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

Group 6

George Herbert Walker School of Business & Technology, Webster University

CSDA 5130: Social and Ethical issues in data analytics

Professor Will J Rivers

May 4, 2023

# Introduction

Data is known as the new oil. This statement carries a lot of weight as we all know from our experiences in the information age 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 is formed when the data systematically favors certain groups or outcomes over others. The bias can be introduced in any stage of the data from the gathering of the data to analysis or interpretation. As such we have to always keep in mind the ramifications of the data and that the implications of data bias are far-reaching and can lead to unequal treatment, injustice, and even harm to individuals and marginalized communities. Furthermore, this can perpetuate stereotypes and discrimination, leading to unfair treatment in multitudes of fields such as law enforcement, healthcare, and even employment.

This paper aims to evaluate and synthesize the most recent research on the moral and social ramifications of data bias and introduces important topics linked to data bias and examines how it affects many disciplines and this essay tries to investigate the various facets of data bias, its effects on diverse fields, and its ethical implications. In today's world, using data is becoming more and more common and the utilization of data is becoming more crucial across a range of societal areas, including healthcare, education, and finance, and the profession of data science is expanding quickly and it is now an essential component of decision-making procedures. Data is quickly becoming an indispensable part of our daily lives in the current digital era. In recent years, concerns about the social and ethical implications of data bias have risen, especially with artificial intelligence (AI) and machine learning (ML) technologies.

Data bias has been a long-standing issue in the realm of data science and artificial intelligence (AI), which has become increasingly significant in the modern world. Data bias can be described as the presence of unintentional or intentional partiality in data that may affect the results and conclusions drawn from it. Data bias, or the systematic and unfair underrepresentation of groups in data, has grown to be a major issue, nevertheless. Data bias has, however, also emerged as a result of the widespread use of data in decision-making processes, which has important social and ethical ramifications. Data can be used in ways that have social and ethical ramifications because it is not necessarily fair and objective. Particularly data bias is a major worry because it might provide false results and sustain social inequality. In order to reduce the effects of data bias, it is crucial to comprehend their implications and establish appropriate mitigation measures.

With the growth of big data, data analytics, and artificial intelligence, the need for accurate and unbiased data is becoming increasingly important. Data bias occurs when data is collected, analyzed, and reported in a way that systematically favors certain groups over others. This can lead to unfair treatment of individuals or groups, perpetuating social and economic inequalities. This paper aims to review the current scientific literature on the social and ethical implications of data bias and provide examples of how it affects society.

In this paper, we explore the social and ethical implications of data bias, focusing on its impacts on individuals, society, and the field of data science itself. Key terms used in this paper include data bias, machine learning, artificial intelligence, fairness, accountability, and transparency.

# 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

Algorithms are increasingly being used to make better decisions that were previously left to humans, and they now mediate social processes, business transactions, governmental decisions, and how we perceive, understand, and interact with our environment (Mittelstadt et al., 2016). Algorithms are inescapably value-laden (Brey and Soraker, 2009; Wiener, 1988). Operational parameters are specified by developers and configured by users with desired outcomes in mind that privilege some values and interests over others (cf. Friedman and Nissenbaum, 1996; Johnson, 2006; Kraemer et al., 2011; Nakamura, 2013). However, there is a gap between the design and operation of algorithms and our understanding of their ethical implications.

