Algorithmic Bias: A New Age of Racism

Presented to the S. Daniel Abraham Honors Program
in Partial Fulfillment of the
Requirements for Completion of the Program

Stern College for Women
Yeshiva University
April 30, 2021

Shira Schneider

Mentor: Professor Alan Broder, Computer Science

Table of Contents

INTRODUCTION

As technology use has exponentially increased, computer scientists have adopted machine learning algorithms in an effort to teach computers to make decisions without having to specifically program instructions. These algorithms are fed training data and taught to execute certain tasks based on that data. They are useful as they are efficient and allow decision makers in private and public sectors as well as researchers to make smart and definitive, computer-made decisions. Kim Darrah, a British data journalist, says that these algorithms are popular because "they have a reputation for being infallible, neutral and fundamentally fairer than humans, and as a result are quickly making their way into both commercial and public decision-making." [1] The use for machine learning algorithms continues to rapidly increase, spanning from suggestions on social media platforms and product recommendations to speech recognition, medical diagnoses and criminal identification.

Although there are many positive aspects to these machine learning algorithms, there are also tremendous risks. This paper discusses different types of unintended biases that exist within machine learning algorithms and the negative consequences they have produced. In particular, this paper focuses on algorithms that have unconsciously been created with racial biases. Rather than removing prejudice from these fields, the algorithms reproduce it on a massive scale.

As this field continues to grow at rapid rates, it has become increasingly harder to control these algorithms and protect people from their repercussions before the algorithms' risks are completely understood. After analyzing these issues, it is evident that people must adequately educate themselves both in the technology and how it works but also in the sociological and psychological effects these algorithms can have.

MACHINE LEARNING ALGORITHMS

Algorithms are procedures that allow effective and definitive instructions to be carried out efficiently by a computer. Machine learning is a branch of artificial intelligence which focuses on algorithms that learn from a dataset and constantly adjust to provide better results. A resulting model is created when the algorithm runs, which represents the output learned from the algorithm. Models are generally divided into two categories: inference and prediction. Inference models help understand the world and uncover relationships. Predictive models attempt to estimate outcomes [2]. There are three main types of algorithms that produce such models: supervised learning, unsupervised learning and reinforcement learning. Supervised learning algorithms are presented with training data, containing pairs of inputs and their predicted outputs. The training data in these algorithms consists of labeled data, or data where each example is tagged with one or more labels identifying certain essential characteristics. The labeling of data is important in establishing a foundation for reliable learning patterns. A strong algorithm is one that has high accuracy and quality labels. High accuracy reflects how close the labeling is to ground truth and evaluates it against real world occurrences. Quality refers to the consistency across a dataset. The goal of supervised learning is for the algorithms to model relationships between the output and input from the given data sets, learning to calculate predicted output values for new data. This is in contrast to unsupervised machine learning where the computer is trained with unlabeled data. Unsupervised machine learning algorithms are used in detecting patterns by identifying rules in the input data. It is also used for descriptive modeling which provides insight into relationships between factors responsible for real world events. Reinforcement learning uses both labeled data and incoming data to constantly improve results

through experience. It uses and saves rewards signals based on each action so that it can take actions that would maximize the reward and minimize the risk [3].

These different types of algorithms all rely entirely on the data they are given. However, there are tools available for developers to adjust their datasets in an effort to change how the algorithm will predict future outcomes. This is useful when the training data is not representative of the population to which the algorithm is meant to be applied. The algorithm will be less accurate in predicting outcomes for data that is underrepresented in the training data because it tries to generalize from a smaller sample. One method to fix that error is oversampling. Oversampling allows the data modeler to replicate existing rare records to ensure those cases are more prominent. Another method is undersampling, which removes the common records, allowing the dataset to be more balanced. Both of these methods create a more evenly distributed dataset, but they alter the original dataset to train the model. Reweighting is an additional technique that does not require a new population. Rather than duplicating or removing information, weighting assigns certain weights to each record as a new variable in the training dataset. Small sample probabilities are assigned to overrepresented or standard records while a larger sample probability is assigned to non-standard records. This allows some records or features to be weighed more heavily than others, teaching the algorithm to more accurately predict outcomes for those records [4].

BIASES IN MACHINE LEARNING ALGORITHMS

Machine learning algorithms have rapidly become more common because they learn to improve over time without human intervention, they increase efficiency, and they mathematically make predictions and identify patterns. Additionally, many algorithms have been largely used under the guise that they are objective and unaffected by human biases. However,

there exists overwhelming data that these algorithms exist with underlying biases. Data is used to train machines for certain outcomes, but if the data chosen is skewed, the results will be biased. The program is only as accurate as the test data trains it to be. Cathy O'Neil, the founder of O'Neil Risk Consulting & Algorithmic Auditing, questions "whether we've eliminated human bias or simply camouflage it with technology" [5]. This false sense of neutrality most likely stems from the fact these algorithms are mathematical. However, O'Neil explains that "embedded within these models are a host of assumptions, some of them prejudicial" [6]. There are many different factors that can cause an algorithm's result to be biased. Three main biases that arise within algorithms are interaction bias, latent bias, and selection bias.

Interaction Bias

Interaction biases are developed as a result of how the user interacts with a particular algorithm. A system is created with a specific vision and the intent for it to learn based on what it experiences through its users. For example, Microsoft released an artificial chatter bot via Twitter in 2016 with the hope that it would learn to be conversational through its engagement with people. The problem, however, was that the system's interaction with humans resulted in it falling prey to human prejudices. While we assume robots to be impartial, Microsoft's bot, coined Tay, was taken down after 16 hours and 96,000 tweets due to the offensive nature of its tweets. It started posting racist and anti-Semitic tweets, learning from the people with whom it interacted [7]. Tay is one example of the problem with using human data to train algorithms; inherent human biases will emerge if developers allow this interaction to train their model. This may be an extreme example, but it proves the point. Machines can be taught to be biased based on the data to which they are exposed.

Latent Bias

A latent bias is a bias that implicitly mirrors a preexisting prejudice in society. It incorrectly identifies some factor as desired based on its training data and produces results accordingly. Many hiring algorithms have fallen prey to this bias. Amazon, for example, uncovered in 2015 that its recruiting algorithm was biased against women. The system was taught from a pool of resumes from the company over the previous ten years. The intention was for certain patterns to be determined based on those resumes in an effort to find the most fitting candidates. Due to the tech industry being male-dominated, however, the algorithm learned to view males as preferable and chose only male applicants. It reproduced the demographics of the existing workforce and excluded women [8]. Many other companies' recruiting algorithms have also faced this particular bias against women. But, the bias extends further than that, often targeting other minorities and reflecting other inequalities in various models.

