A STUDY OF ALGORITHMIC BIAS WITH A FOCUS ON MITIGATION PRACTICES AND AN ANALYSIS OF DISCRIMINATION CONSCIOUS DATA MINING

by

ASHNI WALIA

Presented to the Faculty of the Honors College of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

HONORS BACHELOR OF SCIENCE IN INFORMATION SYSTEMS

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2022

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my supervisor, Dr. Jennifer Jie Zhang for her invaluable advice, continuous support, and patience. Her immense knowledge and plentiful experience have encouraged me to successfully complete my thesis. I would like to thank the staff at the Honors College as well. It is their kind help and support that have made my years at The University of Texas at Arlington a wonderful time. Finally, I would like to express my gratitude to my parents, my sister, and my friends. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to achieve my goals.

November 19, 2021

ABSTRACT

A STUDY OF ALGORITHMIC BIAS WITH A FOCUS ON MITIGATION PRACTICES AND AN ANALYSIS OF DISCRIMINATION CONSCIOUS DATA MINING

Ashni Walia, B.S. Information Systems

The University of Texas at Arlington, 2022

Faculty Mentor: Jennifer Jie Zhang

Algorithmic bias is a moral error within computer systems that is often left undetected due to a lack of set procedures. The aim of this study was to find the source of this bias leading to possible procedural solutions that can be applied widely. A meta-analysis, case study, and sample interview statistics are used to understand the multiplication of such bias into generated outputs. The study concluded that lack of diverse data leads to bias in output, in addition to a lack of awareness about the existence of such bias. This ignorance is amplified by the myth surrounding deep learning algorithms. The study recommends government intervention to set standards for AI development and further peer-reviewed research in the context of larger societal impact in the future.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In this paper, research on algorithmic bias is presented. The paper includes the origins of this bias along with methods to mitigate the risks of various kinds of algorithmic bias. Case studies and analysis of successful data mining systems that are discrimination conscious are used to prove the feasibility and importance of such practices.

1.1 Origins of algorithmic bias

The chapter begins by exploring how a neutral algorithm can lead to user bias due to the interface. Followed by a discussion about the emergence of algorithmic bias through the inherent bias of the designer, and the negative effect of biased inputs and data, ending with a case study about algorithmic bias in health systems.

1.1.1 Users and bias

The emphasis on the origins of algorithmic bias is often placed on the designer and data sources. However, user bias has been the most problematic aspect of algorithmic bias as found in recent studies. As user interfaces become more interactive, the algorithms lead to indirect inflation of bias. According to Jason Chan et al. (2015, p. 2973), "In 2013 alone, over one million workers and nearly half a million new businesses have joined Elance, an online labor platform that generated $1 billion in worker earnings through 3.5 million posted jobs since its inception." The scale of this market leads to an equal scale of social impact. If the headhunters on these platforms have biased preferences, the website algorithm reflects that. This further amplifies the bias by only showing the hiring managers their preferred population sets. This can lead to even more inequitable access to job opportunities. This is how users can engage with an interface with subconscious gender bias.

1.1.2 Designers and bias

Artificial Intelligence (AI) is the closest thing known to have human intelligence. It uses algorithms to solve problems in a similar way the human brain would process them. AI can therefore be reactive and self-aware. The combination of smart machines capable of human reactions and artificial intelligence can be both a boon and a bane. The baseline assumption for any algorithm is GIGO- Garbage In, Garbage Out. Artificial Intelligence might be capable of using techniques where algorithms uncover or learn associations of predictive power from data however, the algorithms themselves are built by humans and are prone to human error. Any human bias can potentially lead to algorithmic bias. In the previous section, it was discussed how users can engage with an interface with subconscious gender bias. In this section, the discussion leads into how this bias translates into algorithms.

The most tangible form of AI is machine learning, which includes a family of techniques called deep learning that relies on multiple layers of representation of data and is thus able to represent complex relationships between inputs and outputs. However, learned representations are difficult for humans to interpret. Therefore, the application of Artificial Intelligence has remained minimal. Many users still consider AI as merely an algorithm pooling from an infinitely large amount of data and making associations to inference. A popular conclusion is that despite being classified as an advanced technology, it still is not advanced enough to replace humans.

There is an existing lack of peer-reviewed research on the practical use of these Artificial Intelligence algorithms. Many studies have, however, pointed out concerns regarding the existence of algorithmic bias. These algorithms are then reflecting back an amplified bias. For example, most robo-advisor algorithms (an algorithm used to replicate the work of financial advisors) do not factor in important realities for their female investors, such as pay gaps, career breaks, and longer average lifespans. Additionally, in the United States, more African Americans have been denied loans or granted longer prison sentences compared to their Caucasian counterparts. It has been summarized that algorithmic bias occurs when the application of an algorithm compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability, or sexual orientation to amplify them and adversely impact equity.