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 often justified by an alleged lack of bias in algorithms (Bozdag, 2013; Naik and Bhide, 2014). Bias always manifests in algorithms and therefore, they inevitably make biased decisions. An algorithm’s design and functionality reflects the values of its designer and intended uses, if only to the extent that a particular design is preferred as the best or most efficient option. As a result, ‘‘the values of the author [of an algorithm], wittingly or not, are frozen into the code, effectively institutionalising those values’’ (Macnish, 2012: 158). The outputs of algorithms also require interpretation (i.e. what one should do based on what the algorithm indicates); for behavioral data, ‘objective’ correlations can come to reflect the interpreter’s ‘‘unconscious motivations, particular emotions, deliberate choices, socio-economic determinations, geographic or demographic influences’’ (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). These biases are often a reflection of societal values and discriminatory practices, and highlight the need for greater transparency and inclusivity in the development and use of machine learning algorithms**.** As such, because these algorithms are created by humans, they inevitably — and often unconsciously — reflect societal values, biases, and discriminatory practices (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) highlights 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 emphasizes how crucial it is to comprehend how data bias affects healthcare. Biased algorithms can reinforce and amplify already-existing healthcare disparities, 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). The study also highlights the importance of addressing data bias in algorithm development. The authors suggest that incorporating diverse perspectives and data sources, testing algorithms for bias, and making the algorithm transparent and accountable can help mitigate bias (Obermeyer et al., 2019).

There is an increasing amount of literature on the social and ethical implications of data bias. Scholars have highlighted the potential for algorithms to perpetuate and amplify existing social and economic inequalities, particularly for marginalized groups and furthermore, there are concerns that algorithms may be used to reinforce discriminatory practices and undermine free will of the people. Moreover, there is a growing recognition that addressing data bias requires interdisciplinary approaches that draw on diverse perspectives and expertise (Crawford et al., 2019). Scholars from a range of fields, including computer science, ethics, law, and sociology, are working to develop frameworks for addressing data bias and ensuring that algorithms are fair and equitable.

Overall, the article by Obermeyer et al. (2019) highlights the need for addressing data bias in the healthcare industry and the importance of ensuring that all industries ensure that algorithms are fair and equitable for the users and the affected parties.

## Machine bias: predicting future criminals

Angwin et al. (2016) conducted a study on COMPAS, a widely used software program in the criminal justice system and found that it demonstrated significant racial bias. The study showed that the software was more likely to mistakenly classify black defendants as higher risk than white defendants, thereby raising concerns about the fairness of the criminal justice system (Angwin et al., 2016).

This issue of algorithmic bias in predictive policing has been widely studied in the literature. Ensign et al. (2018) found that predictive policing algorithms often rely on biased data, which can perpetuate racial and socioeconomic disparities. Similarly, Noble (2018) demonstrated how the use of predictive policing software can have a chilling effect on communities of color and lead to a breakdown of trust between law enforcement and the communities they serve.

In conclusion, Angwin et al. (2016) highlight the need for greater transparency and accountability in the use of predictive policing software. The study underscores that algorithmic bias is a pressing issue that requires urgent attention to ensure a fair and just criminal justice system.

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

The study by Datta, Tschantz, and Datta (2015) investigates the impact of ad privacy settings on users' online experiences. The authors conducted an automated experiment on Facebook and found that users' ad preferences were opaque and difficult to change. They also found evidence of discrimination in ad targeting, as certain groups were shown different ads based on their perceived interests.

The literature, Acquisti and Grossklags (2005) suggests that ad targeting can perpetuate discriminatory practices, such as racial bias, by allowing advertisers to selectively target specific groups of individuals. These practices can have lasting effects on individuals, leading to the reinforcement of existing social and economic biases. Scholars have called for increased transparency and accountability in ad targeting to address these issues and ensure that individuals' privacy rights are protected. The writer has emphasizes the importance of addressing privacy and discrimination concerns in ad targeting to promote fair and equitable outcomes for all individuals.

The study by Datta et al. (2015) underscores the need for greater transparency and control over ad privacy settings. The authors suggest that platforms should provide users with clear and accessible information about how their data is being used for ad targeting. They also suggest that users should be given greater control over their ad preferences and the ability to opt out of targeted advertising altogether.

In recent years, there have been several efforts to address the issue of online privacy and discrimination in ad targeting. For instance, the European Union's General Data Protection Regulation (GDPR) requires companies to obtain explicit consent from users before collecting and using their data for ad targeting (European Union, 2016). Similarly, the California Consumer Privacy Act (CCPA) requires companies to provide users with the ability to opt out of the sale of their personal information (California Legislative Information, 2018).