Selection Bias

Selection biases occur when the training data for a particular algorithm is not reflective of society. It occurs when the data overrepresents one group over others. This bias has affected many different domains but mostly discriminates against minorities. Because many of these algorithms have been developed in the Western world, they are exposed to those norms. Therefore, algorithms are more likely to recognize Western-style wedding photos over African weddings for example. When given pictures, an algorithm might label a woman in a white dress and a man in a tuxedo as a bride and groom at their wedding when it might not correctly identify traditional African wedding attire [9]. Or, the algorithm could be exposed to pictures of heterosexual couples at their wedding receptions and would not be able to identify homosexual couples in similar photographs. This bias highlights that an algorithm can mirror the developer's own perception of societal norms. The data to which the system is exposed will shape the

algorithm's ability to make predictions and output correct results. Despite a machine being calculated and mathematical, it is not neutral due to the influence of the developer and the data they use to train it.

RACISM

There has been rising tension around racial issues in America as a demand for equality has yet to be met. Despite the Civil Rights Movement having fought and won equal rights for Blacks in the U.S., there is continued discrimination. While algorithms are thought to be unbiased, there do exist biases and many of those biases target Black people. Shalini Kantayya, director of the film *Coded Bias*, states that "we could essentially roll back 50 years of civil rights advances in the name of these machines being neutral when they're not" [10]. Though many of the developers have no malicious intent, historical biases and the implicit biases that they may hold are reflected in their code. Such prejudice becomes problematic because such algorithms have become widespread and are continually learning from their stream of data. These algorithms not only reinforce racist beliefs but can also be extremely detrimental and even dangerous to those whom it discriminates against as society becomes more reliant on algorithms.

I. FACIAL RECOGNITION

Facial analysis algorithms are one example of machine learning algorithms that have been discovered to discriminate against minorities and exhibit racist tendencies. The development of facial recognition technologies is typical of machine learning algorithms; they are given a set of training data and are taught to identify faces. First, the system is tested to be able to distinguish between faces and other objects in the image. Then, they are told to differentiate one person's face from another. As the program develops, it is able to process new photos and determine a name for the face from the faces already in its memory [11].

In order to achieve accurate results, the program must be presented with a diverse group of faces that are representative of society. However, research shows that the opposite is true. Training datasets have been dominated by lighter-skinned subjects, resulting in the inability of the machine to identify people with darker skin. Because the overall system is highly accurate, many times programmers initially do not notice this problem. Nevertheless, these concerns are pronounced when the error rates are analyzed along demographic lines [12].

Joy Buolamwini, a researcher at MIT Media Labs, uncovered that facial recognition code is biased against darker skinned people, particularly females. Buolamwini analyzed three commercial classification algorithms: Microsoft, IBM, Face++. The algorithms were tested using two published datasets known as IBJ-A and Adience, as well as a new Pilot Parliaments Benchmark (PPB). The IJB-A dataset, released in 2015, collected images of 500 subjects and Adience, released in 2014, tested 2284 faces. PPB, developed by Buolamwini and her team, was released in 2017 and consisted of 1270 individuals. While analysis of IJB-A and Adience indicated an overrepresentation of lighter males and an underrepresentation of darker females, PPB represented a more diverse representation of gender and skin-type. PPB consisted of 53.6 percent lighter-skinned individuals in comparison to the 79.6 percent in IJB-A and 86.2 percent in Adience. A more even distribution would presumably decrease the error rate in identifying dark-skinned subjects because the algorithm would have a larger sample size from which to generalize [13].

While the PPB dataset was able to identify faces more accurately than IBJ-A and Adience, the study found that all three classifiers, Microsoft, IBM and Face++, still overall perform better on light-skinned faces than darker-skinned ones (11.8% − 19.2% difference in error rate). Microsoft and IBM classifiers perform best on lighter male faces (error rates of 0.0%

and 0.3% respectively), yet Face++ classifiers perform best on darker male faces (0.7% error rate). Though the PPB was an improvement to the previously published datasets, this research still provides empirical support for racial biases in facial recognition software. This analysis helped prove the importance of inclusive dataset compositions and forced increased transparency and accountability for companies [14].

A further study was conducted by the ACLU in 2018 testing Amazon's facial recognition software, Rekognition. This technology was advertised as providing "highly accurate facial analysis, face comparison, and face search capabilities" [15]. The experiment tested Rekognition, comparing images of members of the House and Senate with a database of 25,000 publicly available mugshots. Amazon's system falsely matched twenty-eight members of Congress as other individuals who have previously been arrested for crimes. And among the incorrectly identified faces, about 40 percent were people of color despite the fact that they compose only 20 percent of Congress [16]. This test highlights the deep-rooted racism within these facial recognition algorithms. Some of the most well-respected Americans, but particularly those of color, were misidentified as criminals.

Google Photos

The reality is that facial recognition software exhibits unfairness. The biggest problem, however, with these results is that as these algorithms become more prevalent they continue to spread bias on a more massive scale and at a rapid pace. In 2015, Google was criticized for its facial recognition algorithm being racist because two African Americans were mislabeled as gorillas in its photo app. This tag was particularly harmful as Blacks have been called apes in the past as a derogatory term, implying that they were less evolved than other humans. Though Google quickly issued an apology, they still have yet to fully solve the issue. Google declared

that to avoid the issue, they would no longer use gorilla as a potential label in identifying their input data. They also removed similar words from their lexicon, such as apes, monkeys and primates. However, Google has not made strides in finding a longer-term solution even a few years later. Training a system properly is difficult, and it is even more difficult to control that system once it's alive. Therefore, companies such as Google must be more careful with their algorithms and ensure that their algorithms do not promote unintended biases [17].

Criminal Identification

Private companies are not the only ones utilizing facial recognition software. There has been an increase in the use of such algorithms by law enforcement as part of their investigative procedures. They have partnered with developers such as Google and Amazon to help track and identify suspects more efficiently. Many departments prefer facial recognition technology to DNA evidence as it does not require the suspect to have left DNA at the crime scene; it only requires an image of the suspect. This method has led to the successful capture of many criminals that were previously unable to be found. However, this method has also led to people, mainly people of color, being wrongfully accused of crimes. In Detroit in the summer of 2019, Michael Oliver and Robert Williams were identified inaccurately by a facial recognition algorithm. After running footage of the crime through the system, police arrested the men on the spot and in front of their family and neighbors. Upon further examination, it would have been clear that they were not the criminals. Both men, as well as another Black man outside Detroit, have sued the police on the basis of needlessly arresting them due to a false match. Detroit Police Chief James Craig admitted that "if we were just to use the technology by itself, to identify someone, I would say 96 percent of the time it would misidentify" [18]. The degree to which police should rely on these models should be questioned, particularly if the algorithm has a

preexisting racial bias. Though this technology has its benefits, it cannot be employed without restrictions. Some companies, such as Microsoft and IBM have started refusing to sell these technologies to the police because of potential harm [19].