1.1.3 Data, inputs, and bias

According to Langston J et al. (2019, p. 2), "A consumer study of an image search on a popular search engine revealed that 11% of results for the term "CEO" were female." But according to Silberg J et al. (2019, p. 2), "At the time, 20% of CEOs in the US were women." The above example is a classic case of biased input that results in a biased output. Even if developers are conscious of algorithmic bias, if the data itself is biased, the results would be reflective of that. The issue is that raw data is the perfect metric for signifying a population's mentality. If there are any biases in a population data point, it transfers over. The truth about only 20% female CEOs translates into only 11% female CEO pictures on a search engine. The data discrepancy is also reflective of how algorithmic bias picks up on social bias and emphasizes it even more. This issue in AI development could lead to larger problems in society if left unchecked.

1.2 Case study: Algorithmic bias in health systems

Within the health community, there are concerns about how AI will abide by the standards that doctors have. Some algorithms used in the past for health systems have given biased results. The question of whether the algorithm is biased or if the results were simply reflective of the biased data remains. Without a quantitative metric to judge an algorithm's fairness, algorithms become subjective to embedding systemic inequities.

Additionally, medical professionals get a deep understanding of the diverse population they serve. They serve these different groups of people using their cognitive abilities and the standards of the profession taught to them during cohesive and rigorous education. Even if an AI model is made to be diversity-conscious, historically there is not an equal amount of data available for such diverse groups of data. According to Panch, Trishan, et al. (2019):

> Deep learning involves the transformation of data from the real world, such as the pixels of an x-ray, into multiple layers of numbers that are combined to create an output of a diagnostic category. In practice, there are up to 100 layers and the relative influence of different elements in each layer is established in the process of learning. This byzantine process yields powerful results, but exactly how it does so is difficult to establish. (p. 2)

Doctors are easily able to talk to the patients and their families with details of the diagnosis, the how, why, and when. However, some AI diagnoses don't allow for that reasoning.

Despite the issues AI can have, if those who will end up using it can trust it, it can be useful within health research and development. However, to get there, the issues discussed earlier need to be combated. Firstly, there seems to be a difference in the outlook of those who make these algorithms versus those who use them and those who are directly impacted. Earlier, the lack of diverse data was discussed. Since that discrepancy cannot be changed for historical data, developers must pay attention to data science technologies that can be used to make the data more appropriate for the end-functioning of the algorithm.

Developers often are focused on optimizing the performance of their algorithms with little regard for algorithmic bias. According to Munro R, et al. (2019):

> Often, there will likely be a trade-off between the speed of algorithm deployment and algorithmic bias. A reasonable control mechanism to counter this trade-off is to create 'human-in-the-loop' systems, where algorithmic outputs are passed to a human decision-maker with necessary caveats and the human is the ultimate decision-maker. (p. 3).

Currently, there is no such process of control mechanism because most algorithms don't go through an extensive peer review. A push to create such a review mandatory would further help mitigate algorithmic bias during development.

Developers are somewhat bound by the data given to them by data scientists. Any bias in this interconnected ecosystem of professions will cause algorithmic bias. According to the 2018 Interim report of the National Academies of Sciences, Engineering, Medicine, "there is a critical need for data scientists in health systems and the development of graduate training in health data science is timely." (p. 2) It must also be added that the universal acknowledgement of diverse teams being better for development of algorithms applies here as well. In addition to representation, awareness for the diverse populations that the data is collected from or for, will make data science teams less amiable to bias. These teams can be made more diverse by adding medical professionals who can help with recognizing the variable characteristics and parameters that would make the data more suited for an algorithm.

Secondly, it was earlier discussed how deep learning algorithms aren't capable of reporting on their internal process which results in a lack of transparency which is considered an important virtue in the medical community. This issue can be countered by possibly reverse engineering the process. If a specific change in input results in a corresponding change in a specific part of the output, it can help with the understanding of the process.

Lastly, regulation and intervention of the government could help set standards for AI. This would increase the trust people have in these algorithms which would further increase the investment in the research for the reduction of algorithmic bias.

CHAPTER 2

LITERATURE REVIEW

In a paper by Jason Chan and Jing Wang (2015), the hiring preferences in online labor markets are discussed. “Sub analyses show that women are preferred in feminine type occupations while men do not enjoy higher hiring likelihoods in masculine-typed occupations.” (p. 1) The study proves that gender bias is not only harmful to women but to men as well. It was also interesting to see that female employers were more likely to have a favorable bias towards hiring females. Despite the popularity of such platforms, a lack of understanding of this issue persists. These systemic biases create inequitable opportunities and have large societal and economic impacts due to their global presence.