In conclusion, the study by Datta et al. (2015) highlights the need for greater transparency, control, and accountability in ad privacy settings. The authors' findings underscore the potential for discrimination in ad targeting and the need to address this issue through policy and regulation. The literature suggests that there have been efforts to address these issues, but further research is needed to evaluate their effectiveness.

# Analysis

Data bias is a significant issue that affects the validity and reliability of data analysis, resulting in discriminatory outcomes that can adversely affect individuals or groups. From the literature review, I agreed with the authors that data bias is a complex issue that needs a multifaceted approach to address it. In particular, I agree with Burrell (2016) that the current emphasis on machine learning algorithms as a solution to data bias overlooks the role of social and historical contexts in shaping data and its interpretation.

However, I did not agree with some of the authors who argue that data bias is an inherent feature of the technology and cannot be eliminated. I believe that, while it is difficult to achieve complete bias-free data, it is possible to minimize the effects of bias by employing diverse teams in data collection, analysis, and interpretation.

In our opinion, one example of data bias is the use of facial recognition technology by law enforcement agencies. Studies have shown that these systems have a higher error rate for people of color and women, which can lead to false arrests and wrongful accusations. For instance, in 2020, a Black man was falsely arrested and held for 30 hours because of a faulty facial recognition algorithm used by the Detroit Police Department (Harwell & Timberg, 2020).

To address data bias, there is a need to develop policies and regulations that prioritize fairness, transparency, and accountability in data collection, analysis, and interpretation. For instance, companies and organizations should be required to disclose the data they collect and how it is used, as well as provide avenues for people to correct or challenge the data that affects them.

A 2019 study by Obermeyer and the team on the “Dissecting racial bias in an algorithm used to manage the health of populations” examined racial bias in a population health management algorithm. The algorithm used by healthcare professionals in the USA to identify high-risk patients who might benefit from additional medical support was examined in the study. When we dive into the results taken from the study, it is evident that the algorithm underestimates the black patients' healthcare needs, this is the cause we have highlighted to show that the algorithm was discovered to have a racial bias (Obermeyer et al., 2019). The study emphasizes the significance of investigating data bias in algorithms that are more frequently used to manage population health. Furthermore, this study sheds light on the possible social and ethical repercussions of data bias in population health management algorithms (Obermeyer et al., 2019). In order to identify high-risk patients, the algorithm in question used information from prior healthcare expenditures. But because of the low expenditures and the inequality of the health care system the percentage of black patients which prefer and go to hospitals have reduced and therefore the accurate values cannot be traced back to a source as the feedback loop for the cause and effect are intertwined and have brought us to this resulting situation.

In addition to this research, we would like to suggest that a controlled study be conducted on the groups in question which include the patients who are receiving treatment from the hospitals. For the controlled study we can consider the current procedures followed by the hospital and a new controlled group selected and offer equal availability and treatment for both the black and white patient study groups. With the results of this we can consider the cost for the treatment and the availability of treatment for both groups and improve on the algorithm to remove the racial biases which are discussed in the article (Obermeyer et al., 2019).

The article "Machine Bias: Risk Assessments in Criminal Sentencing" by Angwin et al. (2016) provides insight into the issue of bias in algorithmic decision-making, particularly in the context of criminal justice. The writers have looked into a popular piece of software called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), which makes the claim that it can forecast a person's propensity to commit a crime in the future. According to the study, the software has a bias against African-American defendants and tends to overstate their likelihood of recidivism when compared to defendants of other races. (Angwin et al., 2016). The paper emphasizes the consequences of such bias because it can lead to harsher punishments and maintain systemic racial injustices in the criminal justice system. As an example for the racial bias in this analysis we can look into some of the examples form the study, which include the comparison between 2 people with shoplifting arrests, a white individual with Prior Offenses such as, 1 domestic violence aggravated assault, 1 grand theft, 1 petty theft, 1 drug trafficking was classified as a low risk individual of risk level 3, and a black individual with Prior Offenses such as, 1 petty theft who was classified as medium risk with a risk level of 6. Furthermore, after the risk analysis for future crimes, the white man had escalated his crimes into 1 grand theft, whereas the black man had no new offences. Similarly, along these lines another instance was recorded where the crime was committed by both groups together the risk showed a high risk for the black person and the white person low risk. (Angwin et al., 2016). We can see that we look into these events that the issue is caused not by the data but by the underlying algorithms of the programs.

