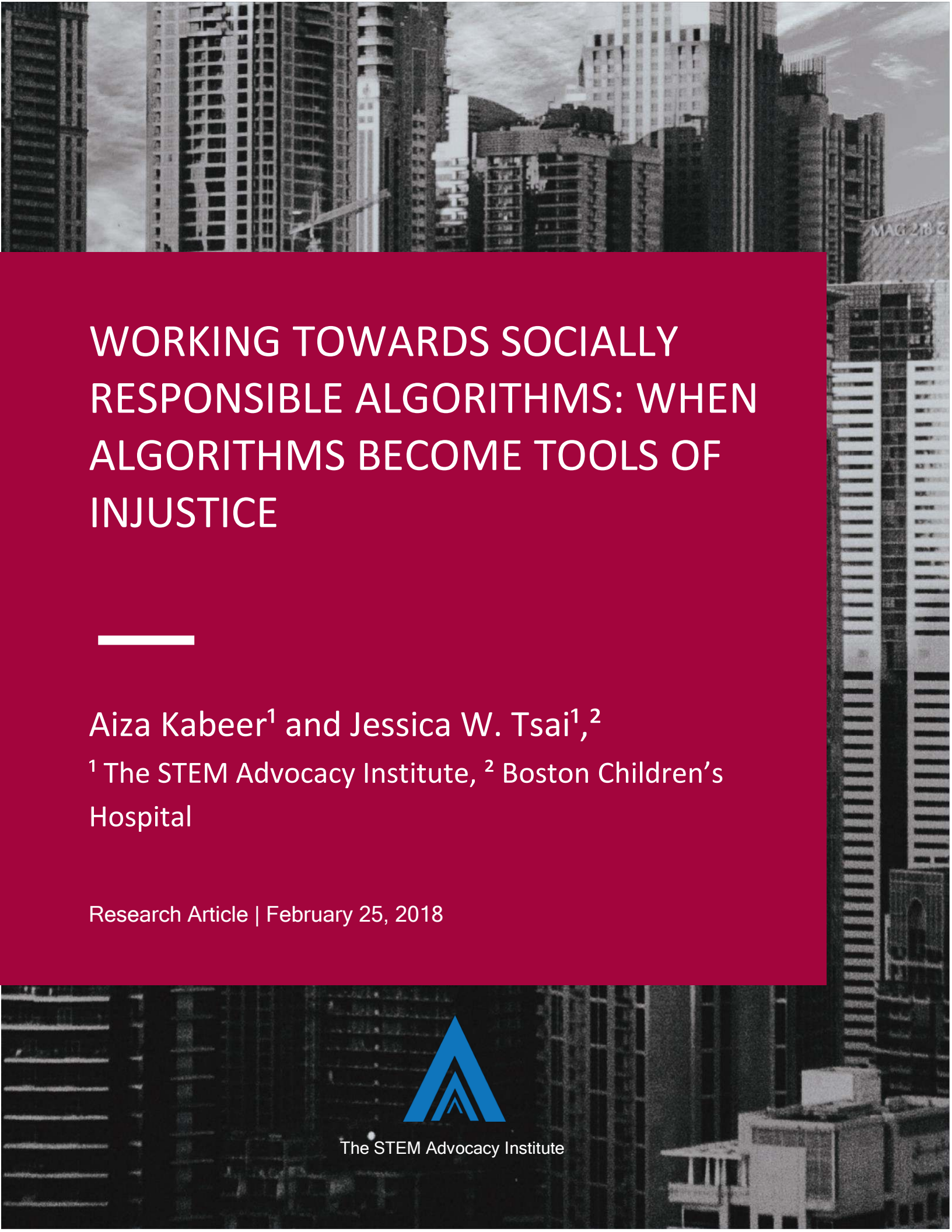

WORKING TOWARDS SOCIALLY RESPONSIBLE ALGORITHMS: WHEN ALGORITHMS BECOME TOOLS OF INJUSTICE

Aiza Kabeer[1] and Jessica W. Tsai[1,2]

[1] The STEM Advocacy Institute, [2] Boston Children's Hospital

Research Article | February 25, 2018

The STEM Advocacy Institute

Abstract

From criminal justice to financial markets, algorithms are now embedded in the fabric of our society. While algorithms benefit humanity, they can also result in discrimination and reinforce injustice. This article begins by defining algorithms and several types of algorithmic bias. The ways in which algorithmic bias can affect society are described. Potential solutions are discussed including further research, diversity initiatives, and education. Changes in education include changes to existing undergraduate computer science curricula as well as workshops or trainings in the tech industry. Beyond these solutions, policy measures will also be necessary for long-lasting change. Algorithmic bias is a serious danger to society. Algorithms are essential to the world as we know it, and we are responsible for the social implications of their use. If we do not prevent them from harming groups susceptible to discrimination, then we are at fault.