When looking at the paper by Chan et al. (2015), it is important to remember that it is limited to online labor markets. They are different from the traditional markets by way of “types of jobs offered, worker composition, and quality assurance mechanisms” (p. 20), and hence the findings of this paper cannot just be applied to other markets. Another limitation to look at here is the possibility of the ratio of available applicants for a certain job. If there are more women applying for a certain job, statistically a hiring bias in their favor cannot be proven. It mostly goes to show the unfair societal norms that encourage women to only aim for certain jobs. Even if a hiring bias exists, the reasoning behind it must not be concluded prematurely by confusing causation and correlation. “It is possible that females receive an advantage in terms of higher hiring likelihood but suffer a penalty in wages received in the online market.” (Chan et al., 2015, p. 20)

In another paper authored by Panch, Mattie, and Atun (2019), the implications of gender bias in health systems are discussed. “For many, the concern is not only that “algorithms are for the most part reflecting back the bias in our world”, but that they are doing so at potentially massive scale and without due oversight” (p. 1). It also explains the challenge of the lack of a clear standard of fairness which makes it harder to make changes to AI systems. It proves how a general AI model cannot be implemented as there is not an equal amount of information available for different socio-economic groups. It is important to understand the drastic impacts this can have within the health sector. For example, pain in the left hand has been known as a symptom of heart attacks for years, but research shows that heart attack symptoms are in fact different for men and women. (Heart Attack Symptoms in Women, n.d.). This propagates a lack of knowledge of female health. This paper adds to the inherent risks of bias in Artificial Intelligence. By adding the perspective of health systems and implications to the larger health of the society, it creates an ethos with the reader. More papers like these would expand on the much-needed call to action to mitigate the bias. Additionally, this paper acknowledges that bias already exists in the world. It adds to the idea that these biases trickle into algorithms and our daily lives, further expanding the presence of such a social bane. This paper helps to establish that we must not only talk about algorithmic bias but also how to end it.

A study by Anja Lambrecht and Catherine Tucker (2019) discusses gender-based discrimination in the display of STEM career ads. An algorithm that was coded to be neutral ended up having a bias against women as fewer women saw a STEM career ad than men. “This happened because younger women are a prized demographic and are more expensive to show ads to." (p. 1) Therefore, cost-effective algorithms were proved to be inherently biased.

Lastly, Megan Garcia wrote a groundbreaking article that has made the topic of algorithmic bias significantly more popular than before. She uses a single case study of Tay, an algorithm that was designed to learn from Twitter users, to prove how algorithmic bias can be a by-product of the most neutral developments. "It was a friendly start for the Twitter bot designed by Microsoft to engage with people aged 18 to 24. But, in a mere 12 hours, Tay went from upbeat conversationalist to foul-mouthed, racist Holocaust denier." (Garcia et al., 2016, p. 112). Even though this experiment was dissolved within 24 hours and did not have any at-large impacts, it helped provide evidence that algorithmic bias exists in even the most nuanced codes including big tech and governmental institutions which have the ability to have drastic socio-economic impacts.

This paper by Garcia et al. (2019) emphasized governmental intervention which allowed for traction. It argued for algorithmic auditing as a way for human intervention whenever biased activity is detected in algorithms. It also expanded on the implications of deep learning that were discussed earlier. The paper focused on how certain algorithms were able to change their own code after the accumulation of new data. Self-programming algorithms make it harder to detect causation because it could be the data or the self-written code. Companies like Google are adopting deep learning algorithms at an increasing rate which makes the call for further research on mitigation of bias more essential than ever before.

In a paper by Diekman, Brown, Johnston, and Clark, (2010) the discrepancy in STEM careers between genders is argued. The paper talks about the lack of representation of women in the fields of science, technology, engineering, and mathematics (STEM). They state stereotypes as the main reason for this disproportionately. This paper expands the existing bias in society which translates into statistical data through respondent bias. Such big data poses a huge threat to the outputs of algorithms.

A very interesting term was coined in a study by Claude Draude, Goda Klumbyte, Phillip Lücking, and Pat Treusch (2019). They described a sociotechnical approach to algorithmic bias. "With reference to the term "algorithmic culture," the interconnectedness and mutual shaping of society and technology are postulated. A sociotechnical approach requires translational work between and across disciplines." (p. 326) Their research was instrumental in bringing a gender and diversity studies perspective to this information systems issue.

This paper also debunked the myth that ads are sometimes not shown to people who are less likely to provide further engagement. However, in this case, women were actually more likely to engage than men. They provide evidence for the gender imbalance to prove that women were merely not shown the ads because they are a more expensive demographic. Any algorithm that focuses on cost-effectiveness, therefore, subjects gender bias through algorithmic bias. Gender neutrality in such cases leads to discrimination as a factor of marketing costing methods. This paper is hence trailblazing in adding to the narrative of how discrimination in algorithms isn't always apparent. It requires a nuanced view.