The study's ethical and societal ramifications of data bias in algorithmic decision-making are highlighted, which is important. It raises questions regarding the fairness and precision of algorithms, which are increasingly being used to make judgments that have an effect on both people and society. The study also emphasizes the necessity of human oversight and intervention to minimize and reduce bias, as well as the significance of openness and accountability in the design and implementation of algorithmic systems. Furthermore, as the developers of the said algorithms it is in our responsibility to make the algorithms fair and unbiased. With respect to my opinion on the improvements which can be done on the existing algorithm, a unbiased study has to be performed for the social groups based on race, ethnicity, region and all other concerning factors and the foundation for the algorithm has to be made transparent for the public as they are the main stakeholders of the solution. Moreover, for improving on this a feedback system with the updated data for the convicts can be taken per period of time and the predictions made by the risk analysis system can be updated with the feedback. A machine learning model with such capabilities would be self-learning and correct mistakes made. But for this to be successful a stable and unbiased foundation code for the algorithm has to be made. So I hope that as technology and analysis, predictive methods are improving such technologies would be used to improve the quality of living and reduce risk in society for all.

Acquisti and Grossklags (2005) has shown in the article that ad targeting can perpetuate discriminatory practices, including racial bias, by allowing advertisers to selectively target certain groups of individuals (Acquisti & Grossklags, 2005). So according to the author discrimination on social media based on gender and race may impact the users in drastic ways which may be carried forward and which will affect the mental and emotional state of the users. Also, in the article by Datta et al. (2015) “Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. Proceedings on Privacy Enhancing Technologies”, the authors have conducted independent studies on social media to perceive the ways that biases are incorporated into the advertisements on social media. They have conducted an automated experiment on Facebook and found that users' ad preferences were opaque and unchangeable. They also found evidence of discrimination in ad targeting, as certain groups were shown different ads based on their race, opacity and perceived interests (Datta et al., 2015).

One of the strengths of the article by Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. is its comprehensive and balanced analysis of the various perspectives on the ethics of algorithms. The authors provide a nuanced assessment of the benefits and risks associated with algorithmic decision-making and highlight the need for a cautious and critical approach to the development and use of algorithms. I totally agree with authors in various challenges posed by algorithms and the potential solution to those challenges.

The second point to acknowledge is that defining "fairness" is not a straightforward task, and cannot be achieved through a simple mathematical formula. Fairness is a dynamic and social concept, not solely a statistical issue. Furthermore, what is considered "fair" can vary greatly across different countries, cultures, and application domains. Therefore, it is crucial to have practical definitions of fairness that are appropriate for different contexts, along with datasets specific to each domain for developing and evaluating methods. It is also important to go beyond the usual training and testing evaluation approach and consider the potential outcomes of interventions aimed at promoting fairness, to ensure the long-term well-being of various groups. Finally, because perceptions of fairness change over time, there is a growing need to explore whether it is feasible to train models on historical data and apply them to current fairness-related problems.

From the book “Algorithms of oppression: How search engines reinforce racism” first off, search engines like Google favor and emphasize the prevailing viewpoints and voices in society, which frequently originate from those with privilege and power. As a result, racial hierarchy and discrimination may be strengthened. Through search results and advertising, search engines frequently support prejudices and derogatory depictions of minority groups, especially people of color. These populations may become even more marginalized as a result, which could be harmful. Last but not least, search engines can support discriminatory behaviors by limiting opportunities and information for particular groups based on characteristics like race or socioeconomic position (Noble, 2018).

According to the analysis with the data age increasing the data taken from the users the companies are using the above collected data to customize the searches, the results on the browsers and even the type of music recommended. So, in doing so if a biased algorithm is present the uses may get dangerous or even criminal search results. So, it is important to highlight that greater transparency and control over ad privacy settings and the algorithms which control the said processes and be shared with independent studies to make surfing the web safer and more effective for everyone.