1. What are algorithms and what is algorithmic bias?

An algorithm is code that defines how to do a task by giving a computer instructions. An algorithm takes some input, runs commands, and produces some output.[1,2] For example, an algorithm might determine what advertisements are shown to a user on a website. More sophisticated algorithms can be used to automate a decision or make a prediction that a human would otherwise make. Deciding who is eligible for a loan, choosing the best route in Google maps, or making a critical prediction in the justice system might use such algorithms. Algorithms are also used to create artificial intelligence (AI).

AI is the science of making machines that do tasks in an intelligent manner that usually only a human can do. There are different types of AI, but the focus is on creating intelligent technology.[2] It can involve the use of algorithms and what is called machine learning. Machine learning is the ability of a system or a machine to learn and improve based on experiences without being explicitly programmed. In other words, machine learning allows systems to behave like a human by drawing conclusions from knowledge and experiences. By feeding algorithms data, they can learn and improve their output in repeated iterations.[3] The data used to teach an algorithm is called training data.

While this is an amazing technological advancement, the use of algorithms and AI is not foolproof and many complications exist. Algorithmic bias is particularly concerning. This usually refers to when the use of an algorithm negatively impacts minority groups or low-income communities in a discriminatory manner [1]. However, algorithmic bias can actually be classified into five types, as described in a paper written by Danks and London.[5]

---

[1] It is important to note that bias is not defined the same in statistics, social science, and law. Algorithmic bias has broad social and political implications, and the exact meaning of what bias is may become blurred across these fields. When discussing this issue, we are usually referring to bias beyond statistical usages.[4]

Type	Description	Example
Training Data Bias	When the input data an algorithm learns from is biased, the algorithm will retain those biases, but we only see the final model/learned behavior.	The training data for a self-driving car is only from one city, but the car will be used throughout the country.
Algorithmic Focus Bias	Information that should not be used for statistical, moral, legal, or other reasons is used to develop or train an algorithm.	Using information that is not legally permissible for certain types of judgements may cause an algorithm to deviate from legal standards.
Algorithmic Processing Bias	The algorithm itself is biased, sometimes intentionally, to compensate for other types of bias. This might be done to help an algorithm make ethical choices.	An autonomous weapons system is coded not to fire at perceived enemies if they are close to a UNESCO protected heritage site.
Transfer Context Bias	An algorithm is used in a context outside of its intended purpose	A healthcare algorithm that was developed for a research hospital is used in a rural clinic
Interpretation Bias	An algorithm's outputs are misinterpreted by the user or the broader autonomous system the algorithm functions in.	An algorithm that is used to make a prediction generates results that are misinterpreted by the user who is making a decision.

As we can see in the table, algorithmic bias has positive and negative effects. Algorithms are problematic when they result in biases that are unethical or discriminatory. Many of these cases involve training data bias where the input data is already biased. Datasets can reflect human biases since data is sometimes labeled by hand. Datasets may also exclude certain populations or otherwise be non-representative.[4]

Algorithms are in active use in our world today, and many are reinforcing systemic injustice. Some are unintentionally resulting in discriminatory practices in our society. We often assume that an algorithm always presents the best solution, but as the following examples illustrate, this is not always the case.

2. How can algorithmic bias affect society?

The journalistic organization ProPublica recently produced investigative pieces on the effects of algorithmic bias. In one of these studies, algorithms in the criminal justice system were found to reinforce racial disparities. The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a risk assessment system which uses algorithms to determine the likelihood of a defendant or convict committing a crime again in the future. In ProPublica's study, the algorithm was almost twice as likely to mislabel a black defendant as a future risk than a white defendant.[6]

In another study, ProPublica focused on an algorithm that determines online prices for Princeton Review's tutoring classes. The study showed that people living in higher income areas were twice as likely to pay a higher price than the general public. However, individuals living in a zip code that had a high density Asian population were 1.8 times more likely to have a higher price, regardless of their incomes.[7]

