

Algorithms Have Built Racial Bias in Legal System-Accept or Not?

Junkai Zhang^{1,*}, Yuxuan Han¹

¹Qingdao NO.58 high school, Shandong, China

¹University of Toronto, Toronto, Canada, M1C 1A4

*Corresponding author. Email: zhangjunkaiwill@126.com

ABSTRACT

Algorithms have been applied in various fields and increased efficiency in most workplaces. Correctional offender Management Profiling for Alternative Sanctions (COMPAS) and Level of Service Inventory-Revised (LSI-R) are the two algorithms that are used to do risk assessment in the legal system. However, both algorithms (COMPAS and LSI-R) have built-in racial bias, and the risk scores (the result of algorithms) are affected by racial bias, which finally affects judges' decisions because risk score is a factor that can influence defendants' sentencing and bail. This paper describes the mechanism of COMPAS and LSI-R, analyses the reasons why algorithms have built-in racial bias: developer embed their inherent racial bias into the algorithm, and many static factors that are included into consideration in algorithms can lead to racial bias; why built-in racial bias that appears from algorithm in the legal system is not acceptable; and the four methods of eliminating the racial bias in algorithms and make sure the algorithms are impartial with high efficiency.

Keywords: Algorithm, COMPAS, LSI-R, Racial Bias

1. INTRODUCTION

With the high efficiency of artificial intelligence and the improvement of its ability to deal with complex problems, the algorithm is gradually being applied in various fields. During the COVID-19 worldwide pandemic, artificial intelligence has used to track the spread of the virus and identify close contacts [1]. Also, contextual story-related content in computer games, such as animation and audio, involves the application of algorithms [2]. The use of artificial intelligence by human beings has been further developed in different fields, and the world is moving towards the era of artificial intelligence; several fields are benefited from the utilization of algorithms. Artificial intelligence can also be used in education; for example, many universities in China add algorithms to physical education to increase students' interest in physical exercises and improve teaching efficiency [3].

Although people benefited from the usage of Algorithms to a great extent, however, there are many issues when people rely on algorithms in decision making. One of the most common and severe problems of using the algorithm is because it can cause bias.

Since 2015, Amazon has been used machine learning engine to review resumes of job applicants, whereas this engine can cause gender bias. Amazon lets machine learns the resume from people who are recruited, analyze their characteristics, to sum up, the algorithm that used to review the applicant's resume. As a result, the system has concluded that male candidates are preferable to female candidates. Managers and developers are confused why this would happen. In all the data that the machine learned, the number of resumes for male employees was larger than the number of resumes for female employees, so the machine concluded that men are better suited to work in the technology industry. Although Amazon edited the algorithm so that it can be fairer, how the algorithm works are not only a "Black Box" to employer and employee but also the developer.

Improper use of algorithms, in an ignoble way, resulting in scandals are not uncommon. During the 2016 presidential election, Cambridge Analytica is accused of illegally obtaining information from nearly 50 million Facebook users, dividing them into different psychological profiles, and then selectively targeting them with political ads to change their choices. Under algorithms, users' preferences, beliefs, and interests can

be displayed [4]. Can we call it democracy when people who make choices might be influenced by the discriminated news that utility by algorithms they read?

In spite of the problems of applying algorithms with discrimination in employment and the 2016 U.S. election; the effect of algorithm discrimination in the legal system would be devastating. Among the most common exploitation of Artificial Intelligence in the field of legal systems in the United States would be focused on the use of algorithms in criminal procedure. In criminal prosecutions, algorithms are normally used to predict whether an individual is likely to re-offend, or whether he or she is likely to appear in court on the day of court, which can finally affect parole, sentencing, and bail; COMPAS and LSI-R are the algorithms that over 50% of the states in the United States [5]. Like algorithms in other fields, COMPAS and LSI-R are biased; both of these two algorithms have evidence that they have built-in racial bias. Even when racial bias occurs in other fields such as company and school is not acceptable, the racial bias that occurs in the legal system is definitely unacceptable; the judicial system is seen as impartial because it has public power, so any bias or unfairness can be seen as partial. If racial discrimination is treated as a “true verdict” by a justice system with public power, can it be called impartial deliberation? When applying algorithms in the legal system (under current technology) would cause racial bias, then lead to inequality between people; therefore, any algorithms that employed in criminal procedure that could or suspect that have built-in racial bias with current technology is not sufficient to be a trustworthy tool in the legal system. This article focus on the problems caused by algorithms that using in the legal system (COMPAS and LSI-R), which they have built-in racial bias and output the result (risk score) are biased; to prove that the COMPAS and LSI-R produce racial discrimination, this article analyzes the data that produced by algorithms and compares the results between different races; and also describes the causes of built-in racial bias and the solution of eliminating those bias.

