Analysis of Algorithmic Bias and its Interaction with Society

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Mark Chitre

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kent Wayland, Assistant Professor, Department of Engineering and Society

Background

With the advent of the internet and the World Wide Web, there are countless new technologies that have been made to improve human experience. From research on drones to social media, technology has been used every which way and continues to shape the way that humans, society, political entities, and nature interact. The usage of computers is essentially required in today's world, making it so that the ideas implemented in technology affect almost everyone in the world. One idea that has been at the forefront of technical discussion for a while has been the ethics of algorithms and other automated programs. Algorithms (a set of instructions given to a computer) can determine a great deal of things- whether someone is selected for a job interview, if someone becomes popular on social media, and many more things that affect daily life. Although most algorithms are created to help people, the way some algorithms are developed and trained can cause issues with them. The most common of these is algorithm bias, where the output of algorithm misrepresents sets of data either because of the data or produces an inaccurate result. Even processes as refined as face recognition (which uses a series of algorithms) have bias because of the way the algorithms are trained. This is very important because algorithms can grant access to the personal information of a person and perpetuate issues with privacy. The usage of algorithms is crucial to processes in everyday life and as a result there should be great pressure to refine and improve their accuracy. The issue of algorithm bias affects everyone using software because it can perpetuate the issues that are present in society. This is partly due to technology being a reflection of society, but these issues can also stem from poor design and development tendencies.

Although it is very complex, algorithmic bias can be remedied through several different forms of change in the future. This ranges from changes in CS education to using social issues to push for alteration in how algorithms are developed. Tracing how biases are generated in the creation of algorithms will be valuable to addressing this issue and confirming specific modifications to the current development process of algorithms will be very helpful. Using the lens of a socio-technical framework will also contribute to analyzing the interaction between algorithms and social change. Specifically, this paper aims to parse through and analyze several cases and then provide support to different ways of resolving algorithmic bias, showing how social movements have brought issues in technology to the forefront of the public eye. Ultimately, through new data sources and extra measures applied during development the issues brought out by algorithmic bias can be minimized and even possibly prevented.

Literature

There are several fields in which algorithmic bias is prevalent, but the one that this literature analysis will focus on is the algorithmic bias within healthcare. The journal article by Panch et al. goes into depth on how societal disparities in the current world are also reflected in health systems that utilize complex algorithms as well. This journal article was chosen because contextualizes the issues brought up in the problem frame and follows a problem, challenges, solution flow giving a well-rounded description of algorithmic biases in the case of health care systems. This journal article shows that the situation is dire and has been addressed but there is still a great deal of work to be done. By using a specific case, this article clarifies the situation in which algorithmic biases arise and are found. This supports the argument that there is a way for algorithmic biases to be resolved by giving suggestions and past uses. There are three major problems that are addressed in this journal: a lack of definitions and standard of “fairness”, insufficient contextual specificity and the “black-box” nature of algorithms (Panch, 2019). In addressing the first problem, the paper explained that in the world not everything is equal using the example of female CEO's. A search engine's results for "CEO" were 11% women, but across the US only 20% of CEO's are women (Panch, 2019). There is no conclusion for whether this is fair or not because while it is not equitable it is closer to proportionate. There is a lack of training data for different socio-economic groups as well, so when models are developed there is implicit bias woven into the model. Lastly, because these models are very complex, it is hard to know which layers and standards contribute to the output compared to the other layers. Since it is difficult to pinpoint, this serves as a challenge within the development and analysis of these algorithms. These issues highlight the challenges of resolving algorithmic bias while also bringing attention to specific parts of society and how they interact with these issues.

The journal also discusses possible actions that can be taken in order to combat these issues, further analyzing how alterations in development can combat the root of these problems. The four methods that are stated are: establishing context in which algorithms are developed/deployed, establishing processes to counter the risks of bias, balancing development in the discipline of health data science, and transparency/explainability in algorithm development (Panch, 2019). Using these main methods, building a framework around their suggestions will be valuable to use when addressing other cases in the field. This new framework would have major topics through which other cases will be able to be analyzed more concisely. The topics include contextualization, transparency, and countering the risks of bias, all which have been discussed heavily in this article.

Using the specifics of the methods, this can be applied to algorithms in a very beneficial way. Establishing context opens up different ways to address algorithms by explaining that algorithmic bias is not solely a technical issue that requires an engineering solution (Panch, 2019). This echoes the fact that developers need to be cognizant of the needs to different groups in order to take a preliminary step to limit bias. The process of development has to be reviewed and regulated by many teams with different interests to account for society's differences. The approach of also including a human-decision maker as an intermediary in algorithms counters issues that arise with algorithms, although reducing the speed of development (Panch, 2019). This falls under establishing processes to counter the risks of bias. The article also states that there is a need for more than software developers in creation of these algorithms. In the case of healthcare, it is important that healthcare professionals are part of this process to improve judgement, which adds to the diversity of professionals involved (Panch, 2019).