Additionally, Google has been guilty of using systems that fall prey to algorithmic bias. Google Images mislabeled an image of two black people as "gorillas." In another incident, Google showed ads of high paying jobs to men more often than women. Both of these examples are the result of algorithmic bias.[8,9]

The legal doctrine of disparate impact makes cases of unintentional racial discrimination illegal, including those that involve algorithms. In June of 2015, a statistical analysis of housing patterns showed that Texans were essentially segregated by race as a result of a tax credit program. In *Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, the Supreme Court used disparate impact theory to rule against housing discrimination. Disparate impact theory is an example of a policy that regulated an unfair algorithm, however it is limited to issues of housing and employment. We do not have a legal process to regulate the use of algorithms in all applications.[10]

In general, we lack overarching solutions or processes to deal with algorithmic bias. Society is more focused on the future effects of AI than algorithmic bias (See this NYT article for a relevant discussion).[11] Most of the action being taken revolves around discussion and research. Discussions are gaining momentum, but we have yet to see tangible solutions take form.

3. What is being done to address the issue?

There are some advocacy and research groups working on the problem of algorithmic bias. For example, the organization Data & Society researches social and cultural problems arising with data related technological advances, and covers topics relating to algorithmic bias.[12] More recently, groups like the AI Now Institute are coming together specifically to understand the social implications of AI.[13] Beyond these initiatives, there is a push to raise awareness of the issue. General discussions of algorithmic bias include pursuing more research and increasing diversity in tech fields. Dealing with this problem requires us to pursue solutions across many paths, including research, diversity initiatives, policy changes, and new ideas. More work is needed in all these areas.

3.1 Research

Algorithms can recognize patterns and have the capability to learn from those patterns. As discussed, they take in data and use instructions to result in meaningful output. From biology to financial lending, algorithms are commonly used in industry, academia, and in the military. Algorithmic bias also spans many disciplines. Although algorithms are developed within fields like computer science and mathematics, they impact issues tied to social problems and law. The questions we must answer about algorithmic bias also involve many fields. Can we program algorithms so that they do not cause discrimination? How can we ensure the data we use to teach algorithms is not biased? What role does the law play in dealing with instances of algorithmic bias? These are only a few of the questions that can be tackled through research, and they clearly cannot be answered by computer scientists alone. Considering the complexity of the situation, it may be important to include social scientific

and humanistic research practices into the development of algorithms, particularly AI research.[4]

Computer science research itself certainly has a role to play. Despite the widespread use of algorithms, we do not understand how algorithms learn and it is often unclear why AI makes one decision over another. Understanding how algorithms function might help us handle algorithmic bias when they are developed.

That said, research to clarify how algorithms work is a daunting task that will take time.[14] The good news is that there are policy and education based solutions we can implement without immediate answers from research in computer science. We may not know that much about how algorithms work, but it is inaccurate to say that we must understand them before investing in regulatory mechanisms, as mentioned in this article (e.g., policy solutions, regulatory bodies, ethical processes, etc).[15] While it is important to know why code results in certain decisions, we do not have unlimited time to wait for an answer. We need to work on solutions that can be implemented in the absence of conclusive results from computer science. Research related to policy, education, and other areas are vital in this respect.

Data and Society and the AI Now Institute are among some groups that are beginning to ask important questions and raise awareness about problems related to social science, ethics, policy and law. Organizations like Pervade are working to put together ethical processes for big data and computational research.[16] But there is more to be done, especially regarding algorithmic bias. We need widespread research that translates into tangible action before algorithmic bias hurts our society any further.

3.2 Diversity in Tech

Increasing diversity in tech fields is another strategy to combat algorithmic bias because it might result in more conscious programmers.[17] For example, the Algorithmic Justice League (AJL) was founded by a Black student because her face was not identified by facial recognition software.[18] Diverse students, and eventually diverse computer scientists, could mitigate bias.

The current landscape of computer science clearly lacks diversity. The National Science Foundation (NSF) provides data with explicit breakdowns by degree received and the race and gender of degree recipients. As per 2011 data, of the total number of students receiving Bachelor's degrees in computer science (U.S. citizens and permanent residents), only 10.6% were awarded to Black and African American students, and 8.5% to Hispanic or Latino students. For women of all races this percentage was 17.7%.