# Conclusion

Data bias is a pervasive issue that requires a multifaceted approach to address it. While machine learning algorithms can help minimize the effects of bias, social and historical contexts play a crucial role in shaping data and its interpretation. Therefore, policies and regulations that prioritize fairness, transparency, and accountability are necessary to ensure that data is used ethically and responsibly.

Identifying and exposing the different ways in which biases can be present and reinforced in data-driven decision-making processes is referred to as "unveiling the social and ethical implications of data bias" in this paper. It involves looking at the intricate and frequently subtle ways that biases can be present in algorithms, models, and datasets, producing unfair, discriminatory, and detrimental results for particular groups of individuals. By bringing these challenges to light, we may seek to create more just and equitable systems that are unaffected by prejudice and discrimination.

In the paper we have discussed the issue of bias in algorithmic decision-making in different contexts. The racial bias in a population health management algorithm, which underestimates the healthcare needs of black patients due to the low healthcare expenditures, the bias in the COMPAS software used for criminal sentencing, which overstates the likelihood of recidivism for African-American defendants, how ad targeting can perpetuate discriminatory practices, including racial bias, by allowing advertisers to selectively target certain groups of individuals and influences from external factors to affect the algorithms to be inherently or by the use of erroneous data to be biased are some of the issues that we have referred to in the paper. The importance of these issues highlighted is that we can be mindful in creating algorithms that mitigate the risks and would not affect the users and the stakeholders in a way that would drastically influence their lives negatively.

The texts highlight the ethical and societal ramifications of data bias in algorithmic decision-making, which points to one important point which is the need for fairness, precision, transparency, and accountability in the design and implementation of algorithmic systems, as well as the necessity of human oversight and intervention to minimize and reduce bias.

# Research Limitations

Within the analysis we have liked about the improvements and our views on how to mitigate the errors in the existing algorithms and to provide a basis for the future machine learning algorithms to operate from. Suggestions for improving existing algorithms include conducting unbiased studies for social groups based on race, ethnicity, region, and other concerning factors and making the foundation code for the algorithm transparent to the public. Feedback systems with updated data and self-learning machine learning models are also suggested.

However, one limitation of the article by Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. is that it was published in 2016, and the field of algorithmic ethics has since evolved rapidly. In particular, the article does not address some of the more recent developments in the field, such as the growing concerns over the use of facial recognition technology and the increasing focus on algorithmic accountability and explainability.

So in our opinion, the risks of the current algorithms are shown in the analysis and we hope that the paper would be useful in paving the way to a better future where the algorithms are fair, precise and transparent.

## Future research opportunities.

People, who involved in creating AI, must consider the impact of their design choices and assumptions on potential biases. Studies have shown that biases can occur due to inadequate awareness of certain categories during development. People from privileged groups may not even be aware of the existence of categories, while those from underprivileged groups are more likely to notice their differences. To combat this bias, increasing diversity in development teams and subjecting algorithms to outside scrutiny are two promising strategies, such as allowing some forms of reverse engineering for algorithmic accountability.

# 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

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.>

Bogenrief, T., & Schutte, M. (2020). *Can artificial intelligence be ethical? A case for financial services*. Journal of Financial Perspectives, 8(2), 23-33.

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*. <https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375>

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., ... & Reddy, S. (2019). *AI Now Report 2018.* New York: AI Now Institute.

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: A tale of opacity, choice, and discrimination*. Proceedings on Privacy Enhancing Technologies, 2015(1), 92-112. doi: 10.1515/popets-2015-0006

Dressel, J., & Farid, H. (2018). *The accuracy, fairness, and limits of predicting recidivism*. Science Advances, 4(1), eaao5580.

European Union. (2016). *General Data Protection Regulation (GDPR)*.

<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.

Lambert, T. (2020). *Algorithmic bias in lending: A review of the literature*. Journal of Financial Services Research, 57(1), 1-25.

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: NYU 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.

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>

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.*