This scholarly literature also discusses the interaction that is analyzed in this paper between societal issues and algorithmic bias. After each of the methods that were stated, there was a general recommendation with further implementation challenges to show the implications of attempting to resolve algorithmic bias in that particular way. This paper concluded that there is always an inherent risk with using algorithms in health systems, and biases in society will manifest themselves in algorithms until mitigation features are optimized to prevent this.

This research showed a specific case in algorithmic bias with health systems and used it to highlight how societal issues are perpetuated in the form of these algorithms. This adds to the notion that society and technology such as algorithms interact heavily and can affect each other to different extents. This work assists with the understanding of this concept in the provided context, additionally showing that this can be applied further to several different fields and pieces of work.

Methods

The data collection done in this paper will be the corroboration of multiple pieces of scholarly research and literature. This also includes the compilation of research data presented in scholarly papers concerning algorithm bias and general conclusions that are made in those papers as well. These papers would include research on health-care systems and facial recognition technology among other topics. Using the main sources of literature and research data to collect evidence will help bridge the gap between empirical data and the social implications of algorithms as well. The data collection is rigorous because not only are several cases of healthcare algorithm biases explored, other cases in different fields are also accounted for in this matter. After the collection of this data, the analysis will provide further insight on trends and correlations that will help discuss the problems presented in a new light.

Theory

There are several different perspectives from which these issues can be looked at, which are all valuable information in this discussion. The scholarly article that was just discussed addresses algorithmic bias from the perspective that existing societal biases are perpetuated in these algorithms. The theory that developments in society shape technology and its advancements is highlighted heavily in this scholarly article, with a great importance put on social justice issues contextualizing how algorithms have affected the world. From this angle, it appears that there needs to be a great change in social dynamic with resolutions to algorithmic bias following. While the article approaches from this perspective, it also provides recommendations that support the opposite of this approach. The recommendations given show that mitigating bias can start at the root of the problem in the development stages of the algorithms, by focusing on social issues during the development stages and confirming that empirical data is properly represented and accurate to a predetermined extent. While alteration of algorithms during development is the easier solution, the misrepresentation of data has to continue to be addressed in order to alleviate future inaccuracies in artificial intelligence models.

Cases

There are several fields in which algorithmic bias is present, including healthcare and its auxiliary practices. Out of the general topic of healthcare there are several specific instances where algorithms are used, including disease detection. For melanoma, Convolutional Neural Networks (CNNs) have been used to determine whether skin lesions are a form of melanoma or something else. A CNN takes in an image, runs many algorithms on it, and then determines the result- whether a patient has melanoma. To train this network to be able to predict results on new images, many pictures have to be fed into the model so it can form a baseline for what is melanoma and what is not. The training and results of this model show a huge lack of inclusion of black patients: only 5-10% of the training data contains skin lesions on black patients and as a result the model is about half as accurate for them (Norori, 2021). This is alarming because the survival rate from melanoma is far less for black patients than white patients (94% for white patients, 70% for black patients) (Norori, 2021).

While on the topic of image recognition algorithms, this racial bias is perpetuated by some facial recognition algorithms as well. This is important to bias in healthcare systems because it discusses major issues in computer vision algorithms which are very commonly used in healthcare to analyze things like X-rays. Amongst the facial recognition algorithms with the highest classification accuracy (Amazon, Google, IBM, Face++, Kairos), these algorithms were still not performing as well as for colored females as they were for lighter skinned males; they were performing about 34% worse for darker-skinned females (Najibi, 2020). In response, these companies have changed their data collection methods and testing cohorts.

Perhaps the most significant study concerning algorithm bias in healthcare systems came out in 2019 discussing the algorithms that determine risk scores for patients. Although the software system sold by Optum was considered to be race-blind (not including race in calculations), there was still a great deal of embedded bias that was uncovered (Obermeyer, 2019). This widely used algorithm would display risk scores based on a plethora of compiled data; at a given risk score, black patients would be considerably sicker than white patients making it so white patients would receive additional help at a rate far more than black patients (Obermeyer, 2019). The reason for this bias is because the system predicts cost of healthcare and not illness, but it perpetuates spending less money for black patients than white patients (Obermeyer, 2019). The corroboration of these three sources shows a disparity between gender and mainly race, showing that the social effects of algorithms need to be gauged before they are developed because of their magnitude.