According to a publication of the U.S. Census Bureau based on 2011 data, among computer occupations, approximately 7.3% of workers are Black or African American and 6.0% are Hispanic (Figure 1). In both cases, this is around half of their overall representation in the U.S. population.[19] Of all races, 26.6% of workers are female (Figure 2). The broad group of computer occupations includes a variety of tech related jobs. When looking at a further breakdown of this group, the percentages of Black/African American and Hispanic workers remains low across various job types, but there is more variation in the percentages of female workers.[20] Regardless of this variation, it is clear that Black/African American and Hispanic/Latino populations are not fairly represented.

Moreover, large tech companies like Google are being called on to employ more diverse programmers.[21] Some well-known companies have ethics boards to prevent bias, but this is difficult to implement in smaller companies.[22] Industry employees need a better understanding of the social implications of algorithms, so that individual engineers, programmers, and data scientists consider the impacts of their work. One way to introduce

this awareness is by creating a diverse population in tech. Admittedly, increasing diversity in student populations and the workforce is a long-term solution. It will take time to get more minorities into computer science, and even longer to get them into industry.

While current developments are heartening, the dilemma remains. Algorithmic bias is operating in our society today, and we have yet to prevent it from contributing to structural injustice. Focusing on both research and building diversity is important, but we need a multi-pronged approach to addressing algorithmic bias.

4. Can an ethics-based education make a difference?

Besides research and increasing diversity, what else can we do? Another idea is to modify computer science education. Undergraduate computer science programs require a course that teaches algorithms, whether it is Data Structures and Algorithms, Introduction to Algorithms, Algorithm Design, or something similar. Incorporating ethics into the computer science curricula and highlighting algorithmic bias might have potential benefits. Some institutions currently offer courses on ethics in computer science, but such courses are not always a requirement of the major. Including ethics as an integral part of computer science programs and in continued education outside of academia might have potent results. While ethics in computer science covers a broad range of topics, we need to ensure that discussions of technology's implications for social inequality and discrimination also occur. This is key for an ethics-based education to ameliorate algorithmic bias.

4.1 Workshops

Workshops or short courses can educate students on the ethics of algorithms. Such a workshop would need to discuss the types of algorithmic bias and highlight the importance of training data. At the very least, computer science majors ought to participate in dialogue on algorithmic bias, especially in courses relating to algorithms. It is also pertinent to consider whether students in related fields (statistics, mathematics, etc.) should also be given exposure to these topics since these students may also work with algorithms.

4.2 Courses

Beyond simply holding workshops or making minor changes to course curricula, creating requirements for a course specifically dealing with ethics in computer science, AI, and related fields could have powerful effects. Many colleges and universities offer electives in this area, but they are typically not required. The general idea of Computer Ethics (CE) has been explored in the past as a core component of computer science curricula. A 2002 paper suggested including a course on CE and five relevant knowledge units in each year's curriculum. These units were classified as History of Computing, Social Context of Computing, Intellectual Property, and Computer Crime.[23] Admittedly, these units would have to be adjusted given social changes and technological progress since 2002, but it is a reasonable starting point. The Social Context of Computing would be most relevant to the issue of algorithmic bias.

More recent literature calls for including discussions of data, ethics, and law in computer science curricula. Barocas *et al*. discuss a research agenda focused on dealing with injustice and algorithms. Beyond research, their paper includes a suggestion to "weave" conversations of ethics and law into data science curricula and highlights the importance of a national conversation on these topics.[24] This holds true for all disciplines that use algorithms.

4.3 Ethics of algorithms incorporated into computer science education: examples

There are some programs that have ethics requirements. The computer science program at the University of Massachusetts Lowell is an example. However, a requirement in ethics courses is not universally acknowledged as a core part of computer science curricula. Even where ethics requirements exist, they must discuss discrimination and algorithms to be relevant to algorithmic bias.

As of 2018, a handful of universities are beginning to implement courses in ethics with the intent to educate their students on the potential consequences of emerging

technologies.[25] Harvard and MIT are currently offering a joint course dealing with ethics in AI, and the University of Texas at Austin is offering a course on ethics with plans to make it mandatory for all computer science majors.[25, 26, 27] These developments are wonderful. However, we need courses dealing with the social implications of technology to be normative in all computer science programs.

