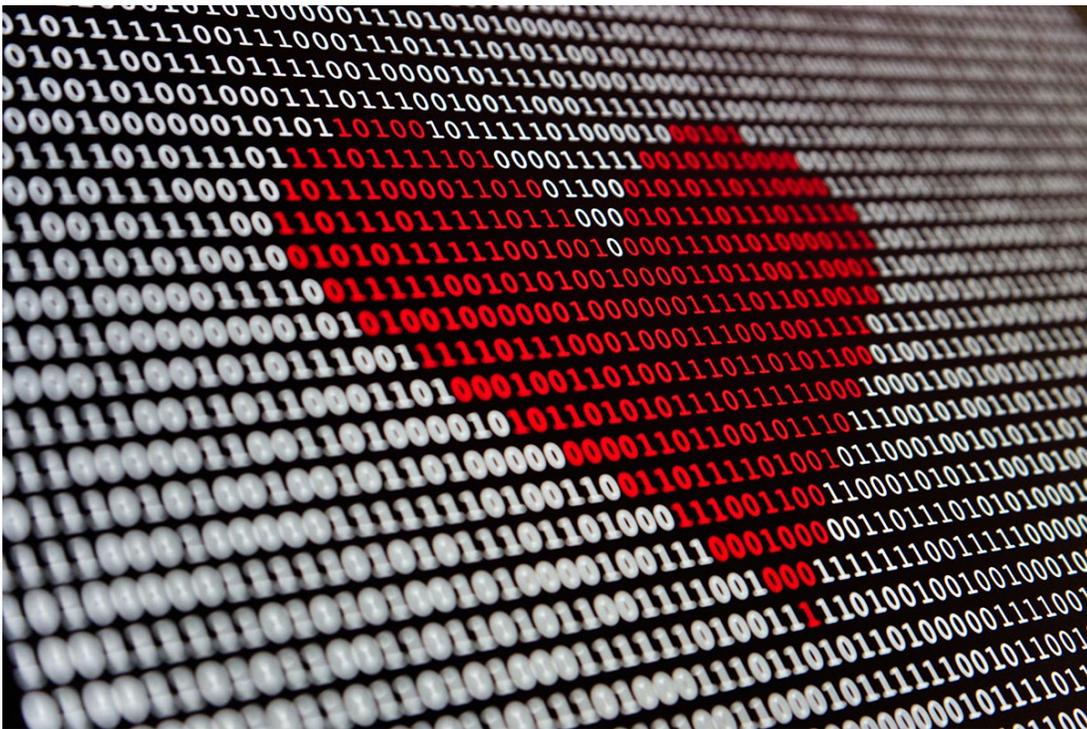


TECHNOLOGY

Algorithmic Bias: Why Bother?

by Damini Gupta and T S Krishnan



With the advent of AI, the impact of bias in algorithmic decisions will spread on an even wider scale.

✓ **INSIGHT** | FRONTIER 17 Nov 2020

The hue and cry raised about bias in algorithmic decisions does not mean humans never made biased decisions. If the biases always existed, then why bother now? Why should one care about bias in AI decision-making? Because earlier the impact of the biased decisions made by humans was localized and geographically confined. With the advent of AI, **the**

impact of such decisions is spread on a much wider scale. The very concept of geographical boundaries is breached when we speak about an AI algorithm used to take critical decisions, the algorithms that hosted on the World Wide Web and are accessed by many. For example, if a single judge was racist his decision impacted only a few unfortunate individuals appearing in his court. On the other hand, cyber courts powered by biased AI judges adversely impacts everyone across the country. Thus, the biases in AI models have resulted in much larger impact, adversely affecting far larger groups of consumers and potential employees. Few of these examples are listed below and range from gender to racial bias.

- **Colored, Latino consumers are charged more vis-à-vis White consumers to purchase and refinance mortgages**
- **Corporate hiring/recruitment biased in favor of men**
- **Prison sentencing risk assessment biased against African Americans**
- **Cameras with facial recognition identifying eyes of Asians as 'closed'**
- **Emotions (in facial recognition systems) are interpreted differently based on the person's race**

Historically, human decision-making has been biased. In a study done about two decades ago, **exactly same descriptions of a female (Heidi Roizen) and male (Howard Roizen) managers were circulated to B-school students.** For exhibiting the same behavior, Heidi was judged as aggressive & selfish whereas, Howard was judged as assertive & likeable. In summary, the impression was they did not like Heidi, did not want to hire Heidi, and did not want to work with Heidi. Very different impressions were created on the same data with only a single difference of gender.