Both COMPAS and LSI-R are risk assessment tools that assess the offenders’ relevant factors such as characteristics of the situation that offenders are facing, and output a final score to measure the likelihood the offender will re-offend or not; the reason why the United States employ algorithm in the legal system is that they hope to obtain more accurate risk assessments and lower detention rates [5]. However, risk assessment itself is not a new field in the legal system; this assessment has been a long time since it was dependent on “judges’ sense of justice, intuition and aptitude” [5], whereas the risk assessment depends mainly on algorithms. Before the conviction in the criminal system, whether individuals are going back to integrate the society or sentencing, these risk assessments

(COMPAS and LSI-R) play a vital role. Moreover, using these algorithms (COMPAS and LSI-R), judges will receive a level of risk of offenders, which helps judges make more appropriate decisions when decides to sentence [6]. Because of the unique nature of algorithm and software, the weight of each individual factors that contained in algorithms are undisclosed; as a result, it would be a “black box” to everyone despite the developer of the algorithms.

2. RACIAL BIAS

2.1. Racial Bias in LSI-R

Level of Service Inventory-Revised (LSI-R) is a third-generation risk assessment algorithm that was developed in the 1980s, and it is the most common tool used in the jurisdiction in the United States [7]. LSI-R contains fifty-four items (ten domains), from criminal history to alcohol or drug issues, from education history to family background [8]. The algorithm synthesizes the answer from offenders and provides different levels of risk that judges can refer to, which could affect judges’ decisions. On LSI-R its own, the algorithm should be a relatively useful tool compared to the traditional risk assessment tool, because it is more efficient and combines as many as factors possible.

However, there are many statistics show that LSI-R has built-in racial bias. Research shows that when dealing with racial minorities, LSI-R does not perform an adequate task; the result of LSI-R would have higher accuracy with White offenders than other non-White offenders; Caucasians and Hispanics would have fewer errors in classification from algorithm during the risk assessment [9]. Compared with the final score from LSI-R, African Americans are more likely to receive a higher score than White Americans; as a matter of fact, the actual risk of African Americans is usually disproportionate to White Americans [10]. According to the research by Fass et al., the author did a comparison between the result of African American participants and Caucasian participants, and Caucasian participants receive significantly lower LSI-R scores, compare to African American participants [9]. Thus, many statistics and research show that racial discrimination in the use of LSI-R is not a contingency, but rather a high probability.

2.2. Racial Bias in COMPAS

Similar to Level of Service Inventory-Revised (LSI-R), Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a fourth-generation risk assessment algorithm that introduced by Northpointe in the year 2000; 137 questions in COMPAS covers different areas of information of offenders, such as criminal history,

social history, and attitudes toward criminal [9]. For example, offenders' grades in school, criminal history of offenders' friends, or does offenders' friends have any history of use of the illegal drug [6]. One characteristic that is unlike LSI-R is that the result of COMPAS states three separate risk scores: "risk of violence", "risk of recidivism" and "risk of failure to appear" [9]. The result of "risk of violence" indicates whether the offender can commit violent crime afterward; the result of "risk of recidivism" obviously indicates whether the offender would re-offend or not in the future; finally the score of "risk of failure to appear" indicates whether the offender will show up to the trial or not [6].

Same as Level of Service Inventory-Revised (LSI-R), COMPAS also has built-in racial bias. ProPublica did a study comparing the result of COMPAS, and the results indicate that the score of "risk of recidivism" has racial bias, which is more discriminated against of African Americans and in favor of White Americans [4]. Compare to Caucasians and Hispanic offenders, African American offenders have a higher probability to receive the medium or high level of risk scores, and there are around 57.6% of African American offenders receive a high risk under COMPAS, whereas only 33.1% of Caucasians get a higher level of risk by using COMPAS; for the offenders who do not re-offend, the percentage of receiving high score for White Americans is only 22.01%, whereas Non-White Americans are 42.34%; also, black defendants have the lower possibility to get a low score than white defendants under COMPAS [6]. COMPAS not only has a built-in racial bias but there are also concerns about accuracy. Study shows that the accuracy of COMPAS is only about 66%, but COMPAS is an important factor influencing judges' decisions [6]. How do we accept low-accuracy algorithms in the court?