Autonomous Cars

In another study that is not yet peer reviewed, the authors claim that autonomous vehicles use facial recognition software to detect objects. Using a large dataset consisting of pedestrians with different skin tones, they found a significant disparity; the algorithm had greater precision with light-skin types than darker-skin types. This experiment tested two datasets, MS COCO and BDD100K, for assessing person detection on various models. The preliminary results proved that there was an imbalance in detection across all models, so the researchers examined possible causes. This work tested to see if the discrepancy would be affected by time of day, occlusion or disproportionate distribution among the subjects in the training data. Different times of day present different levels of contrast between the subject and the background, which could have been a possible cause for the inequality. However, the data shows that there is no evidence that time of day is the source of the discrepancy. Blockage, another possible source of the predictive inequity, was also disproved as a responsible factor after the results were consistent with only unoccluded subjects. The third possible source of the gap between detection of light-skin types and dark-skin types was that the training data underrepresented certain groups of subjects. Because the data consists of more light-skin types, about 3.5 times more than dark-skin types, the results are skewed towards those subjects. The majority group has more examples, so the algorithm is able to learn to perform well on them; however, it struggles finding patterns in the minority classes as they are underrepresented. To counteract this imbalance, the researchers used a reweighting technique. They assigned a higher weight to the dark-skin types in an attempt to

give them more focus, so the algorithm can better learn to predict them correctly. The results show that the reweighting improved the detection of darker skinned pedestrians and that they were now close to the results of the detection of light skinned pedestrians in the unweighted examples. The analysis highlights that reweighting can better the performance of the dark-skin types without sacrificing the functioning of the algorithm on the light-skin types. In fact, the detection of the light-skin types also increased under the new weighting. These findings suggest that the original bias stems from the inadequate and disproportionate data given in the training data but that there are ways to correct that bias [20].

There are many benefits to facial recognition technologies. They allow for greater levels of efficiency, better safety and are able to perform tasks that people could not. However, they are not neutral, and these studies show that believing that they are unbiased can be harmful. These algorithms have not been trained and tested adequately to recognize dark-skin types. These algorithms have been let out into the world with systematic discrimination against and dangers to Blacks and other minorities. However, these studies have also proven that there are fixes. These algorithms are only as strong as their training data. A more diverse training dataset and reweighting data are two ways to improve results, reduce racism and even potentially save lives.

II. RISK ASSESSMENT INSTRUMENTS (RAIs)

Algorithms are quickly becoming widespread across the criminal justice system. Facial recognition is being used to identify suspects, as mentioned above. But beyond that, PredPol and HunchLab are two technologies that estimate the most likely location of a specific crime over a period of time. PredPol uses crime type, crime location and crime date and time to help reduce crime rates and manage resources more effectively [21]. HunchLab is similar but more extensive as it also incorporates the use of non-crime data as variables because it believes that predictions

will be skewed if they are based only on crime related datasets. For example, it includes data regarding vaccines, weather, social events and school calendars [22]. Patternizr is another algorithm that recognizes patterns in the police database, helping detectives quickly explore a broad pool of data and find related crimes during an investigation [23]. District attorneys have also turned to predictive technology to discern high-risk cases. And, risk assessment instruments have been developed to predict a defendant's future risk for misbehavior.

Risk assessment instruments (RAIs) have been developed in an effort to increase objectivity and fairness in the judicial system and reduce mass incarceration. They allow for the judges and prosecutors to defer to these machines in order to make unbiased decisions and have been adapted to assist through various stages of the judicial process. This type of software is being utilized during the pretrial phase, for sentencing decisions, probation and parole requirements. We assume them to be less fallible or predisposed than a subjective human judge; they have the potential to be consistent, accurate and bring transparency to these life-altering decisions made in court. However, studies show that they exhibit racial biases. Rather than using their power to reform our society, they have further entrenched the racial stratification because of the way they are trained [24].

Historical Background

The criminal justice system has historically targeted Black people over white people. The percentage of Blacks in Alabama prisons rose from two percent to 70 percent between 1850 and 1870. The rates were so high because Black men were convicted of various crimes so that they could be put to work as cheap labor. This leasing of convicts persisted through the 1940s, adding to the increased number of Blacks in the prison system. And despite the Civil Rights movement in the 1960s, criminalization of Blacks continued to increase. From the 1970s through the 1990s,

low-income Black communities had increased arrests and control by the police. The War on Drugs, especially under President Reagan, exacerbated mass incarceration targeting Blacks. Blacks have been convicted for marijuana possession at significantly higher rates than other groups though they have been estimated to consume marijuana at approximately the same rates [25]. Though the imprisonment rate for Black Americans has finally been declining in recent years (34 percent decline since 2006) [26], there still exists a significant imbalance in the prison system. Even today, one in three Black men is expected to be incarcerated during their lifetime [27]. And a University of Maryland analysis found that prosecutors in Harris County, the largest country in Texas, are three times more likely to suggest the death penalty for African Americans than for whites convicted of the same crimes [28]. Racism has been and continues to be prevalent, resulting in an unequal distribution in the criminal justice system.

Bias Risk Assessment Instruments (RAIs)

The reality that there is a historical imbalance in the prison system is one that cannot be easily corrected. And, it is highly difficult to remove such biases because of the legacy of unequal criminal justice practices. However, a well-produced algorithm could possibly prevent this from continuing given that it is an objective. Unfortunately, many existing systems have been created with racial discrimination because they are based on the historical practice. If the training data of a RAI algorithm is based on historical data, it could implicitly learn to give Black people higher risk scores. The model might reproduce the unfair realities. For example, a model might rate Blacks as higher risks for possession of marijuana simply because their race has historically been convicted for marijuana possession at higher rates. Despite the rates of marijuana possession being roughly equal among race groups, Blacks have been convicted more frequently [29]. This latent bias is developed from existing prejudices in society, specifically the

racism embedded in American history and the prison system in particular. Rather than minimizing the racial issues in the judicial system, such algorithms can exacerbate the disparities, putting more Blacks in prison for longer [30].

