**Unveiling the Social and Ethical Implications of Data Bias**

Daniiar Esenbaev, Madhawa Dissanayake, Sravani Thamatam

George Herbert Walker School of Business & Technology, Webster University

CSDA 5130: Social and Ethical issues in data analytics

Professor Will J Rivers

May 18, 2023

# Introduction:

“Data is known as the new oil” (Humby, 2006, as cited in Andrade, 2022). This statement carries a lot of weight as we all know from our experiences in the information age that is called data is power. Similarly, raw data is nearly unusable without refinement and analysis. But in these cases, data biases are formed, and these may have implications in the social and ethical significance.

Data bias appears when some groups or outcomes are permanently preferred over others in the data (Mittelstadt et al., 2016). The bias can be introduced at any stage in the data process, from data collection to analysis or interpretation. Therefore, we must always care about the adverse consequences of the data, because the implications of data bias might lead to unequal treatment, injustice, and even harm to individuals and marginalized communities. Furthermore, this can enhance stereotypes and discrimination, leading to unfair treatment in multitudes of fields such as law enforcement, healthcare, and even employment (Liddell & O'Flaherty, 2018). This paper's goal is to evaluate and compile the most recent findings on the moral and societal effects of data bias. Firstly, it examines how data bias affects a lot of fields and introduces major subjects related to it. Then, we make an effort to investigate all of its facets, including its moral ramifications.

However, if we use bias in a broader meaning, it could also be referring to data that does not include characteristics that accurately represent the variable or factors which we wish to anticipate. Data also includes information created by humans that may be biased against certain groups of people (Prabhakar Krishnamurthy, 2019). An example for this would be an AI filter, which was designed and trained only for people from 5 feet tall and taller. So if we consider a person who is shorter than the trained height of 5 feet the AI filter may not work as the algorithm is developed to as the data used to train the algorithm only runs until 5 feet (Prabhakar Krishnamurthy, 2019). So this is an example where the humans have only trained the dataset for the group mentioned and thus is biased to that group of people. According to this criterion, most organically created datasets are biased, with the exception of data obtained via properly designed randomized studies. If the studies from which the data taken to train and develop the model are incomplete and have missing values, this will result in the training dataset and the algorithm developed to be biased and does not depict the data accurately (Prabhakar Krishnamurthy, 2019).

By the rise of technology, the use of data plays started playing crucial role in many societal sectors such as healthcare, education, and finance. Following this, the field of data science is growing swiftly and becoming a core of decision-making processes. That’s why, data is rapidly becoming an essential component of our daily life. However, concerns about the social and ethical implications of data bias gradually becoming an acute problem, especially with artificial intelligence and machine learning technologies (Ntoutsi et al., 2019). For instance, according to reports, digital photo technology has lately been found to use racist algorithms. Black people were purportedly labeled as "animals" or "apes" in May 2015 by Flickr's picture recognition algorithm (Yapo et al., 2018). Apart from that, web camera software from HP had trouble distinguishing dark skin tones, and camera software from Nikon mistook Asian persons for blinking (Yapo et al., 2018).

Data bias has been and remains an indecisive issue in the fields of data science and artificial intelligence, which has become increasingly significant in the modern world (Mittelstadt et al., 2016). We can depict data bias as the presence of unintentional or intentional partiality in data that may affect the results and conclusions drawn from it (Mittelstadt et al., 2016). The systematic misrepresenting of groups in the data, known as data bias, has grown to be a significant issue. Constant data use in decision-making has made data bias possible, with significant societal and ethical ramifications. (See, for example, Friedman and Nissenbaum (1996); Johnson (2006; Kraemer et al. (2011); Nakamura (2013)). Data can be used in many ways in social and ethical consequences as it is not necessarily fair and objective. Data bias causes major worry because it would provide false information and sustain social inequality. It is worth noticing that understanding data bias’s implications and take appropriate mitigations help avoid unwanted consequences. As there is a rise in big data, data analytics and artificial intelligence, the need for sorted and unbiased data becomes important.

# Literature Review

## Bias in data-driven artificial intelligence systems – An introductory survey.

Biases in humans enter AI systems, which can reproduce or even increase existing inequalities or discriminations (Ntoutsi et al., 2019). Biases in data collection often lead to over- or under-representation of certain groups, perpetuating discrimination and disadvantage. Furthermore, AI algorithm that are trained on collected data with biases reproduce or even increase existing societal and ethical problems like inequality and discrimination. The article emphasizes the need to address bias in AI systems to ensure fairness and avoid discriminatory decision-making. In order to prevent biases, several methods should be used for dataset and algorithms: balanced data set, fairness constraints and regularization and using adversarial techniques to adjust the model’s output (Ntoutsi et al.,2019).