Analysis

Although individual pieces of data, there are a lot of common patterns amongst what has been explored. There is a strong underrepresentation of minority groups in training data which is the cause of some algorithmic bias issues. As shown in the melanoma study and the algorithm that determined risk scores, the model was systematically wrong due to the data that was fed into it, even when the data did not “include” race. The lack of data for certain groups seems to be a common scapegoat for many companies, which is normally followed by efforts to increase inclusion and improve data collection.

Parallels to the algorithm bias in the field of healthcare exist when looking at other cases as well. Although these parallels have different contexts and information, they contribute to the argument that there are poor tendencies in algorithm development perpetuating unjust ideals and representation. In light of this, another perspective that should be addressed when looking at this data is from a social justice perspective because it relates heavily to what the data concerns. Around May of 2020, the Black Lives Matter movement picked up steam, causing numerous people and media sources to address the disparity in existing technological systems. The social issues of the movement affected many fields of science, increasing the resources put together to search for inequality and non-representation. The article discussing racism in facial recognition technology directly references these events and goes on to confirm that minority groups such as Latinx and Black people make up 99% of "suspected gang affiliates", a group which has no requirements (Najibi, 2020). This is one instance of how biases in society are reflected in algorithms because the data fed into these algorithms does not represent the population correctly. This demonstrating how social issues have pushed the criticism of certain technology to the forefront of issues. While there has been increased public attention directed towards the racial injustices presented in the field of healthcare, disproportionate spending towards groups still occurs to this day but to a lesser extent due to the research done by Obermeyer et al. The timeline of the events shows that technology has pushed forward social issues to an extent, but there is a far greater impact on social justice issues pushing forward technological advancement.

Comparing the three cases (facial recognition, melanoma, bias in healthcare systems) presented systematically provides a new insight that should be explored. The literature with algorithmic biases in healthcare shows a great deal about development and the societal biases that are reflected because of the training data provided to algorithms. The facial recognition technology about suspected gang affiliates is similar to this because of the misrepresentation of data leading to biases, but this stems from choices that people make in society rather than data collection underrepresenting groups. Both of these cases do show that there are tremendous impacts of society in the development of technologies such as algorithms. The melanoma case also has clear parallels with the issues that are brought up in the journal article about algorithmic bias as a whole. Both go into detail about how software can skew data unknowingly and this shows that healthcare algorithms have to prove a lot before they can be used to make life-changing decisions. Speaking of the decision, the case with the model that assigned severity scores to make a queue of patients getting help also confirms the fact that biases reflected in algorithms need more attention. No matter what model is used in any of the cases, they all exploit the flaws of algorithms in decision-making in healthcare.

Throughout this analysis, it is valuable to apply the framework that was created as a result of the literature review to more deeply look at what needs to be improved/what is missing. When looking at each of the cases in terms of their context, the main theme is that the developers were cognizant of the groups that they were creating their product for, but did not take into account the differences that would result in poor performances of their algorithms. This can be seen through the melanoma case heavily because even though African Americans are at higher risk for malignant melanoma, the image analysis algorithm performed worse for them. Another aspect of this framework is transparency, where the process in which the algorithms are developed are publicly shown as to not hide anything within the algorithms. This part of the framework was interesting to address because most of the cases showed that the development process was unbiased; this leads to a major conclusion that how data is collected is the root of many negative effects dealing with algorithmic bias. Nearly all of the cases showed that the data which was fed into the algorithm was in some way not represented correctly, which adds to the validity of data being a large issue. This part of the framework also ties into another part, countering the risks of bias. Most of the cases did not do a sufficient job of addressing this, as shown by the backlash they have all received after rolling out the software. This aspect makes the testing before release of an algorithm important to address future risks. Through the lens of this newly created framework, it was apparent what was wrong in the cases listed and it was also apparent what could be improved.

The cases and research discussed in this paper show how algorithms cause and solve problems. The melanoma case shows that algorithms are overall beneficial for use in that domain, but they are not good enough because they do not perform fairly. This also confirms what many of the cases brought up- algorithms can only be developed fairly by predicting incorrect outcomes and compensating for them. By pushing this argument, algorithms should be given data that not only represents all groups but also purposely skews data so that they are fair. The socio-technical frameworks used also add to this because in some cases the outputs of technology need to fall in line with social justice or there will be problems. What this entails is that if current technological innovation does not properly reflect what society deems as a pertinent issue, there will be outcry about the technology not taking into account popular issues that people in society have just learned about. Additionally, recency bias with social issues plays into this, as many of the blind spots of technology have been searched for as a result of the popularity of certain issues. Overall, the cases discussed show how algorithm bias can be prevented by accounting for the socio-technical interaction as well. This overarching theme means that in order to mitigate the major sources of algorithm bias, one cannot solely look at technology but has to make sure to factor in the relationship between technology and society.