In addition to latent biases in RAIs, they can also contain selection biases. These biases in RAIs can be derived from the fact that the diversity of a population and rate of recidivism varies throughout the country and even across different counties. So, a model can be trained on data that does not reflect the population to which it is applied. For example, the Ohio Risk Assessment System (ORAS) has been adopted nationwide but is trained on data of 1800 prisoners in Ohio. Therefore, the RAI could produce inaccurate predictions if the distribution of defendants in other counties differ from those in Ohio. This method does not consider the differences between jurisdictions and customize for local populations [31]. Studies on the development and validity of the ORAS argue that although the data was collected from offenders across Ohio, it is unwise to assume that results reflect the tendencies of individual counties in Ohio. They question whether these results can be applied broadly in Ohio, let alone nationwide [32]. Algorithms that make such impactful decisions must be further evaluated and more localized to ensure that the outcomes are effective and neutral.

The COMPAS Assessment

The use of these algorithms has been largely contested due to concerns about the influence of these biases on the algorithms, in addition to other objections. In *State v. Loomis*, a Wisconsin Supreme Court case, Mr. Loomis challenged the use of the COMPAS RAI as part of the criminal justice process. In 2013, Loomis pleaded guilty to having driven a stolen car and fleeing from police. The COMPAS assessment identified him as an individual who was at high risk to the community, and he was sentenced to eight and a half years in prison. Loomis

protested the use of the score as a breach of his due process rights. The state ruled in favor of the judge, claiming that the use of the score can be considered in addition to other factors. And, some research has also argued that such RAIs do, in fact, have roughly accurate results [33]. Northpointe, the for-profit company that created COMPAS, insists their algorithm does not have racial biases. The score COMPAS generates is based on 137 questions that the defendant answers or are pulled from records, and race is not one of them. The founder, Brennan, and two colleagues conducted a study in 2009 testing the validity of their product. With a sample of 2,328 people, they found that the score was accurate 68 percent of the time. And, the study also found that the score was only slightly less accurate for black males (67 percent) than for white males (69 percent) [34]. Another RAI was tested in a 2016 study by Skeem and Lowenkamp which found that Blacks did, on average, receive higher scores but that the disparity was not on account of bias. A 2012 statistical study done by New York City also analyzed the effectiveness of RAIs. They tested more than 16,000 probationers and found that the system was 71 percent effective. However, these studies did not specifically evaluate for racial differences [35].

In reaction to Northpointe's study, ProPublica analyzed the COMPAS algorithm evaluating whether it was biased against particular groups and found that it was both racially biased and inaccurate. COMPAS scores range from one to ten, where ten was the highest risk. Scores under four are labeled by COMPAS as low and scored above eight are labeled as high. The subjects in the study were composed of more than 10,000 criminal defendants in Broward County, Florida. They were observed over a two-year period to test whether their recidivism rate was consistent with their predicted scores. They tested both general recidivism and violent recidivism rates, and they found that overall the score was accurate 61 percent of the time for general recidivism but only correct 20 percent of the time with violent recidivism. When they

evaluated specifically for race, ProPublica found that the distribution of the COMPAS scores for black and white defendants differed significantly. In the general recidivism data, the majority of white defendants have scores below five with a few outliers having higher scores. The black defendants on the other hand, had an even distribution across scores. The violent recidivism model shows a more equal distribution for Blacks, but there were still racial disparities. Their research shows that Blacks are more likely to be predicted as high-risk while white defendants were more likely to be given lower scores [36].

ProPublica also produced a logistic regression model based on COMPAS to test for further inequalities by specific variables. In the general recidivism model, age was found to be the greatest factor in predicting a high-risk score. People younger than 25 years old were expected to get higher scores 2.5 times more often than middle aged offenders. However, race also proved to be a determining factor. Though Black defendants originally had higher scores, they were still 45 percent more likely to get a higher recidivism rate than whites when the data was controlled for other factors. The violent risk model found age to be an even more influential indicator as young defendants were 6.4 times more likely to receive higher scores. In this model, race was also a stronger factor as the findings showed that it was 77.3 percent more probable for black offenders to get a higher score. ProPublica's results of this study, controlled for specific variables, show empirical evidence that Black offenders were more likely to be labeled as high risk but not re-offend, and white defendants were more often labeled as low risk and did re-offend. Blacks were twice as likely as white recidivists to be misclassified as a higher risk of violent recidivism while white defenders were misclassified as low risk 63.2 percent more often than Blacks [37].

It is difficult to separate the data from the historical biases. There has been so much racism in the history of our country, specifically in the criminal justice system. It is dangerous to think that these algorithms, trained on historical criminal databases, are reducing imprisonment and acting as objective alternatives. People who are real threats could receive low scores and walk free, and innocent people might receive higher scores and be wrongly accused or given too severe of a punishment. This system is flawed, and people's lives, and liberty are at stake. However, this can be corrected. A 2016 article in *The Washington Post* commented on the debate between Northpointe and ProPublica saying,

> "Algorithms have the potential to dramatically improve the efficiency and equity of consequential decisions, but their use also prompts complex ethical and scientific questions…. We must continue to investigate and debate these issues as algorithms play an increasingly prominent role in the criminal justice system" [38].

Effective RAIs are appealing because they can make criminal sentences more consistent and less likely to be swayed by the prejudices of a judge. Additionally, they save money by minimizing the length of the average sentence. Nevertheless, developers must hold themselves accountable and ensure that the RAIs they create are neutral and fair. These algorithms should be more localized systems which use targeted training data and do adequate testing to verify that biases do not influence the algorithms. The RAIs which currently exist perpetuate racial disparities rather than fix them.

III. HEALTH CARE

Many health systems rely on predictive algorithms to identify patients with complicated health needs. Applying such algorithms to health care provides hope for unbiased diagnoses and treatments. Machine learning algorithms have the potential to impartially interpret data in

medical records and assist health care providers in the clinical decision process. They are able to review all data in electronic records which would be tedious and expensive for humans to do. Hospitals and insurers use algorithms to aid in administering care for about 200 million people in America every year. These algorithms are trusted to discern patients who need "high-risk care management" programs which provide additional resources and greater attention to the patient. However, studies have shown that many existing predictive models reproduce racial disparities in their efforts to identify who will derive the most benefit from the programs.