## The ethics of algorithms: Mapping the debate

Nowadays, we frequently employ algorithms to enhance human decision-making in areas that have historically been handled by machines. Or, to put it another way, algorithms are now involved in mediating social processes, commercial transactions, political choices, and influencing our perceptions, comprehensions, and interactions with our surrounds (Mittlestadt et al., 2016). Furthermore, it is impossible to separate algorithms from the values that they embody (Brey and Soraker, 2009; Wiener, 1988). Moreover, developers establish the operational parameters of algorithms, while users configure them based on their desired results. Sadly, this process privileges some values and interests over others (cf. Friedman and Nissenbaum, 1996; Johnson, 2006; Kraemer et al., 2011; Nakamura, 2013). In light of this, there is a disconnect between our knowledge of the ethical consequences of algorithm design and operation.

Operations within accepted parameters does not guarantee ethically acceptable behavior. Much algorithmic decision-making and data mining relies on inductive knowledge and correlations identified within a dataset (Mittelstadt et al., 2016). However, acting on correlations can be a doubly problematic when it is population level decision. The automation of human decision-making is frequently justified in terms that the algorithms are free from bias. (Bozdag, 2013; Naik and Bhide, 2014). According to this viewpoint, bias is an inherent feature of algorithms, and they are bound to make biased decisions. The design and operation of an algorithm are influenced by the values and intended purposes of its creator, even if a particular design is chosen for its efficiency or effectiveness. As a result, the values of the algorithm's author are embedded into its code, potentially institutionalizing those values. (Macnish, 2012: 158). Algorithms do not only produce results but also require interpretation, which involves deciding what action to take based on the algorithm's output. In the case of behavioral data, even seemingly objective correlations can be influenced by the interpreter's unconscious motivations, emotions, deliberate choices, socio-economic status, geographic location, or demographic characteristics (Hildebrandt, 2011: 376). Every algorithm which collects data or make a decision should thoroughly undergo moral principles.

## Ethical implications and accountability of algorithms

Algorithms are implemented with the hope of being more neutral (e.g., Barry-Jester et al. 2015), thereby suggesting that the decisions are better than those performed solely by individuals. Data bias can either reinforce or violate ethical principles of the decision context. The input data is a fuel for every algorithm, and every model relies either appropriate or inappropriate attributes of data set. Especially in criminal justice system, criminal sentences must be based on facts, the law, the actual crimes committed, the circumstances surrounding each individual case and the defendant’s history of criminal conduct rather than unchangeable factors that a person cannot control (Holder, 2014). Developing accountable algorithms requires identifying the principles and norms of decision making, the features appropriate for use, and the dignity and rights at stake in the situated use of the algorithm (Martin, 2018).

## Ethical implications of Bias in machine learning

The use of machine learning algorithms has brought numerous benefits, but recent research has uncovered instances of bias in these algorithms that can have deleterious consequences. Examples include gender bias in Google search results, racist algorithms in digital photo technology, and Facebook's distribution of fake news and divisive content (Yapo, 2018). To be honest, these biases appear in societal values and discriminatory practices. Therefore, it gives us signal not to create algorithm based on certain group’s opinion. By doing this, we can reach greater transparency and fairness in the development and use of machine learning algorithms (Centre for Internet and Human Rights, 2017). Then, what kinds of controls and regulations do we need to minimize AI’s potential harm to society and maximize its benefits?

## Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries

The use of social data in digital form has become a vital component of various applications and platforms, and it has drawn the attention of many researchers. Social data can provide insights into people's opinions, behaviors, and relationships (Otleanu, 2019), enabling better decision-making in fields such as public policy, healthcare, and economics. Social data opens unprecedented opportunities to answer significant questions about society, policies, and health, being recognized as one core reason behind progress in many areas of computing (e.g., crisis informatics, digital health, computational social science) (Crawford and Finn, 2015; Tufekci, 2014; Yom-Tov, 2016). However, using social data has inherent biases and inaccuracies, and it can also introduce methodological limitations, ethical concerns, and unexpected consequences. In social data analysis different steps should be taken for check in order to avoid data quality issues. As other media, social media contains misinformation and disinformation. Misinformation is false information unintentionally spread, while disinformation is false information that is deliberately spread (Stahl, 2006). Both types of false information can distort social data, sometimes in subtle ways.