While there are few examples where ethics workshops have been made mandatory, the University of Nevada asked incoming graduate students across a range of engineering disciplines to take an ethics workshop in 2015. The workshop covered four major topics: Research Ethics, Computer Coding Ethics, Publishing Ethics, and Intellectual Property.[28] Only 7% of participants were from computer science, therefore it would be difficult to measure what impact the workshop had with regards to ethical and social issues like algorithmic bias. However, this demonstrates that ethics workshops in CS are certainly feasible.

In the absence of structured educational experiences that give students exposure to the social ramifications of algorithms, faculty members might take it upon themselves to expose students to these problems. For example, a professor at Dartmouth teaching a course on AI included references to articles about algorithmic bias and its challenges.[29] It is a small gesture, but one that nonetheless may raise awareness to the discriminatory effects of AI.

4.4 Education in the ethics of algorithms beyond academia

Continuing this education beyond the boundaries of academia is imperative. Individuals in industry might not have exposure to ethical issues. This is unfortunate, since cases of algorithmic bias also come from algorithms created and utilized by industry. It is essential that professionals believe this is a problem worth addressing. We can encourage companies to include trainings and workshops similar to those given to students.

If makers of algorithms are aware of potential pitfalls they can try to find ways to correct them. For example, Google has recently implemented a crowdsourcing

feature for Google translate. It is a practical way for users to correct mistakes and biases the algorithm might make.[30] Although user input will not always be enough for true accountability, it is a step in the right direction. Informed professionals may be able to find ways to circumvent bias, whether it is through using representative data or asking for user inputs. Workshops or job trainings in industry can provide this awareness.

Besides the workshop at the University of Nevada, there are few records of such mandatory programs being implemented on a wider scale. In this sense, it is hard to predict or measure what impact such changes might have. Awareness is the first step to solving the problem, and an education in ethics can help create it. Future actions in this area include creating workshops and implementing curriculum changes.

5. Can policy changes make a difference?

Research, increasing diversity, and changing education are still not enough to solve all problems related to algorithmic bias. How can we control the use of algorithms by those who have authority and power? Since algorithms are in active use today, we need policies to define bias, and hold programmers and companies accountable for their work. For example, many complex algorithms are known as black boxes. These are predictive systems that utilize machine learning and make crucial decisions, but the public cannot see the code used to build them. This has led to calls for code transparency, so that we can see how the system came to a certain conclusion. However, this may not always be feasible since, as discussed, we do not always understand how the code behind such complex algorithms work.[14] While transparency may not always be enough, other policy solutions can allow us to enforce accountability.

In the case of housing discrimination in Texas, the legal doctrine of disparate impact theory allowed the courts to hold those who made the algorithm accountable.[10] We need legal and policy solutions like this that apply to all algorithms. New York City recently passed a bill that will result in the creation of a task force to monitor algorithms.[31] This is an important and groundbreaking step for policy based solutions that could be replicated by other cities.

Recent research has found that the use of copyright laws could reduce the occurrence of bias. A paper written by Amanda Levendowski suggests that principles of traditional fair use align with the goal of mitigating bias.[32] Even so, Levendowski still highlights the need for AI programmers, policymakers, and lawyers to define what is ethical for an algorithm to do. This ties back to the importance of discussing and teaching ethics.

6. Can we prevent algorithmic bias?

Testing could be used to prevent algorithmic bias. If we run "pre-release" trials of complex algorithms and AI systems, we might determine whether algorithms are causing bias or if the training data might cause problems. Companies can monitor the results of their algorithms even after they are released and used in different contexts or communities.[4]

Another more obvious prevention technique is to prepare training data more carefully. This ties back to our discussion of education. Providing education on ethics and algorithmic bias would create awareness of the problems associated with training data bias. If students and industry professionals know the effects of training data bias, we hope they will use good, representative training data. Understanding the potential ramifications of using "poor" datasets might motivate individuals to prepare training datasets more carefully. This could result in significant improvements in the short term.

7. Conclusion

Algorithms and artificial intelligence hold great potential, however algorithmic bias is an alarming problem that reinforces injustice. Allowing discrimination to occur and waiting on a future solution to appear is insufficient. This is not just a problem of technology, but a moral quandary at the societal level.

There are multiple fronts through which we can fight the unjust effects. We should expand work on long term research projects across different disciplines simultaneously, while

also pushing initiatives to increase diversity in tech related fields. We might be able to improve education by teaching students the social implications of algorithms. Including discussions of the ethics of algorithms in computer science curricula and in industry can have great potential. Whether this occurs through new course requirements, changes to existing courses, or workshops, learning about different problems of ethics in computer science is only in keeping with the times. It is also crucial that we create policy measures to deal with algorithmic bias.