Racial Discrepancy in Health Care

Studies show that racial and ethnic minorities do not have equivalent care relative to other groups in the United States. The causes for these disparities range from income, education and socioeconomic status to public policy and societal norms. Blacks generally have lower levels of health insurance coverage. They are more likely to lack continuity in health care, have public health insurance coverage and receive their care in nonoptimal organizational settings. And, a 2000 study demonstrates that between 1977 and 1996 this imbalance between Blacks and whites has not lessened over time [39]. Beyond healthcare coverage, minorities also face challenges in regard to the quality of their care. Research suggests that implicit bias plays a role in health care disparities. A 2018 study tested this utilizing the Implicit Association Test, and the results imply that most health care providers across all fields and levels exhibit implicit biases against Blacks [40]. Additionally, research indicates that minority patients are more likely to be treated by less proficient physicians and that the treatment of racial and ethnic groups differs even within a facility. Evidence also reflects that Black patients are more likely to have their pain underestimated and be undermedicated for their pain in comparison to white patients. Furthermore, access to better healthcare differs by residential area based on population. For

example, areas with larger Black populations tend to provide worse treatment to myocardial infarction patients of all races and ethnicities than areas with small percentages of Blacks [41]. Scientists speculate that this differential treatment and discrimination has led to mistrust of the healthcare system. Even today, Black Americans stand out as the racial and ethnic group least inclined to get vaccinated for COVID-19. Only 42 percent of Black Americans said that they would be willing to take the vaccine in November 2020, compared to 63 percent Hispanic and 61 percent white adults [42].

Bias Algorithms

A study by Ziad Obermeyer, who researched machine learning and health care management at the University of California, dissected the racial bias in algorithms that manage health care and found that the use of the cost of care as a label created a bias because of the reality that Black patients overall spend less on their care. The design of these algorithms assumes that those with the greatest health needs will also have the greatest total medical expenditures in a given year. However, the use of this label skews the results because it does not reflect the reality that the model should predict. Obermeyer and his colleagues were given access to the inputs, outputs and eventual outcomes of the algorithm and were able to thoroughly understand it and isolate issues. All primary care patients from 2013 to 2015 within a large academic hospital were identified and categorized based on race. The main sample consisted of 6,079 Black patients and 43,539 white patients. The overall health status of each patient was determined, and they were separated by race based on the risk score given by the algorithm. Depending on a given score, individuals were offered to participate in a program that provides extra care and resources. Results show that at the same score, Blacks had significantly more illnesses than whites. At the 97th percentile (the score at which patients are automatically

selected for the program), Blacks displayed 26.3 percent more chronic illnesses than whites. A simulated scenario was created to remove the bias by replacing healthier white patients above a specific threshold with less healthy Black patients below that threshold until the marginal patient was equally healthy. This change increased the percentage of Black patients for all thresholds above the 50th percentile, and at the 97th percentile, the fraction of Black patients rose from 17.7 to 46.5 percent [43].

Their results underscore the importance of finding a suitable label on which the algorithm can be trained. Cost prediction is a common approach, being the accuracy metric for the current ten most widely used health care algorithms. However, the data reveals that this method creates bias. Therefore, Obermeyer and his colleagues developed three new predictive models, two of which tested new labels. They studied total cost in a specific year, avoidable cost in a specific year and a measure of health in a specific year (the number of active chronic illnesses). The percentage of Black patients at or above the 97th percentile was 14.1 for total cost, 21.0 for avoidable cost and 26.7 percent for the active chronic conditions' predictor. The greatest difference in composition of Black patients was nearly doubled. The results demonstrate that the choice of label has a strong impact on the outcome of the algorithm in terms of bias. The researchers worked together with the algorithm manufacturers to develop a better model. They designed a system where the label combined health prediction with cost prediction which reduced bias by 84 percent. Though using total cost is a seemingly reasonable choice, it still produced a large bias and impacted people's access to better care. Changing the labels chosen and data fed to algorithms is key in using predictive models effectively. Careful consideration is required to reap the benefits of the algorithm while also avoiding risk [44].

A similar study was done by IBM Research on a publicly available and widely used dataset produced by the U.S. Department of Health and Human Services, Medical Expenditure Panel Survey (MEPS). Research found that only 6.8 percent of Blacks are expected to be in the top ten percent of the highest total expenditure for the year compared to the 10.6 percent of whites. The analysis shows that Blacks are underrepresented in the high-risk population, and they have to be sicker than whites to be included. The results of this study mirrored Obermeyer's in that Blacks at a given risk level are typically sicker than whites at the same level and are therefore less likely to be recommended for care management. To mitigate the bias, the experimenters applied a reweighting technique to the training data. Unlike Oberemeyer, this method reassigned weights to pairs in the training data without altering the labels. This adjustment raised the percent of Blacks predicted to be high-expense to 11 percent and whites to 9.9 percent. The new percentages are more consistent and fairer as the Blacks in the group are sicker than the whites on almost all metrics measured. Reweighting allowed the algorithm to account properly for the underrepresented Blacks in the dataset, giving them more equal access to advanced health care programs [45].

A third study, published in JAMA Internal Medicine in 2018, analyzes the potential biases in using electronic health record data for algorithms. The work questions the method of data collection and injustices that can arise from it. Firstly, the use of electronic data limits the algorithm because data can be missing or not accessible. This error can create a bias as the algorithm cannot assign an outcome to people whose data is missing, which is predominantly those with low socioeconomic status. Patients with lower incomes may receive fewer tests, visit several different health care facilities or have limited access to healthcare. Therefore, systems trained at university hospitals tend to document white individuals with higher income and may

not be applicable to lower socioeconomic classes and minorities with incomplete data. And even with complete data for minority groups, they are still vastly underrepresented in the dataset. This can cause the machine learning algorithm to provide biased estimates as it is trying to predict outcomes based on a small sample. The algorithm could learn to treat the majority group with preferential treatment, and low samples sizes and underestimation of minority groups could be misinterpreted as lower disease burden. Finally, studies show that patients of different races and ethnicities receive differential care. For example, women are less likely to be given lipid-lowering medications despite being more likely to present with hypertension and heart failure. Because algorithms learn from the data they inherit even if that data is biased, the algorithm might begin to reflect the health care providers bias and misclassify patients accordingly [46]. This study highlights how easily unintended prejudices can arise in machine learning algorithms and the importance of considering minorities in designing such algorithms.

Studies on various health care algorithms show that predicting the cost of care as a proxy for health needs ultimately creates a bias against minorities as providers spend less on their care overall. These results call for a more socially conscious approach to developing technology. A seemingly reasonable label used in an algorithm, such as health cost, could potentially magnify pre-existing disparities in society and even produce life-threatening results. While historically the fight against racial injustice was more explicit, it is now encoded by designers who are simply not attentive to systemic racism. However, these biases and complications within machine learning models are all manageable once they are identified. Efforts to correct them must be made to ensure accurate and fair results across populations and that all people receive equal access to health care.