These cases go to show that there is a very distinct balance with technology and society, something that is reflected with algorithmic bias in healthcare as well. This ties back to the fact that things have to change with both technology and its development, and society in order for improvements to be made- starting with technology which is easier to control from an engineering standpoint.

Conclusion

When dealing with the field of technology and computer science specifically, there are countless unintended consequences that arise after each project. Far less problems arise when these consequences are considered in development, but there are also many pieces of the puzzle that are missing at that time. With algorithms being used on such a large scale currently, it is very difficult to address all issues while rolling out software in a speedy manner. The socio-technical interaction shown also sheds a new light on how algorithm bias and justice issues have bounced off of each other, looking to prove the larger point that technology and society sway each other.

The exploration of algorithm bias in healthcare highlights a lot of the general issues that lead to algorithm bias, suggesting resolutions along the way that can remedy many of the problems discussed in the following cases. The main issues arise from the data in question and not the algorithm itself (with the exception of the decision-making algorithm assigning severity scores). Whether or not they come from development, these issues lead to unfair treatment and have pushed forward the same problems that are also at the forefront of world news as well.

All in all, there needs to be change in the field of computer science because the issues that incorrect algorithms perpetuate is wrong. By synthesizing the data that was collected, a better view on how this can be resolved in healthcare and auxiliary fields is presented. Adding socio-technical frameworks allows for this issue to be looked at through a lens that brings up problems that can be solved and should be addressed in the future. After collecting and analyzing the research data, society and developers need to take significant steps to remedy the issues of algorithmic bias in parallel.

References

Alake, R. (2020, April 28). Algorithm bias in artificial intelligence needs to be discussed (and addressed). Medium. Retrieved November 1, 2021, from https://towardsdatascience.com/algorithm-bias-in-artificial-intelligence-needs-to-bediscussed-and-addressed-8d369d675a70.

Gordon, A. (2019, July 17). Computer Science vs. Computer Programming: What's the difference? Medium. Retrieved November 1, 2021, from https://medium.com/@adamjgordon24/computer-science-vs-computer-programmingwhats-the-difference-5e3764be9532.

How data structures and algorithms are important for computer science graduates - SYNERGISTICIT. Synergistic It. (2021, April 12). Retrieved November 1, 2021, from https://www.synergisticit.com/how-data-structures-and-algorithms-are-important-forcomputer-science-graduates/.

Hughes, T. P. (2009). Technology and society: Building our sociotechnical future. MIT PRESS.

Klein, H. K., & Kleinman, D. L. (2002). The Social Construction of Technology: Structural Considerations. Science, Technology, & Human Values, 27(1), 28–52. http://www.jstor.org/stable/690274

Matlin, C. (2016, February 2). Defining algorithms-a conversational explainer. Slate

Magazine. Retrieved November 1, 2021, from https://slate.com/technology/2016/02/whats-the-dealwith-algorithms.html. Najibi, A. (2020, October 26). Racial discrimination in face recognition technology. Science in the News. Retrieved November 1, 2021, from from https://slate.com/technology/2016/02/whats-the-dealwith-algorithms.html.

Najibi, A. (2020, October 26). Racial discrimination in face recognition technology. Science in the News. Retrieved November 1, 2021, from 10 https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognitiontechnology/.

NBCUniversal News Group. (2019, March 12). The World Wide Web is 30 years old - and its inventor has a warning for us. NBCNews.com. Retrieved November 1, 2021, from https://www.nbcnews.com/tech/tech-news/world-wide-web-30-its-inventor-has-warningus-n982156.

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10).

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

Panch, T., Mattie, H., & Atun, R. (2018). Artificial intelligence and algorithmic bias: implications for health systems. *Journal of Global Health*, *9*(2). https://doi.org/10.7189/jogh.09.020318

Spitz, J. (2021, July 1). *Why tech executives must embrace diversity as their first line of defense against the business impacts of algorithmic bias*. Forbes. Retrieved February 16, 2022, from https://www.forbes.com/sites/judithspitz/2021/07/01/why-tech-executives-must-embrace-diversity-as-their-first-line-of-defense-against-the-business-impacts-of-algorithmic-bias/?sh=67560a19c6d8

Utah State University Digitalcommons@usu. (n.d.). Retrieved November 1, 2021, from https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=3024&context=etd.

Vegas, E., & Fowler, B. (2020, August 4). What do we know about the expansion of K-12 Computer Science Education? Brookings. Retrieved November 1, 2021, from https://www.brookings.edu/research/what-do-we-know-about-the-expansion-of-k-12-computer-science-education/.