2.3. How Racial Bias Occurs?

It is important to find out why algorithms (COMPAS and LSI-R) have a built-in racial bias, and there are two main sources of the origin of racial discrimination: developer embed their inherent racial bias into the algorithm, and many of the factors that the algorithm considers about the defendants can themselves lead to racial bias. For the first source of racial bias from COMPAS and LSI-R, the developers themselves did not know they have an inherent racial bias, therefore, when they were designing algorithms, they brought that inherent bias into the algorithm, and the algorithm became biased. The more data a developer collects, the less biased the algorithm itself becomes; If developers design COMPAS and LSI-R with a relatively large amount of data from White Americans compared to Non-White Americans, the

algorithm would have higher accuracy of a risk score for White defendants than Non-White American. For example, race would not be supposed to be taken into account when designing the algorithms, however, if developers include race in the algorithms, different races will be given different scores, resulting in racial bias. Besides, if most or almost all of the developers are of the same race, then they would take less consideration of racial issues into account. For instance, if most of the developers are white Americans, then the whole algorithms are designed to stand at White Americans, not Non-White Americans.

The second source of racial bias in COMPAS and LSI-R would be static factors included in the algorithms, which would lead to racial bias. For all the factors that COMPAS and LSI-R taken into account, they can be classified as static and dynamic factors; dynamic factors would change, such as antisocial attitudes or relationships between friends and family; static factors are the factors that cannot be changed, such as defendants' race, work history, education and criminal history [8]. Dynamic factors are the factor that could change in the future, therefore might influence defendants the "risk of violence and recidivism". Static factors are the history of individuals, therefore cannot change; since they will not change in the future, then use the past history of defendants to predict their future action would have high possibility to contain racial bias. Since COMPAS and LSI-R do take static factors into account, then both tools would cause racial discrimination.

Among the static factors, there are some socioeconomic factors, such as the living address, employment, education, income, and debts. For example, individuals have higher education would receive relatively lower risk scores in COMPAS and LSI-R than individuals with no education; defendants with full-time work and high salaries would receive lower risk scores than those with no work and have liabilities. However, many defendants are underprivileged and suffering from socioeconomic marginality because of the racial bias in the current society, therefore they usually receive a higher risk score than privileged defendants [8]. Part of the reason they are underprivileged is not of their own making, but social problems. COMPAS and LSI-R consider the crime rate of the defendant's living address, this is because those rates can tell "risk of violence and recidivism"; but the rate itself may be biased, for example, more police in the district that have more African Americans compared to the district that have more white Americans, then white American district would have lower crime rate because; when these rate run in algorithms, algorithms would not take these background information into account, which finally leads to racial bias in COMPAS and LSI-R. Education is a factor that affects by defendants' families; if

defendants' parents are privileged, then they can have better financial management and afford defendants' tuition; whereas unprivileged parents have no choice but to let their children drop out of school; therefore, the socioeconomic marginal family usually do not have enough education, which finally has higher score when analyzed by algorithms. Furthermore, education can influence employment, which is further causing the built-in bias in algorithms. In the legal system, these factors actually were overrepresented [7]. To be more deeply, the problem is about prediction, because algorithms are based on all the past history of individuals and predict the future of defendants [11].

3. WHY ALGORITHMS RACIAL BIAS IS NOT ACCEPTABLE IN LEGAL SYSTEM?

Algorithms that are utilized in the court need to ensure that the tools are impartial, without any bias; if the tools that are used in the court are proofed that have racial bias, then is not acceptable. In the court, the COMPAS and LSI-R's risk score would affect by socioeconomic factors, then affect sentencing; defendants who hurt from built-in racial bias in algorithms would become more socioeconomic marginal, which finally lead to social inequality [7]. Under this inequality, individuals may hard to reintegrate into society, then cause crime rate may increase; the result goes the opposite direction of the purpose of employment of COMPAS and LSI-R in the legal system (obtain more accurate risk assessments and lower detention rates with the help of algorithms

Due to the special nature of algorithms, it is a "black box" for the public, because everyone does not know the accurate operation mechanism of COMPAS and LSI-R except the developer, which the court hears decisions produced by a mechanism of rules established by a small number of people. In 2016, the Wisconsin Supreme Court had applied COMPAS, and sentenced Eric Loomis for 11 years because he was charged with a drive-by shooting crime; however, there is no method to check the mechanism of how COMPAS and verify the manner [4].