SOLUTIONS

The rate at which machine learning algorithms are becoming so widely used has been rapid. Algorithms are undoubtedly useful as they easily identify trends and patterns, are continually improving, and are more efficient than humans. Furthermore, algorithms can be widely applied to different areas; they have been implemented to assist with finances, healthcare, marketing and the government, among other areas. These technologies have quickly become embedded in everyday life and developers haven't had a chance to understand if they are as harmful as they are helpful. Though algorithms are innovative, it is unclear how well they truly serve their purposes and how well they serve society. This paper has only begun to showcase the biases that exist within machine learning algorithms. Beyond racial biases, many other groups are the target of unjust algorithms, such as women, Hispanics and people of lower socioeconomic status. But as studies have begun to further analyze this problem, researchers have identified some solutions.

Diversity

A primary solution to the injustice rooted in various machine learning algorithms is more inclusivity. Training data that is more heterogeneous allows the algorithm to work effectively on its task in determining a certain outcome without making prejudiced decisions. More diverse datasets would prevent an algorithm from accidentally learning that a majority group is preferable to a minority group or having high error rates in underrepresented groups. Stronger datasets would be more common with a more diverse group of developers. Studies show an improving trend, but minorities remain underrepresented in computing jobs. In 2017, only 24.4 percent of people employed in computing occupations were women, 8.3 percent were Blacks and 7.2 percent were Hispanic. Though these are improvements from the previous few years, there is an obvious imbalance in the technology sector [47]. This gap introduces more frequent selection

biases as designers naturally represent themselves or their view of society in their projects. If white males are predominantly creating algorithms, white males will be represented as the majority. And, many such datasets have become adopted as standard training data which perpetuates the bias. Furthermore, latent biases can also be corrected through diversity because the more aware developers are of the struggles of minorities, the more they will understand how to ensure the algorithm produces fair outcomes. Unfortunately, software has inadvertently incorporated biases as a result of the experiences of their developers. The diversity of the datasets is imperative and that can come about with a more diverse group of programmers.

Weights

Collecting better training data is not always feasible, and there are alternative solutions to mitigating bias. One technique is reweighting the inputs, similar to the solutions tested in the study of the disparities between light-skin and dark-skin individuals in the facial recognition technology of autonomous vehicles and IBM's research on the racial biases in health care. For the algorithms discussed in this paper, this method could assign a stronger weight to Blacks in certain facial recognition and health care datasets as they are generally underrepresented. On the other hand, RAI datasets could be overpopulated with Black records, so those records might be assigned a lower weight in order to balance the distribution and reduce the bias.

New Technology

In addition to solutions that an individual developer or data modeler can perform, various companies have made efforts to provide services that will recognize racial biases in machine learning models. IBM, for example, has launched an open-source software toolkit, AI Fairness 360 toolkit (AIF360), aimed at spotting and removing bias. It allows customers to see how their algorithm makes decisions and which factors are being used to determine the outputs. They also

added a “check fairness” button to examine the correlation of a system’s results with specific attributes such as ethnicity or ZIP code [48]. Similarly, Microsoft developed the Explainability Boosting Machine (EBM) which is intended to reduce bias through producing explanations to the algorithm's decisions. Google also announced they are planning to release new AI ethics services that give advice on detecting racial biases and establishing other ethical guidelines [49]. Companies have begun to take on more accountability and have started taking steps towards creating more impartial algorithms.

Laws

Because this technology is new and growing quickly, judges have been hard pressed to respond to ethical concerns and adapt to the reality of such technologies. For example, the Fourth Amendment protects citizens from unreasonable arrest. Given the fact that many facial recognition software has been shown to be biased and unreliable, perhaps the Fourth Amendment should prevent police from using such technology for identification of suspects [50]. U.S. lawmakers are also taking further steps to regulate the use of these algorithms and mitigate the effects of their bias nature. In 2019, policymakers released the Algorithm Accountability Act. It requires big companies to audit their algorithms for discrimination and correct them accordingly. It is the first step in Congress addressing new technological realities [51].

CONCLUSION

This paper surveyed multiple manifestations of algorithm bias in different areas and highlighted the impact of incorporated bias. Facial recognition technology is used for a range of purposes, including photo tagging, identifying suspects and even autonomous vehicles. Risk assessment instruments are used throughout the criminal justice process to determine decisions such as length of sentencing and probation. Health care algorithms are used to distinguish sicker

individuals to offer additional care and resources. The studies shown in this paper suggest that there are biases embedded in each of these algorithms. Because of their wide scope, racism has been reinforced on a massive scale. Algorithms are intended to make life easier, but these types of algorithms are concerning because they amplify systemic racism. Society should question how much we should rely on such technology until better solutions are found and regulations are put into place.

Endnotes

[1] Kim Darrah. 2017. The troubling influence algorithms have on how we make decisions. (November 2017). Retrieved April 1, 2021 from https://www.theneweconomy.com/technology/the-troubling-influence-algorithms-have-on-how-we-make-decisions

[2] Eric Hart. 2021. Machine Learning 101: The What, Why, and How of Weighting. Retrieved April 29, 2021 from https://www.kdnuggets.com/2019/11/machine-learning-what-why-how-weighting.html

[3] Cloudfactory. The Ultimate Guide to Data Labeling for Machine Learning. Retrieved April 29, 2021 from https://www.cloudfactory.com/data-labeling-guide#:~:text=While%20the%20terms%20are%20often,consistent%20with%20real%2Dworld%20conditions.

[4] Eric Hart. 2021. Machine Learning 101: The What, Why, and How of Weighting. Retrieved April 29, 2021 from https://www.kdnuggets.com/2019/11/machine-learning-what-why-how-weighting.html

[5-6] Kathy O'Neil. 2016. *Weapons of Math Destruction.* Crown Publishing Group, New York, NY, 29.