In another study by Teodorescu, Morse, Awwad, and Kane, the risk of adding bias by using machine learning tools is discussed. Using algorithms to perform repetitive tasks works only if it is not a complex situation. The paper suggests that therefore, "human augmentation of ML tools is necessary" (2019, p. 1). It pushes for more research in this subject due to limited current research. This adds to the point of how algorithmic bias is prevalent and expanding and hence needs more peer-reviewed research. Information Systems research needs a more dynamic approach as it does not occur in a vacuum. There is bound to be some social and environmental intersection. This paper is also particularly helpful by characterizing the "typology of augmentation for fairness consisting of four quadrants: reactive oversight, proactive oversight, informed reliance, and supervised reliance" (2019, p. 1). It also points out how Machine Learning tools do not abide by the traditional Information Systems assumptions, theories, and concepts. This further signifies the importance of a new wave of research in this subject.

Research done by Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan, discusses whether machine learning can improve human decision making. It uses bail decisions as a case. Bail decisions are a concrete task with a large available data set which makes it a programmable task. They use past decisions made by judges for releasing defendants. The paper also uses econometric strategies to account for the personal preferences of judges. "Moreover, all categories of crime, including violent crimes, show reductions; these gains can be achieved while simultaneously reducing racial disparities" (Klienberg et al., 2017, p. 1). The study suggests that if statistically driven predictions are used in policy problems such as bail decisions, it can improve decision making. This study supports the idea that algorithms and their outputs are heavily dependent on the data set. In this example, if the sample set of decisions by judges chosen is biased or flawed, it would directly impact the result of the study. Hence, the origination of such a data set needs to be extracted with a focus on the common statistical fallacies and the resulting algorithmic bias.

In her book, “Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy”, Cathy O'Neil, a former Wall Street Quantitative Developer, discusses algorithms and mathematical models that are capable of systematically changing our way of life. She begins by listing the ways that mathematical models affect our everyday lives including, “where we go to school, weather we get a car loan, [and] how much we pay for health insurance” (2016, p. 271). O’Neil then points out that unfortunately these models are “opaque, unregulated, and uncontestable, even when they’re wrong” (p. 271) which mean they are biased, unfair, and do not “[judge people] according to the same rules” (p. 271). One example she uses to reinforce this idea is of a poor student who “can’t get a loan because a lending model deems him too risky (by virtue of his zip code)” (p. 271) this model, which reinforces discrimination, has far reaching impacts including “[cutting him] off from the kind of education that could pull him out of poverty” (p. 271). O’Neil concludes her argument with a powerful assertion that “Models are propping up the lucky and punishing the downtrodden, creating a ‘toxic cocktail for democracy’” (p. 271). This book supports the idea that developers must take more responsibility for their algorithms and that policy regulation is essential. The popularity of this book has led more and more people to ask questions and demand change about not letting big data negatively affect their lives.

Another study by Scott Morton, Zettelmeyer, and Silva-Risso, discusses how consumer information affects the pricing of new cars to women and minorities. “Online, we find that minority buyers pay nearly the same prices as do whites controlling for consumers' income, education, and neighborhood characteristics." (Morton et al., 2001, p. 65) This conclusion expands on the need for discrimination conscious data mining systems. Since offline transactions are not based on a diverse data set, they present a biased result. However, online transactions that were based on a cohesive non-discriminatory data set led to fair decisions.

CHAPTER 3

METHODOLOGY

The study used a meta-analysis to concur findings. Interview statistics were also collected to assess the respondent's understanding of algorithmic bias.

3.1 Interview Statistics

Through an online poll, people were asked various questions regarding their knowledge of algorithmic bias. 37% of females agreed that they mostly saw consumer advertisements on social media. This was followed closely by 34% who saw information from news outlets. Only 24% of women agreed that they mostly saw financial advertisements, and lastly, six per cent said they see job opening advertisements most frequently. Most males, 40%, said they mostly saw financial advertisements on their social media platforms. A close 33% said that they mostly saw consumer advertisements. 16% said that they saw frequent articles from news outlets and another ten per cent said the same for job opportunities.

Table 3.1: Most frequent advertisements seen on social media platforms based on gender

	Consumer Advertising (%)	News outlet (%)	Financial Tools / Education (%)	Job opening (%)
Female	37	34	24	6
Male	33	16	40	10

These results should be studied with an understanding of the different platforms used by the population sample. If a student frequently visits LinkedIn, they are more likely to see job opportunities in comparison to someone using Instagram more often who may then mostly see consumer advertisements. Additionally, respondent bias must also be considered when looking at these statistics.

Table 3.2: Understanding of algorithmic bias based on gender

	I know about and think it impacts me (%)	I have read about it (%)	Never heard about it (%)
Female	27	24	49
Male	22	25	54

As shown in Table 3.2 above, for females, 27% agreed that they are aware of algorithmic bias and think it impacts them, followed by a close 24% of female students who are aware of such bias. However, 49% of female respondents had never heard of such bias. In addition, most males (54%) said they had never heard of algorithmic bias. 25% said that they had read about it and only 22% said that they are aware of it and think it impacts them.