4. MEANS OF SOLVING THE PROBLEMS OF COMPAS AND LSI-R BUILT IN RACIAL BIAS

To solve the problem of COMPAS and LSI-R have built-in racial bias, there are four ways: have a third party check how the algorithm works to make sure there is no bias; increase the diversity of developer; reduce the static factors in the algorithm as much as possible; and make the results of the algorithm less influential to the judge in the courtroom. Since algorithms may have many copyright issues, then show all the details and explain to the public would not be feasible, therefore, if

there is a third party that can check every logical manner of the algorithms, then they can avoid the occurrence of racial discrimination. Second, the more diverse race join in the design of the algorithm, the less unconscious racial bias in the algorithm. The third way is to reduce the static factors that COMPAS and LSI-R include, such as education and income, and this can avoid the overrepresentation problem. The last method is to weaken the importance of the risk score of COMPAS and LSI-R in the legal system, which judges can affect less by the built-in racial discrimination in algorithms.

5. CONCLUSION

Although artificial intelligence can increase efficiency and reduce error in most fields of work, the outcome of built-in bias and error would be devastating. It is common to see COMPAS and LSI-R in the legal system, and judges' decisions would affect by the risk score from these two algorithms. However, many statistics and evidences show that both COMPAS and LSI-R have built-in racial bias, which could lead to social inequality. Under algorithms, it is believed that algorithms remain high objectivity if there is built-in racial bias in the COMPAS and LSI-R, and the legal system still pretends they are high objective, then use of algorithms in the legal system will amplify the inequality in the whole society [11]. Therefore, the algorithm currently is not an ideal tool in the legal system, because there is built-in racial discrimination embed in the algorithm; if there is an algorithm that can ensure no erroneous and bias, then it would be a useful tool in the legal system. In the modern society, racial prejudice is still a serious problem. The suggestions to solve this kind of question are that: firstly, make friends in terms of their characteristics and personality instead of the nation they belong to or the skin that born with him. Secondly, cooperate with others more while people have their tasks to finish, find something in common between you and the person you meet, share your values or beliefs with him. Last but not the least, be confident with your nationality, stand on the side of your nation and support your ethnic nation forever.

AUTHORS' CONTRIBUTIONS

This paper is completed by Yuxuan Han and Junkai Zhang together.

ACKNOWLEDGMENTS

Several people encouraged me during the whole research. I would like to express my thanks to the supervisors, for their suggestions, and guidance. Also, I am extremely grateful to my parents, for both financial and mental support.

REFERENCES

- [1] Raju Vaishya, Mohd Javaid, Ibrahim Haleem Khan, Abid Haleem. Artificial Intelligence (AI) applications for COVID-19 pandemic [J] *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 2020, 14(4).
- [2] Pannu A. Artificial intelligence and its application in different areas [J] *Artificial Intelligence*, 2015, 4(10): 79-84.
- [3] Yu S. Application of artificial intelligence in physical education [J] *The International Journal of Electrical Engineering & Education*, 2021: 0020720921996604.
- [4] Katyal S K. Private accountability in the age of artificial intelligence [J] *UCLA L. Rev.*, 2019, 66: 54.
- [5] Liebman. Artificial Intelligence in Judicial Practice in the United States: Problems and Challenges [J] *China Law Review*, 2018(02):54-56.
- [6] Brackey A. Analysis of Racial Bias in Northpointe's COMPAS Algorithm [D] Tulane University School of Science and Engineering, 2019.
- [7] Van Eijk G. Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality [J] *Punishment & Society*, 2017, 19(4): 463-481.
- [8] Duwe G, Rocque M. A jack of all trades but a master of none? Evaluating the performance of the Level of Service Inventory–Revised (LSI-R) in the assessment of risk and need [J] *Corrections*, 2016, 1(2): 81-106.
- [9] Fass T L, Heilbrun K, DeMatteo D, et al. The LSI-R and the COMPAS: Validation data on two risk-needs tools [J] *Criminal Justice and Behavior*, 2008, 35(9): 1095-1108.
- [10] Lowder E M, Morrison M M, Kroner D G, et al. Racial bias and LSI-R assessments in probation sentencing and outcomes [J] *Criminal Justice and Behavior*, 2019, 46(2): 210-233.
- [11] Mayson S G. Bias in, bias out [J] *Yale LJ*, 2018, 128: 2218.