[7] James Vincent. 2016. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. (March 2016). Retrieved April 9, 2021 from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

[8] Rachel Goodman. 2018. Why Amazon's Automated Hire Tool Discriminated Against Women. (October 2018). Retrieved April 29, 2021 from https://www.aclu.org/blog/womens-rights/womens-rights-workplace/why-amazons-automated-hiring-tool-discriminated-against

[9] Annie Brown. 2020. Biased Algorithms Learn From Biased Data: 3 Kinds Biases Found In AI Datasets. (February 2020). Retrieved January 10, 2021 from https://www.forbes.com/sites/cognitiveworld/2020/02/07/biased-algorithms/?sh=3e0a4fc776fc

[10] Shalini Kantayya. 2020. Your undivided attention: Bonus - Coded Bias. Podcast. (8 April 2021). Retrieved April 9, 2021 from https://www.humanetech.com/podcast/bonus-coded-bias

[11-12] Peter N.K. Schuetz. 2021. Fly in the Face of Bias: Algorithmic Bias in Law Enforcement's Facial Recognition technology and the Need for an Adaptive Legal Framework. *Minnesota Journal of Law & Inequality* 39, 1, Article 8 (February 2021), 225-229.

[13-14] Joy Buolamwini, Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR* 81 (2018), 77-91.

[15] Amazon Web Services. 2021. What is Amazon Rekognition. (2021). Retrieved April 14, 2021 from https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html

[16] Jacob Snow. 2018. Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots. (July 2018). Retrieved April 13, 2021 from https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

[17] Tom Simonite. 2018. When It Comes to Gorillas, Google Photos Remains Blind. (January 2018). Retrieved April 14, 2021 from https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

[18-19] Peter N.K. Schuetz. 2021. Fly in the Face of Bias: Algorithmic Bias in Law Enforcement's Facial Recognition technology and the Need for an Adaptive Legal Framework. *Minnesota Journal of Law & Inequality* 39, 1, Article 8 (February 2021), 229-235.

[20] Benjamin Wilson, Judy Hoffman, Jamie Morgenstern. 2019. Predictive Inequity in Object Detection. arXiv: 1902.11097v1. Retrieved from https://arxiv.org/pdf/1902.11097.pdf

[21] PredPol. 2020. Predictive Policing Blog. Retrieved April 29, 2021 from https://www.predpol.com/

[22] GitBook. Predictive Policing. Retrieved April 29, 2021 from https://teamupturn.gitbooks.io/predictive-policing/content/systems/hunchlab.html

[23] Andrew Liptak. 2019. The NYPD is using a new pattern recognition system to help solve crimes. (March 2019). Retrieved April 20, 2021 from https://www.theverge.com/2019/3/10/18259060/new-york-city-police-department-patternizer-data-analysis-crime

[24] Alex Chilas-Wood. 2020. Understanding risk assessment instruments in criminal justice. (June 2020). Retrieved April 16, 2021 from https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/#footnote-16

[25] Georgetown University Law Library. 2021. A Brief History of Civil Rights in the United States. (April 2021). Retrieved April 20, 2021 from https://guides.ll.georgetown.edu/c.php?g=592919&p=4172706

[26] John Gramlich. 2020. Black Imprisonment rate in the U.S. has fallen by a third since 2006. (May 2020). Retrieved April 20, 2021 from https://www.pewresearch.org/fact-tank/2020/05/06/share-of-black-white-hispanic-americans-in-prison-2018-vs-2006/

[27] The Sentencing Project. 2020. Criminal Justice Facts. (2020). Retrieved April 21, 2021 from https://www.sentencingproject.org/criminal-justice-facts/

[28] Kathy O'Neil. 2016. *Weapons of Math Destruction.* Crown Publishing Group, New York, NY, 29.

[29] The Sentencing Project. 2020. Criminal Justice Facts. (2020). Retrieved April 21, 2021 from https://www.sentencingproject.org/criminal-justice-facts/

[30] Alex Chilas-Wood. 2020. Understanding risk assessment instruments in criminal justice. (June 2020). Retrieved April 16, 2021 from https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/#footnote-16

[31] Sharad Goel, Ravi Shroff, Jennifer Skeem, Christopher Slobogin. 2018. The Accuracy, Equity and Jurisprudence of Criminal Risk Assessment. (December 2018).

[32] Edward J. Latessa, Richard Lemke, Matthew Makarios, Paula Smith, Christopher T. Lowenkemp. 2010. The Creation and Validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, 71, 1. (June 2010).

[33] Cindy Anderson. 2020. Risk Assessment Instruments Are Inappropriate for Sentence Reform: Real solutions for reform address racial stratification. *Georgetown Journal of Law & Modern Critical Race Perspectives* 12, no 2 (Fall 2020), 193-194.

[34-35] Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirschner. (2016). Machine Bais. (May 2016). Retrieved April 21, 2021 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[36] Alex. 2018. Racial Bias and Gender Bias Examples in AI systems. (September 2018).Retrieved April 20, 2021 from https://medium.com/thoughts-and-reflections/racial-bias-and-gender-bias-examples-in-ai-systems-7211e4c166a1

[37] Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirschner. (2016). How We Analyzed the COMPAS Recidivism Algorithm. (May 2016). Retrieved April 21, 2021 from https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[38] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel. 2016. A computer program used for bail and sentencing decisions biased against blacks. It's actually not that clear. (October 2016) Retrieved April 23, 2021 from https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/

[39] National Research Council. 2004. Understanding Racial and Ethnic Differences in Health in Late Life: A Research Agenda. (2004). Retrieved April 25, 2021 from https://www.ncbi.nlm.nih.gov/books/NBK24693/

[40] Ivy W. Maina, Tanisha D. Belton, Sara Ginzberg, Ajit Aingh. 2017. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science & Medicine*, 199, (2018), 219-229.

[41] National Research Council. 2004. Understanding Racial and Ethnic Differences in Health in Late Life: A Research Agenda. (2004). Retrieved April 25, 2021 from https://www.ncbi.nlm.nih.gov/books/NBK24693/

[42] Cary Funk, Alec Tyson. 2020. Intent to Get a COVID-19 Vaccine Rises to 60% as Confidence in Research and Development Process Increases. (December 2020). Retrieved April 25, 2021 from https://www.pewresearch.org/science/2020/12/03/intent-to-get-a-covid-19-vaccine-rises-to-60-as-confidence-in-research-and-development-process-increases/

[43-44] Ziad Obermeyer, Brian Powers, Christine Vogeli, Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366 (2019), 447-453.

[45] Moninder Singh, Karthikeyan Natesan Ramamurthy. 2019. Understanding racial bias in health using the Medical Expenditure Panel Survey data. arXiv: 1911.01509v1. Retrieved from https://arxiv.org/pdf/1911.01509.pdf

[46] Milena A. Gianfrancesco, Suzanne Tamang, Jinoos Yazdany, Gabriela Schmajuk. 2018. Potential Bias in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med* 178(11) (November 2018), 1544-1547.