The study took a detailed analysis of preferences and demographic information to produce a controlled study as shown in the following tables. These statistics provide further confirmation of young women being a highly prized demographic for consumer advertising despite their preferences.

Table 3.3: Type of sites most visited by the respondents

	Social Media (%)	News (%)	Financial (%)	Job search related (%)
Female	29	28	20	23
Male	16	23	37	24

Table 3.4: Most frequent advertisements seen on social media platforms based on gender and age

Gender	Age	Consumer Advertising (%)	News outlet (%)	Financial Tools / Education (%)	Job opening (%)
Female	18-24	50	12	31	7
	25-34	61	16	21	2
	35-44	0	52	36	12
	45-54	0	63	25	13
	55-64	15	77	8	0
	65+	20	80	0	0
Male	18-24	39	9	39	12
	25-34	19	19	50	13
	35-44	36	18	32	14
	45-54	31	15	54	0
	55-64	22	22	44	11
	65+	50	50	0	0

Table 3.5: Most frequent advertisements seen on social media platforms based on gender and age and educational level

Gender	Education	Consumer Advertising (%)	News outlet (%)	Financial Tools / Education (%)	Job opening (%)
Female	No high school	75	0	25	0
	High school	68	6	22	5
	Associates	18	18	53	12
	Bachelors	19	54	21	6
	Masters	46	31	20	3
	Advanced	0	100	0	0
Male	No high school	43	0	43	14
	High school	31	31	19	19
	Associates	75	0	25	0
	Bachelors	32	6	53	9
	Masters	29	29	41	0
	Advanced	17	50	0	33

3.2 Meta study statistics

3.2.1 Advertising economics

Table 3.6: Raw data regarding click rates across different demographics

Age group	Female click rate (%)	Male click rate (%)
18-24	0.18	0.15
25-34	0.15	0.13
35-44	0.17	0.12
45-54	0.18	0.13
55-64	0.21	0.15

The results shown in Table 3.6 are taken from the raw data provided by Facebook. It takes into account the click rate (frequency of interaction compared to the number of advertisements shown) across different demographics (gender × age) across all countries. One clear observation is that women are more likely to interact with an advertisement upon seeing it. On average, the click rate for men was 0.131% and 0.167% for women. Due to a higher click rate, the advertisements shown to women become more expensive leading to fewer ads. This can have a negative effect. "If women are not exposed to information on STEM careers, they may never apply for STEM jobs." (Lambercht et al., 2016, p. 2970). Given that women are more likely to interact with advertisements, it is also proven that they are more likely to convert into actual customers as shown in Figure 3.1. Therefore, for retailers, the return on investment is higher upon showing advertisements to women since men are less likely to lead to a sale. Hence, they are willing to pay more to show advertisements to women making them a highly prized demographic.

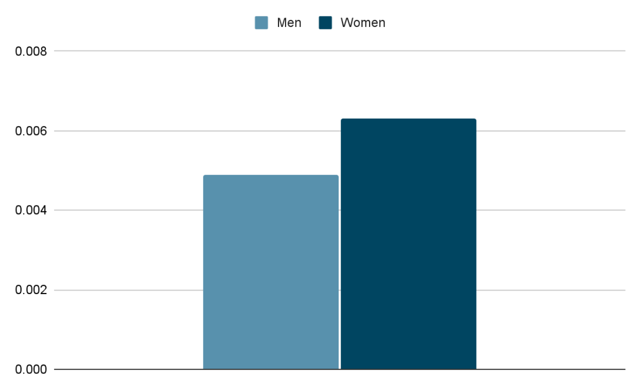

Figure 3.1: Women are more likely than men to convert after clicking (Lambercht et al., 2016, p. 2976)

Table 3.7: Results of Test on Instagram (Lambercht et al., 2016, p. 2977)

	Impressions	Click rate (%)	Cost per click ($)
Female	1560	0.27	1.74
Male	9595	0.59	0.95

So far, the paper has focused on data sets from Facebook. It uses a similar data set from Instagram to prove the general applicability of such results. Even though Instagram is owned by Facebook, it can be considered a vastly different platform with a similar advertising model. As shown in Table 3.7, only 15% of advertisements were shown to women, and men were more likely to click on an advertisement. This is consistent with the fact that women were more expensive to show ads to than men as proven by Lambercht et al. (2016)

3.2.2 AI behind autonomous vehicles

In Autonomous Vehicles (AVs), this sort of discrimination can be sourced from the designers and the datasets being used. Such discrimination can lead to safety risks as listed in Appendix B. Additionally, statistical bias also leads to unfairness with AV training. “For instance, training an AV using data from only one country could result in the AV learning localised patterns and not accurately modelling driving behaviours that apply in other countries or contexts (Lim et al., 2019, p. 7). Under or over representation can be lethal in such cases. “In American anti-discrimination law, discrimination exists when there is disparate treatment, which is the “discriminatory intent or the formal application of different rules to people of different groups”, and/or disparate impact, which is the result that “differ for different groups” (Lim et al., 2019, p. 7).