In another study, **exactly same resumes were circulated among potential recruiters.** The only difference, half of the resumes had white-sounding names (Emily, Greg, etc.) and the other black-sounding names (Lakisha, Jamal, etc.). Black-sounding names received 50% less job interview calls vis-à-vis white-sounding names.

In the studies described above, it took a great deal of time and money (and stealth) to identify these human biases. **Identifying and fixing biases in AI algorithms is relatively easy vis-à-vis human decision-making** i.e. it is far simpler to identify bias in AI decisions

and fix it than trying to make people unlearn behaviors learnt over generations. Rectifying unconscious behavior is far more costly and time consuming than rectifying algorithmic bias.

Though “algorithmic bias” is the popular term, the foundation of such bias is not in algorithms. It is in data. Algorithms are not biased, data is! Algorithms learn the persistent patterns that are present in the training data. Multiple attributes of training data may make an AI algorithm biased. First, is due to bias present in the underlying data (decisions) used to train the AI algorithm. For example, if a judicial system is trained on historical judgements that are more unfavorable to Hispanics or Blacks, it will replicate the same and award harsher punishment to Hispanics and Blacks.

Second, if a certain demographic (or decision) is under-represented in the training dataset, it will result in biased AI algorithm. For example, if an automated speech recognition system is trained on dataset that has disproportionately fewer voice snippets of Asian women talking, it will make a lot of errors while trying to comprehend their voice.

Solving the problem of biased AI algorithms due to errors in training data is far more tractable. Use of more representative training datasets is one way to address the problem. Another, is to eliminate the features from the data set that can be used to identify gender, age, race etc. Third, is to modify the dataset through feature engineering that makes gender, age, race undifferentiated. Fourth, is to mask the features related to gender and race in the dataset by random shuffle. Fifth, is data augmentation.

AI offers an historic opportunity to identify and fix biases at an unimaginable scale. This has encouraged social activists and scholars to raise their voices against biases in algorithmic decision making. These social movements often spearheaded by scholars and backed by in-depth research, have not only created wider public awareness, but have also played an active role in shaping legislation.

Data for Black Lives, founded by Yeshimabeit Milner, aims to use data science to create positive change in the lives of Black people using open data. The organization focuses on use of open data for policy and advocacy campaigns, shift media narratives, and build real

political and economic power. In two years, Data for Black Lives has raised over \$2 million, and has changed the conversation around big data & technology across the US and globally.

A recent documentary film, Coded Bias, illustrates **the racial and gender bias prevalent in commercial facial recognition systems**. This documentary is based on the research done by Joy Buolamwini, researcher at MIT Media Lab and her journey that led to federal legislation being drafted in the US that addresses bias in AI algorithms. She is also the founder of **Algorithmic Justice League, an organization that aims to raise public awareness about the impacts of AI**, equip advocates with empirical research to bolster campaigns, build the voice and choice of most impacted communities, and galvanize researchers, policymakers, and industry practitioners to mitigate AI bias and harms.

She has a supporter in Mutale Nkonde, founder of **AI for the People**, an organization that seeks to use popular culture to educate Black audiences about the social justice implications of the use of AI Technologies in public life. Nkonde was part of the team that introduced **the bill for Algorithmic Accountability Act**, No Barriers to Biometric Barriers Act & Deep Fakes Accountability Act to the US House of Representatives in 2019.

The **AI Now Institute**, a research center at New York University, studies the social dimensions of AI and publishes independent research that influences the formation of Government legislations. Based on its research, **the institute has been providing testimonies to Government departments** and submitting amicus curiae briefs in courts calling out biases in predictive policing, facial recognition, jail sentencing, and so on.

Another organization, **AlgorithmWatch**, analyses the effects of algorithmic decision-making processes on human behavior and points out ethical conflicts. They also maintain AI Ethics Guidelines Global Inventory that lists all ethical guidelines being developed across the world. ¹

In response to social activism, legislations are being passed across the globe, mandating algorithms to be transparent, fair, and accountable. While formulating these legislations, Governments are taking into cognizance work done by these social organizations.