## Racial bias in an algorithm used to manage the health.

The article "Dissecting racial bias in an algorithm used to manage the health of populations" by Obermeyer et al. (2019) depicts and discuesses the social and ethical implications of data bias in healthcare (Obermeyer et al., 2019). The algorithm that healthcare professionals use to identify high-risk patients and allocate resources to them is the subject of this study. The authors discovered that white patients with the same level of risk received more resources than black patients because the algorithm was biased against them.

The study has found out how crucial it is to understand the ways that data bias can affect healthcare. Biased algorithms can reinforce and amplify already-existing healthcare inequalities, resulting in worse health outcomes for disadvantaged groups. Furthermore, relying on skewed algorithms can damage patients' confidence in the healthcare system and make them decide not to seek treatment at all. (Obermeyer et al., 2019). This study by Obermeyer and the other authors have shown what factors can contribute to mainly make bias in algorithms and how the algorithms are affected from the work done to develop the algorithms. The main issue discussed in the article is highlighted in the following citation, “Black patients assigned the same level of risk by the algorithm are sicker than White patients” (Obermeyer et al., 2019) For the analysis of the study the authors have taken a sample of patients from 2013 to 2015 from a large academic hospital and who were primary care patients. “This is because labels are often measured with errors that reflect structural inequalities.” (Obermeyer et al., 2019) In the article the authors have come to the conclusion that social and ethical inequities have been formed through the structure of our society and that the algorithm prediction is based on the results which have been taken with the existing data. Hence the poor and non-white populations will be penalized because the current health conditions are not equal for all.

## Machine bias: predicting future criminals.

Judges, probation officials, and parole officers throughout the country are starting to use algorithms to assess a criminal defendant's chances of becoming a recidivist - a term that refers to offenders who re-offend (Angwin et al., 2016). In the article the authors have analyzed COMPAS, a widely used software program which is used to predict recidivism.

The article has proven the hypothesis that the algorithm used is biased towards black people and therefore gives the black offenders a higher risk of recidivism. This can be used in the legal system to penalize them even more unfairly based on just the output of the software. The examples in the article include an incident where two people have committed a crime together but the difference being that one offender is white and has prior convictions, and the other being black and have no prior convictions. When both were taken into the COMPAS software the repeat offender who is white, has been given as low risk, and the black, non-repeat offender was categorized as high risk. This shows that the software used can be extremely dangerous and can affect human lives unfairly, if bias is not removed from algorithms. Furthermore, a higher transparency in the algorithms could lead to more scrutiny and higher fairness which intern would lead to a more unbiased use of algorithms in high impact software.

## Ad privacy settings: A tale of opacity, choice, and discrimination.

The increase in data and the advancements in analytics in the world today has introduced a novel but serious concern for the privacy of the user data. The collected data through various methods such as online surveys, online searches and even device usage are sold or resold to different parties and used to predict and target the users needs and advertise for different brands (Datta et al., 2015).

The authors have discussed the bias and discrimination of online ads according to the tests and analysis performed. “We use AdFisher to find that the Ad Settings was opaque about some features of a user’s profile, that it does provide some choice on ads, and that these choices can lead to seemingly discriminatory ads” (Datta et al., 2015). According to the research, the authors tested with each gender and found that the pay for the jobs in job searches were lower for females than males. Also, the ads shown had changed when visiting sites with substance abuse but the settings remained unchanged.

In the traditional concepts the users should be given the ability to select the privacy settings and what the users want to share according to the websites. But according to the research conducted by Acquisti and Grossklags (2005) shows that the users are lacking the required information and the knowledge to chose and which details to share and the harmful implications of sharing personal details. Furthermore, the authors have highlighted that the users tend to take the easy way to accept all the conditions and share the details without thinking of the long term implications. “Even with sufficient information, are likely to trade off long-term privacy for short-term benefits” (Acquisti & Grossklags, 2005).

According to the book “Algorithms of oppression: How search engines reinforce racism” search engines are accused of emphasizing the existing views of the users and using this to drag the user deeper into the same ideas. This can be dangerous as, even if a user accesses an illegal or terrorist related website the searches would promote the similar advertisements and search results. “Run a Google search for “black girls” ―what will you find? “Big Booty” and other sexually explicit terms are likely to come up as top search terms. But, if you type in “white girls,” the results are radically different” (Noble, 2018). As shown from the abstract taken from the author the search results according to the even race can be drastically different.

With these privacy concerns growing the international community has issued multiple regulations to restrict and control the collection and usage of consumer data illegally. For example, the General Data Protection Regulation (GDPR) of the European Union has made it mandatory for corporations to seek explicit access, provided from site users before collecting and using their data for ad targeting (European Union, 2016). Similarly, the California Consumer Privacy Act (CCPA) also has made it mandatory for sites to allow users to opt out of the usage and reselling of their personal information to other companies for the data analysis (California Legislative Information, 2018).