Ethical concerns, seen in Appendix B, are also an important concern in AV development. Some algorithms are trained to penalize or protect certain individual-specific characteristics, such as a person's age and gender. Some AVs might choose to protect children even if that means causing harm to a larger number of people who do not have the specific characteristics preferred by the algorithm. These algorithms continue to exist because they maximize profits and because of scarce legal frameworks.

CHAPTER 4

DISCUSSION

4.1 Significance

Given what is already known about the topic, the focus of this research is to prove why being conscious of algorithmic bias should be essential for big tech companies. Using previous papers and a case study of the recent changes that LinkedIn made to combat algorithmic gender bias, in this section it is proven how this is of importance ethically and economically. Information Systems has long been regarded as a back-office or support role. But this study would contribute to establishing it as an essential arm of any business or community.

World politics has been growing more polarized especially with online media's expansion. Other, older forms of media were not known to increase polarization the way online platforms have because online platforms allow for engagement from the entire population. With information readily available and a dramatic increase in access to communication, online platforms reach thousands if not millions of people. Most importantly, this stream of information is organized by algorithms to maximize the goals of the media channel (Sirbu et al., 2019). Showing information in an unbalanced method further expands this polarization (Sirbu et al., 2019). This is essentially how algorithmic bias is generated. A study by Sirbu, Pedreschi, et al, coined the term "algorithmic segregation" (Sirbu et al., 2019, p. 2) for this phenomenon.

Today, more people use online media instead of traditional channels to receive information (Sirbu et al., 2019). Algorithmic bias then uses the number of “likes” on each post to decide which information gets priority promotion. News, then, is not valued for its inherent content but popularity. These algorithms also target information that a user would likely agree with, leading to a lack of diversity in the marketplace of ideas. Apart from “likes”, “shares” also share responsibility for spreading “fake news”. This further enhances the exchange of ideas between individuals with similar ideas. Thus, algorithms indirectly lead to the creation of an echo chamber effect (Sirbu et al., 2019, p. 2). An extension of echo chamber effect is that “algorithmic bias in the evolution of the social network (i.e. by social recommendation systems) demonstrates the glass ceiling effect, where some user groups (e.g. female users) are excluded from the top layers of the social network.” (Sirbu et al., 2019, p. 2).

4.2 Resistance from big tech

Recently, there have been more and more cases of algorithmic bias on popular websites that have brought focus to the subject. These large companies have so far apologized for algorithmic bias by diminishing it to a technical difficulty unbeknownst to the fact that such tech issues lead to marginalization along the lines of gender, race, and socioeconomic status. (Hern, 2020)

Last September, Twitter had a similar situation. Their image cropping algorithm was programmed to choose white faces and ignore black ones. Their public statement after the incident was that they went through their processes to check for bias, but their check systems failed (Hern, 2020). Likewise, TikTok was under fire for suppressing the Black Lives Matter hashtag and continued to call it a tech issue as well. So long as companies do not recognize algorithmic bias as a feature, but as a tech issue, it will be harder to make substantial change. Since the product of actions by big tech firms is marginalization, their actions must be seen as intentional. Allowing white, male, and privileged executives in Silicon Valley to market their companies as an advanced tool for society instead of the detrimental tool that it truly is, ensures the existence of algorithmic bias. Ruha Benjamin, a professor at Princeton University coined the term “The New Jim Code” (Benjamin, 2019) as new technologies that reproduce existing inequities while appearing more progressive than the discriminatory systems of a previous era.

4.3 Case study: Ellevest

To expand on successful companies that have dismantled algorithmic bias and used it to their advantage, a case study on Ellevest, a digital interesting platform for women is discussed. This was launched by Sallie Krawcheck, the former Citigroup CFO. This startup focuses on the issues that women face in investing that are not given much attention because men do not face the same issues.

Ellevest considers various factors in their algorithm including men earning higher salaries than women, and women living longer than men. The former helps women make a financial plan better suited for their needs and the latter is essential when planning for retirement. The algorithm also automatically accounts for women having to take extra time off work for children.

Their algorithm for determining and dynamically adjusting risk capacity is patented and the U.S. Patent and Trademark Office released the following abstract regarding this patent, "A rules-based engine receives the user information and derives risk capacity, wherein the rules-based engine provides the user initial portfolio recommendations that are based on the derived risk capacity and goals as well as automatically rebalancing the derived risk capacity as a time horizon approaches." (Kwan, 2016, p. 1)

4.4 Further research

As discussed earlier, this subject promptly needs further peer-reviewed research to create a checks and balances system for new algorithms. These research studies must particularly work to demystify Artificial Intelligence and to garner support for government interest.