These solutions are interdisciplinary and will involve more than just people who create algorithms. We will need the involvement of policy makers, lawyers, academic institutions and tech companies. This may seem like a formidable task, but we can begin by discussing the ethics of algorithms. If algorithms and AI are to be integral parts of our society, then it is our responsibility to make sure that we use them in a manner that is socially responsible.

References

1. Brogan J. What's the Deal With Algorithms? 2017. Available from: http://www.slate.com/articles/technology/future_tense/2016/02/what_is_an_algorithm_an_explainer.html.

2. Carthy J. WHAT IS ARTIFICIAL INTELLIGENCE? 2017 [Available from: http://www-formal.stanford.edu/jmc/whatisai/whatisai.html.

3. Lipton ZC. The Foundations of Algorithmic Bias 2017. Available from: http://approximatelycorrect.com/2016/11/07/the-foundations-of-algorithmic-bias/

4. Campolo A, Sanfilippo M, Whittaker M, Crawford K. AI Now 2017 Report. AI Now Institute; 2017.

5. Danks D, London AJ. editor Algorithmic Bias in Autonomous Systems. 26th International Joint Conference on Artificial Intelligence; 2017.

6. Angwin J, Larson J, Kirchner L, Mattu S. Machine Bias2016 2016-05-23. Available from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

7. Angwin J, Larson J. The Tiger Mom Tax: Asians Are Nearly Twice as Likely to… — ProPublica2015 2015-09-01. Available from: https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review

8. Spice B. Questioning the Fairness of Targeting Ads Online - News - Carnegie Mellon University2015 November 2017. Available from: http://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html

9. Guynn J. Google Photos labeled black people 'gorillas'2017. Available from: https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/

10. Kirchner L. When Discrimination Is Baked Into Algorithms 2017. Available from: http://www.theatlantic.com/business/archive/2015/09/discrimination-algorithms-disparate-impact/403969/

11. Crawford K. Opinion | Artificial Intelligence's White Guy Problem2016 20160625. Available from: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

12. Data & Society 2017 [Available from: https://datasociety.net.

13. AI Now Institute 2017 [Available from: https://ainowinstitute.org/.

14. Hudson L. Technology Is Biased Too. How Do We Fix It? | FiveThirtyEight2017 2017-07-21T12:20:53+00:00. Available from: https://fivethirtyeight.com/features/technology-is-biased-too-how-do-we-fix-it/.

15. Kirkpatrick K. Battling Algorithmic Bias. Communications of the ACM. 2017;59(10).

16. PERVADE - Pervasive Data Ethics for Computational Research 2017 [Available from: https://pervade.umd.edu/.

17. Yao M. Fighting Algorithmic Bias And Homogenous Thinking in A.I. 2017. Available from: https://www.forbes.com/sites/mariyayao/2017/05/01/dangers-algorithmic-bias-homogenous-thinking-ai/.

18. AJL -ALGORITHMIC JUSTICE LEAGUE 2017 [Available from: http://www.ajlunited.org/.

19. Science and Engineering Degrees, by Race/Ethnicity of Recipients: 2002–12 - NCSES - US National Science Foundation (NSF). 2017.

20. Landivar LC. Disparities in STEM Employment by Sex, Race,and Hispanic Origin. Washington, DC: United States Census Bureau; 2013.

21. Dickey MR. Google taps Van Jones and Anil Dash to discuss race and algorithmic bias2016 2016-12-12 15:46:35. Available from: http://social.techcrunch.com/2016/12/12/google-taps-van-jones-and-anil-dash-to-discuss-race-and-algorithmic-bias/.

22. Edionwe T. The fight against racist algorithms2017. Available from: https://theoutline.com/post/1571/the-fight-against-racist-algorithms.

23. Pretorius L, Barnard A, de Ridder C. Introducing computer ethics into the computing curriculum: two very different experiments 2017. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.128.9214&rep=rep1&type=pdf.

24. Barocas S, Bradley E, Honavar V, Provost, F. Big Data, Data Science, and Civil Rights. Computing Community Consortium. 2017.

25. Singer N. Tech's Ethical 'Dark Side': Harvard, Stanford and Others Want to Address It. Business Day [Internet]. 2018. Available from: https://www.nytimes.com/2018/02/12/business/computer-science-ethics-courses.html.