**Artificial intelligence can be ethical?**

Artificial intelligence can only be ethical if it is designed by ethical principles in mind (Bogenrief, T., Schutte, M.,(2020)). Th authors focuses on artificial intelligence usage in financial services and the importance of ethical principles to be followed. Using AI has many advantages in financial services such as increased efficiency, accuracy and cost effectiveness. There are risks and challenges also because of using AI such as algorithmic bias and lack of transparency mind (Bogenrief, T., Schutte, M.,(2020)). After examining the issues with the usage of AI in financial services, Bogenrief and Schutte suggested a framework. That framework consists of five principles such as transparency, accountability, responsibility, trustworthiness and fairness. And given examples of how to implement each principle in the use of AI in Financial services such as credit scoring and fraud detection. Authors also make a point of involving the stakeholders in the development and implementation of AI such as customers, regulators and employees. The authors also suggest the involvement of stakeholders in this process will help to identify and solve the potential ethical issues and usage of AI will align with social values.

**Algorithmic bias in lending:**

The use of algorithms in lending decisions is becoming common these days and leading to bias against certain groups of borrowers (Lambert, T.(2020)). Lambert in his article starts by explaining the importance of algorithmic bias in lending and its benefits like increased efficiency and reduced human bias. But also focuses on the concerns about algorithmic bias and the reasons for unfair outcomes in lending decisions. The author provides a combination of various studies that examine algorithmic bias in lending, such as studies that explore bias based on demographic factors that are race, gender and age. And the studies that are focusing on other potential sources of bias such as loan types and geographic location. Lambert suggests many strategies to decrease bias such as focusing on data quality, diversity and increased transparency and implementing careful testing and validation process for algorithms.

**Predicting Recidivism:**

The use of predictive algorithms in criminal justice systems is increasing in today’s world and many people are depending on such algorithms for decision making (Dressel, J., & Farid, H.(2018)). The study analyses data from over 100,000 Broward County, Florida and compares the predictions made by a tool which is a commercial risk assessment tool, and this is used widely to get the actual recidivism outcomes. The findings say that the predictive accuracy of the tool is modest at best, with slightly better performance than random guessing. Dressel and Farid’s looked into the accuracy of commercial data sources such as criminal history, demographics and socio-economic factors for predicting recidivism. To test the effectiveness by comparing the predictions made by these algorithms with actual recidivism. The study also focuses on limitations of predictive algorithms in recidivism prediction. The authors also discussed the ethical issues of depending on predictive algorithms in criminal justice systems. The authors raise concern for not maintaining transparency in the development of these algorithms and possible for strengthening existing biases and inequalities.

**Algorithms usage on Big data result in disparate impact :**

The increased dependency on Big data and algorithmic systems has the potential to increase social inequalities (Barocas, S., & Selbst, A. D.(2016)). The authors highlights that how these systems maintain and amplify biases presented in the data they are trained on lead to discriminating outcomes. The article says that these biases are not intentional but the result of how algorithms process and understand the data. Barocas and Selbst investigated legal frameworks such as Fair Credit Reporting Act (FCRA) and Title VII of the civil rights Act which were developed to address discriminatory practices. The authors worked on how these laws can be applied to cases involving algorithmic decision making and discuss the challenges in algorithms responsible for their contrasting impact. The Authors also proposed a framework that consists of statistical methods, causality, analysis and legal standards to assess and decrease the big impact on algorithmic decision making. They argue for these techniques to be followed while design, development and deployment of algorithms to ensure honesty and avoid discrimination.

**This is how computers misunderstand the data:**

Computers have powerful capabilities but fail to truly understand this complex world because of their limitations and misconceptions (Broussard, M. (2018)). Broussard requests readers to fully examine the underlying assumptions and values placed with AI systems as they reflect the options and biases of their human creators. The author highlights the importance of human mistakes and responsibility and suggests a more ethical and responsible approach in terms of developing and deploying AI technologies. The author seeks to bridge the gap between hype and the reality of AI. Broussard focuses on the need for incorporative collaboration, including suggestions from social scientists, ethicists and policy makers to make sure that AI technologies align with human values and serve the wide interests of society. The paper provides a deep examination of AI technologies, and their limitations, biases and societal implications.

# References

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. IEEE Security & Privacy, 3(1), 26-33. doi: 10.1109/MSP.2005.22

Andrade, F. (2022, July 21). *Is Data The New Oil of the 21st Century or Just an Overrated Asset?* Medium. <https://towardsdatascience.com/is-data-the-new-oil-of-the-21st-century-or-just-an-overrated-asset-1dbb05b8ccdf>

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica.