With an increase in the use of Artificial Intelligence, the challenges of algorithmic bias have also increased. Since these algorithms are made by humans, a margin of error must always be accounted for. Automation is not perfection. Yet there are no regulations or common practice standards for oversight in this field. The lack of transparency in some Artificial Intelligence algorithms makes such intervention a hard task but still a necessary one. Therefore, further research would pave the way for fairer systems.

CHAPTER 5

CONCLUSION

The goal of this research was to find the source of algorithmic bias leading to possible procedural solutions that can be applied widely. In the first chapter, the paper discussed the generation of bias by users, designers, and data. It gave a holistic view of the dynamic nature of this subject. Existing literature and meta-analysis were then used to understand how this bias expands after going through an algorithm. A lack of awareness and propaganda from big tech has kept the research on this subject minimal. A case study on the impact on health systems as well as a discrimination conscious platform called Ellevest were used. The study concludes that a system of regulatory oversight is necessary and such government intervention can be achieved through further research.

APPENDIX A

RAW DATA FROM INTERVIEWS

Table A.1: Most frequent advertisements seen on social media platforms based on gender

Gender		Consumer Advertising	News outlet	Financial Tools / Education	Job opening
Female	Percent	37%	34%	24%	6%
	Actual	60	55	39	9
Male	Percent	33%	16%	40%	10%
	Actual	32	16	39	10

Table A.2: Understanding of algorithmic bias based on gender

Gender		I know about and think it impacts me	I have read about it	Never heard about it
Female	Percent	27%	24%	49%
	Actual	44	39	80
Male	Percent	22%	25%	54%
	Actual	21	24	52

Table A.3: Type of sites most visited by the respondents

Gender		Social Media	News	Financial	Job search related
Female	Percent	29%	28%	20%	23%
	Actual	47	45	33	38
Male	Percent	16%	23%	37%	24%
	Actual	16	22	36	23

Table A.4: Most frequent advertisements seen on social media platforms based on gender and age

Gender	Age		Consumer Advertising	News outlet	Financial Tools / Education	Job opening
Female	18-24	Percent	50%	12%	31%	7%
		Actual	21	5	13	3
	25-34	Percent	61%	16%	21%	2%
		Actual	35	9	12	1
	35-44	Percent	0%	52%	36%	12%
		Actual	0	13	9	3
	45-54	Percent	0%	63%	25%	13%
		Actual	0	10	4	2
	55-64	Percent	15%	77%	8%	0%
		Actual	2	10	1	0
	65+	Percent	20%	80%	0%	0%
		Actual	2	8	0	0
	Total	Percent	37%	34%	24%	6%
		Actual	60	55	39	9
Male	18-24	Percent	39%	9%	39%	12%
		Actual	13	3	13	4
	25-34	Percent	19%	19%	50%	13%
		Actual	3	3	8	2
	35-44	Percent	36%	18%	32%	14%
		Actual	8	4	7	3
	45-54	Percent	31%	15%	54%	0%
		Actual	4	2	7	0
	55-64	Percent	22%	22%	44%	11%
		Actual	2	2	4	1
	65+	Percent	50%	50%	0%	0%
		Actual	2	2	0	0
	Total	Percent	33%	16%	40%	10%
		Actual	32	16	39	10

Table A.5: Most frequent advertisements seen on social media platforms based on gender and education level

Gender	Education Level		Consumer Advertising	News outlet	Financial Tools / Education	Job opening
Female	No High school	Percent	75%	0%	25%	0%
		Actual	3	0	1	0
	High school	Percent	68%	5%	22%	5%
		Actual	25	2	8	2
	Associates	Percent	18%	18%	53%	12%
		Actual	3	3	9	2
	Bachelors	Percent	19%	54%	21%	6%
		Actual	13	37	14	4
	Masters	Percent	46%	31%	20%	3%
		Actual	16	11	7	1
	Advanced	Percent	0%	100%	0%	0%
		Actual	0	2	0	0
	Total	Percent	37%	34%	24%	6%
		Actual	60	55	39	9
Male	No High school	Percent	43%	0%	43%	14%
		Actual	3	0	3	1
	High school	Percent	31%	31%	19%	19%
		Actual	5	5	3	3
	Associates	Percent	75%	0%	25%	0%
		Actual	3	0	1	0
	Bachelors	Percent	32%	6%	53%	9%
		Actual	15	3	25	4
	Masters	Percent	29%	29%	41%	0%
		Actual	5	5	7	0
	Advanced	Percent	17%	50%	0%	33%
		Actual	1	3	0	2
	Total	Percent	33%	16%	40%	10%
		Actual	32	16	39	10

APPENDIX B

AI BEHIND AV

Table B.1: Summary of ethical issues (Lim et al., 2019, p. 11)