26. The Ethics and Governance of Artificial Intelligence – MIT Media Lab: MIT Media Lab; 2018 [Available from: https://www.media.mit.edu/courses/the-ethics-and-governance-of-artificial-intelligence/.

27. Syllabus for CS109: University of Texas at Austin; 2018 [Available from: https://www.cs.utexas.edu/~ans/classes/cs109/syllabus.html.

28. Trabia M, Longo JA, Wainscott S. Training Graduate Engineering Students in Ethics. 2016.

29. Palmer CC. CS89 Cognitive Computing with Watson 2017 [Available from: http://www.cs.dartmouth.edu/~ccpalmer/teaching/cs89/Course/CS89-Resources/index.html.

30. Lardinois F. Google Wants To Improve Its Translations Through Crowdsourcing. 2014 2014-07-25 13:53:55. Available from: http://social.techcrunch.com/2014/07/25/google-wants-to-improve-its-translations-through-crowdsourcing/.

31. Coldewey D. New York City moves to establish algorithm-monitoring task force. 2017 2017-12-12 16:43:19. Available from: http://social.techcrunch.com/2017/12/12/new-york-city-moves-to-establish-algorithm-monitoring-task-force/.

32. Levendowski A. How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem. Washington Law Review, Forthcoming. 2017.

Figures

Figure 1: Racial Representation in Overall Population vs. Computer Occupations

Note: Percentage of U.S. population is per 2010 U.S. Census, percentage of workforce is per 2011 data. Source: U.S. Census Bureau

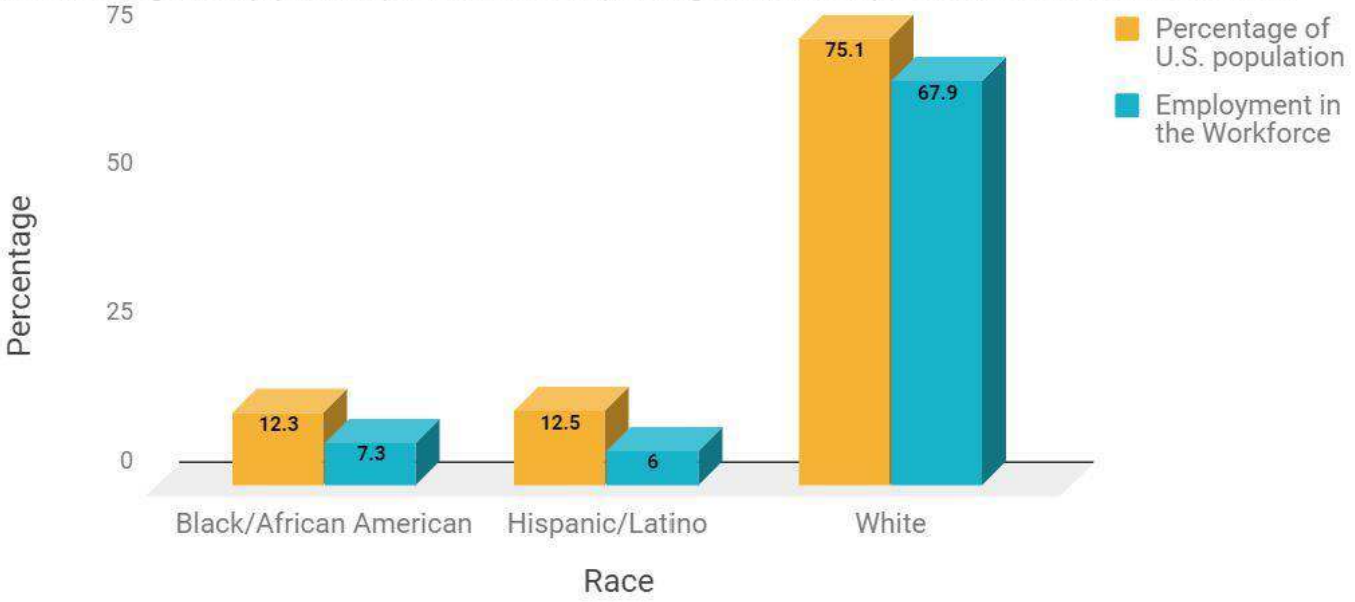

Figure 2: Gender Representation in Computer Occupations

Source: U.S. Census Bureau