Such legislations are being embedded within the broad privacy legislations (Right to explanation in GDPR) or passed as stand-alone ones (Algorithmic Accountability Act). There are also cases where existing legislations have been used in the context of AI (Equal Credit Opportunity Act). These legislations are briefed below to give a fine-grained view of how regulation is applicable to AI algorithms:

- The Equal Credit Opportunity Act (ECOA) is a US Federal Law enacted in 1974 to ensure access to credit without discriminating based on race, color, religion, national origin, sex, marital status, age, or because a person receives public assistance. Using this Act, **the FTC can challenge an AI algorithm that makes credit decisions based on consumers' zip codes**, leading to discrimination of ethnic groups.
- The European Union's **General Data Protection Regulation** enacted in 2016, **provides consumers the right to obtain an explanation of the decision and to challenge the decision**. The right to explanations is applicable irrespective of whether the decisions are made by a human being or an AI algorithm.
- Algorithmic Accountability Act is a 2019 bill that **requires companies to assess their automated decision systems** (i.e. AI algorithms), including training data, for impacts on accuracy, fairness, bias, discrimination, privacy, and security.

Discussions on ethical AI is not limited to the developed countries but has also begun in developing countries like India as well. For example, Capgemini Research Institute's recent survey reveals that **more than 80% Indian companies have faced ethical issues from the use of AI systems**. As a first move, Tamil Nadu (a State in southern India) Government is in the process of **formulating a legislation for ethical use of AI**. In 2019, Indian Institute of Technology Madras organized an industry-academia AI colloquium with a session dedicated to AI and Ethics, Fairness and Explainability.

For businesses, the risks of using biased AI algorithms are regulatory and reputational. **Facebook was ordered in January 2020 to pay \$550 million** to settle a class-action lawsuit over its unlawful use of facial recognition technology. Clearview AI, another seller of facial recognition system to private companies and law enforcement agencies **has been issued numerous cease and desist orders** and is at the center of a number of privacy lawsuits. Some companies like Amazon, IBM and Microsoft are acting preemptively; significantly limiting the use of facial recognition algorithms to the extent that IBM would

no longer offer general purpose facial recognition or analysis software. Examples of reputational losses for tech giants from biased AI are starting to emerge. Amazon employees wrote a letter to CEO Jeff Bezos expressing their concerns when in 2018 its facial recognition system was found to be biased. ² In its 2018 10-K filing, **Microsoft acknowledged that use of AI in its offerings may result in reputational harm or liability**: “AI algorithms may be flawed. Datasets may be insufficient....These deficiencies could undermine the decisions, predictions, or analysis AI applications produce, subjecting us to competitive harm, legal liability, and brand or reputational harm.”

The efforts of social activists to push for legislations that govern AI algorithms, has striking similarity to the American Civil Rights Movement which demanded equality by law for all Americans. David Aberle, an American anthropologist, has coined such an effort as a radical **social movement** dedicated to fundamentally changing the societal value systems.

However, the wind of change is blowing! The wind of change is blowing from social activists who have taken up the issues of gender, racial and socio-economic biases of AI algorithms and are pushing Governments to develop legislations that govern these algorithms. Their efforts are also increasing the cost of reputational losses for organizations by creating public awareness on the impact of biased AI algorithms.

► References

1. This inventory is developed and maintained by AlgorithmWatch, a non-profit organization doing research and advocacy by analyzing algorithmic decision-making process and pointing out ethical conflicts. ↩
2. It incorrectly matched 28 members of US congress to faces picked from public mugshots. ↩



Damini Gupta ([Follow](#))

Damini Gupta, Ph.D. is the Associate Vice President and Lead (AI and Fintech) at Mphasis NEXT Labs in Bangalore, India. She heads the Ethical AI initiative at the labs and has a US patent for the AI solution that aggregates and analyzes data across multiple data sources. She did her MBA from IIM Calcutta and earned her Ph.D. from IIM Bangalore. Her Ph.D. was in the topic of sustainable business practices – specifically focusing on the effect of firms’ toxic waste generation on its valuation, financial performance, and risk.



T S Krishnan ([Follow](#))

T S Krishnan, Ph.D. is Senior Manager at Mphasis NEXT Labs in Bangalore, India. He earned his Ph.D. from IIM Bangalore in Production and Operations Management.