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Barry-Jester, A. M., Casselman, B., & Goldstein, D. (2015). *The new science of sentencing. The Marshall Project*. <https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing.>

Bozdag E (2013*) Bias in algorithmic filtering and personalization*. *Ethics and Information Technology* 15(3): 209–227.

Brey, P. A., & Søraker, J. H. (2009*). Philosophy of Computing and Information Technology. In Elsevier eBooks* (pp. 1341–1407). Elsevier BV.

<https://doi.org/10.1016/b978-0-444-51667-1.50051-3.>

California Legislative Information. (2018). California Consumer Privacy Act of 2018. Retrieved from <https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375>

Crawford, K., & Finn, M. (2015). The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal*, *80*(4), 491–502. <https://doi.org/10.1007/s10708-014-9597-z>

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on AD Privacy Settings. Proceedings on Privacy Enhancing Technologies, 2015(1), 92–112. <https://doi.org/10.1515/popets-2015-0007>

European Union. (2016). General Data Protection Regulation (GDPR). Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Friedman B and Nissenbaum H (1996) *Bias in computer systems*. *ACM Transactions on Information Systems* (TOIS) 14(3): 330–347.

Hildebrandt, M. (2008). Defining Profiling: A New Type of Knowledge? In *Springer eBooks* (pp. 17–45). Springer Nature. https://doi.org/10.1007/978-1-4020-6914-7\_2

Holder, E. (2014*). Attorney General Eric Holder Speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference*. The United States Department of Justice. <https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th.>

Johnson JA (2006) *Technology and pragmatism: From value neutrality to value criticality*. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Available at: http://papers.ssrn.com/abstract=2154654 (accessed 24 August 2015).

Kraemer F, van Overveld K and Peterson M (2011) *Is there an ethics of algorithms? Ethics and Information Technology* 13(3): 251–260.

Liddell, R., & O'Flaherty, M. (2018). Handbook on European non-discrimination law. European Union Agency for Fundamental Rights (FRA).

Macnish K (2012) *Unblinking eyes: The ethics of automating surveillance*. Ethics and Information Technology 14(2): 151–167.

Martin, K. E. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835–850. <https://doi.org/10.1007/s10551-018-3921-3>

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). *The ethics of algorithms: Mapping the debate*. *Big Data & Society*, *3*(2), 205395171667967. <https://doi.org/10.1177/2053951716679679>

Naik G and Bhide SS (2014) *Will the future of knowledge work automation transform personalized medicine?* Applied & Translational Genomics, Inaugural Issue 3(3): 50–53.

Nakamura L (2013) *Cybertypes: Race, Ethnicity, and Identity on the Internet. New York: Routledge.*

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., . . . Staab, S. (2020). *Bias in data‐driven artificial intelligence systems—An introductory survey*. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, *10*(3). <https://doi.org/10.1002/widm.1356>

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. <https://doi.org/10.1126/science.aax2342>

Olteanu, A., Castillo, C. F., Diaz, F., & Kiciman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*. <https://doi.org/10.3389/fdata.2019.00013>

Prabhakar Krishnamurthy. (2019, September 12). *Understanding Data Bias. Medium; Towards Data Science*. <https://towardsdatascience.com/survey-d4f168791e57>

Stahl, B. C. (2006). *On the Difference or Equality of Information, Misinformation, and Disinformation*: A Critical Research Perspective. Informing Science the International Journal of an Emerging Transdiscipline, 9, 083–096. <https://doi.org/10.28945/473>

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *arXiv (Cornell University)*, *8*(1), 505–514. <https://doi.org/10.1609/icwsm.v8i1.14517>

Wiener N (1988) *The Human Use of Human Beings: Cybernetics and Society*. *Da Capo Press*.

Yapo, A., & Weiss, J. (2018). Ethical Implications of Bias in Machine Learning. In *Proceedings of the . . . Annual Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/hicss.2018.668>

Yom-Tov, E. (2016). *Crowdsourced Health: How What You Do on the Internet Will Improve Medicine.*

Barocas, S., & Selbst, A. D.(2016). *Big Data’s Disparate Impact. California Law Review, 104(3), 671-732.*

Dressel, J., & Farid, H.(2018). *The Accuracy, Fairness and Limitations of predicting recidivism, Science Advances, 4(1).*

Broussard, M. (2018). *Artificial intelligence: How Computers misunderstand the world. Cambridge, MA: MIT Press*