	Ethical Issues	Proposed Solutions
Bias	Sources of bias in AV algorithms: Statistical bias and including personal characteristics in the data. Manufacturers and programmers can program algorithms to favour AV users' safety to boost profits. Large-scale replication of algorithmic preferences in AVs can perpetuate systemic discrimination. Challenges of detecting and correcting bias: Algorithmic opacity masks decision-making logic. Data-driven and unpredictable nature of ML-based decisions makes it dicult to predict bias. Humans are excessively trusting of algorithmic decisions due to "automation bias". Difficult to prove discriminatory intent in algorithms.	Proposed solutions: Modify the data, algorithm and output to offset bias. Measure and test for data bias, and identify the affected individuals. Clarify the standards to evaluate bias in algorithms. Increase transparency via traceability and interpretability. Steps taken: AI guidelines that emphasise on fairness, transparency and accountability—Japan, Singapore. Creating design and testing methods to mitigate bias and discrimination from AI—South Korea, UK. Prohibiting the use of sensitive personal data in automated decisions and mandating a right to explanation—EU GDPR.

REFERENCES

Chan, Jason, and Jing Wang. "Hiring Preferences in Online Labor Markets: Evidence of a Female Hiring Bias." *SSRN Electronic Journal*, 2015, doi:10.2139/ssrn.2619922.

Cramer, Henriette, et al. "Assessing and Addressing Algorithmic Bias in Practice." *Interactions*, vol. 25, no. 6, 2018, pp. 58–63., doi:10.1145/3278156.

Danks, David, and Alex John London. "Algorithmic Bias in Autonomous Systems." *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017, doi:10.24963/ijcai.2017/654.

"The Department of Global Health and Population (GHP) Seeks to Improve Global Health through Education, Research, and Service from a Population-Based Perspective." *Department of Global Health and Population*, www.hsph.harvard.edu/global-health-and-population/.

Diekman, Amanda B., et al. "Seeking Congruity Between Goals and Roles." *Psychological Science*, vol. 21, no. 8, 2010, pp. 1051–1057., doi:10.1177/0956797610377342.

Draude, Claude, et al. "Situated Algorithms: a Sociotechnical Systemic Approach to Bias." *Online Information Review*, vol. 44, no. 2, 2019, pp. 325–342., doi:10.1108/oir-10-2018-0332.

Fejerskov, Adam Moe. "Algorithmic Bias and the (False) Promise of Numbers." *Global Policy*, vol. 12, no. S6, 2021, pp. 101–103., doi:10.1111/1758-5899.12915

Garcia, Megan. “Racist in the Machine.” *World Policy Journal*, vol. 33, no. 4, 2016, pp. 111–117., doi:10.1215/07402775-3813015.

“Harvard T.H. Chan School of Public Health.” *Department of Health Policy and Management*, www.hsph.harvard.edu/health-policy-and-management/.

Kleinberg, Jon, et al. “Human Decisions and Machine Predictions.” 2017, doi:10.3386/w231.

Lambrecht, Anja, and Catherine E. Tucker. “Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads.” *SSRN Electronic Journal*, 2016, doi:10.2139/ssrn.2852260.

Morse, Lily, et al. “Do the Ends Justify the Means? Variation in the Distributive and Procedural Fairness of Machine Learning Algorithms.” *Journal of Business Ethics*, 2021, doi:10.1007/s10551-021-04939-5.

Morse, Lily, et al. “A Framework for Fairer Machine Learning in Organizations.” *SSRN Electronic Journal*, 2020, doi:10.2139/ssrn.3690570.

Morton, Fiona Scott, et al. “Consumer Information and Price Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?” 2001, doi:10.3386/w8668.

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

Panch, Trishan, et al. “Artificial Intelligence and Algorithmic Bias: Implications for Health Systems.” *Journal of Global Health*, vol. 9, no. 2, 2019, doi:10.7189/jogh.09.020318.

Sîrbu, Alina, et al. “Algorithmic Bias Amplifies Opinion Fragmentation and Polarization: A Bounded Confidence Model.” *Plos One*, vol. 14, no. 3, 2019, doi:10.1371/journal.pone.0213246.

Teodorescu, Mike, et al. “Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation.” *MIS Quarterly*, vol. 45, no. 3, 2021, pp. 1483–1500., doi:10.25300/misq/2021/16535.

Zettelmeyer, Florian, et al. “How the Internet Lowers Prices: Evidence from Matched Survey and Auto Transaction Data.” 2005, doi:10.3386/w11515.

“A Financial Company, by Women for Women.” *Ellevest*, www.ellevest.com/.

BIOGRAPHICAL INFORMATION

Ashni Walia is a senior majoring in Information Systems with a minor in Mathematics at the University of Texas at Arlington. Upon graduation, she is beginning her career as an Investment Banker in New York City. She has previously, completed a private equity and a tax internship. During her years at UT Arlington, she has held many leadership positions including President of Honors College Council, President of Association of Women in Mathematics, Orientation Leader, Chancellor of Delta Sigma Pi, and Freshmen Leaders on Campus.