[47] Code.org. 2018. Is diversity in computing jobs improving? (April 2018) Retrieved April 26, 2021 from https://codeorg.medium.com/is-diversity-in-computing-jobs-improving-32f30068b7de

[48] IBM Developer Staff. 2020. AI Fairness 360. (March 2021). Retrieved April 30, 2021 from https://developer.ibm.com/technologies/artificial-intelligence/projects/ai-fairness-360/

[49] Tom Simonite. 2020. Google offers to help others with the tricky ethics of AI. (August 2020). Retrieved April 29, 2021 from https://arstechnica.com/tech-policy/2020/08/google-offers-to-help-others-with-the-tricky-ethics-of-ai/

[50] Peter N.K. Schuetz. 2021. Fly in the Face of Bias: Algorithmic Bias in Law Enforcement's Facial Recognition technology and the Need for an Adaptive Legal Framework. *Minnesota Journal of Law & Inequality* 39, 1, Article 8 (February 2021), 236-237.

[51] Karen Hao. 2019. Congress wants to protect you from biased algorithms, deepfakes, and other bad AI. (April 2019). Retrieved April 28, 2021 from https://www.technologyreview.com/2019/04/15/1136/congress-wants-to-protect-you-from-biased-algorithms-deepfakes-and-other-bad-ai/

Bibliography

Amazon Web Services. 2021. What is Amazon Rekognition. (2021). Retrieved April 14, 2021 from https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html

Alex. 2018. Racial Bias and Gender Bias Examples in AI systems. (September 2018). Retrieved April 20, 2021 from https://medium.com/thoughts-and-reflections/racial-bias-and-gender-bias-examples-in-ai-systems-7211e4c166a1

Anderson, Cindy. 2020. Risk Assessment Instruments Are Inappropriate for Sentence Reform: Real solutions for reform address racial stratification. *Georgetown Journal of Law & Modern Critical Race Perspectives* 12, no 2 (Fall 2020), 193-194.

Angwin, J., Larson, J. Mattu, S and Kirschner, L. (2016). How We Analyzed the COMPAS Recidivism Algorithm. (May 2016). Retrieved April 21, 2021 from https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Angwin, J., Larson, J. Mattu, S and Kirschner, L. (2016). Machine Bais. (May 2016). Retrieved April 21, 2021 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Brown, Annie. 2020. Biased Algorithms Learn From Biased Data: 3 Kinds Biases Found In AI Datasets. (February 2020). Retrieved January 10, 2021 from https://www.forbes.com/sites/cognitiveworld/2020/02/07/biased-algorithms/?sh=3e0a4fc776fc

Buolamwini, Joy and Gebru, Timmit. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR* 81 (2018), 77-91.

Chilas-Wood, Alex. 2020. Understanding risk assessment instruments in criminal justice. (June

2020). Retrieved April 16, 2021 from https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/#footnote-16

Cloudfactory. The Ultimate Guide to Data Labeling for Machine Learning. Retrieved April 29, 2021 from https://www.cloudfactory.com/data-labeling-guide#:~:text=While%20the%20terms%20are%20often,consistent%20with%20real%2D world%20conditions.

Code.org. 2018. Is diversity in computing jobs improving? (April 2018) Retrieved April 26, 2021 from https://codeorg.medium.com/is-diversity-in-computing-jobs-improving-32f30068b7de

Corbett-Davies, S., Pierson, E., Feller, A. and Goel, S. 2016. A computer program used for bail and sentencing decisions biased against blacks. It's actually not that clear. (October 2016) Retrieved April 23, 2021 from https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/

Darrah, Kim. 2017. The troubling influence algorithms have on how we make decisions. (November 2017). Retrieved April 1, 2021 from https://www.theneweconomy.com/technology/the-troubling-influence-algorithms-have-on-how-we-make-decisions

Funk, Carly and Tyson, Alec. 2020. Intent to Get a COVID-19 Vaccine Rises to 60% as Confidence in Research and Development Process Increases. (December 2020). Retrieved April 25, 2021 from https://www.pewresearch.org/science/2020/12/03/intent-to-get-a-covid-19-vaccine-rises-to-60-as-confidence-in-research-and-development-process-increases/

Georgetown University Law Library. 2021. A Brief History of Civil Rights in the United States.

(April 2021). Retrieved April 20, 2021 from https://guides.ll.georgetown.edu/c.php?g=592919&p=4172706

Gianfrancesco, M. A., Tamang, S., Yazdany, J. and Schmajuk, G. 2018. Potential Bias in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med* 178(11) (November 2018), 1544-1547.

GitBook. Predictive Policing. Retrieved April 29, 2021 from https://teamupturn.gitbooks.io/predictive-policing/content/systems/hunchlab.html

Goel, S., Shroff, R., Skeem, J. and Slobogin, C. 2018. The Accuracy, Equity and Jurisprudence of Criminal Risk Assessment. (December 2018).

Goodman, Rachel. 2018. Why Amazon's Automated Hire Tool Discriminated Against Women. (October 2018). Retrieved April 29, 2021 from https://www.aclu.org/blog/womens-rights/womens-rights-workplace/why-amazons-automated-hiring-tool-discriminated-against

Gramlich, John. 2020. Black Imprisonment rate in the U.S. has fallen by a third since 2006. (May 2020). Retrieved April 20, 2021 from https://www.pewresearch.org/fact-tank/2020/05/06/share-of-black-white-hispanic-americans-in-prison-2018-vs-2006/

Hart, Eric. 2021. Machine Learning 101: The What, Why, and How of Weighting. Retrieved April 29, 2021 from https://www.kdnuggets.com/2019/11/machine-learning-what-why-how-weighting.html

Hao, Karen. 2019. Congress wants to protect you from biased algorithms, deepfakes, and other bad AI. (April 2019). Retrieved April 28, 2021 from https://www.technologyreview.com/2019/04/15/1136/congress-wants-to-protect-you-from-biased-algorithms-deepfakes-and-other-bad-ai/

IBM Developer Staff. 2020. AI Fairness 360. (March 2021). Retrieved April 30, 2021 from https://developer.ibm.com/technologies/artificial-intelligence/projects/ai-fairness-360/

Kantayya, Shalini . 2020. Your undivided attention: Bonus - Coded Bias. Podcast. (April 2021). Retrieved April 9, 2021 from https://www.humanetech.com/podcast/bonus-coded-bias

Latessa, E. J, Lemke, R., Makarios, M., Smith, P. and Lowenkemp, C. T. 2010. The Creation and Validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, 71, 1. (June 2010).

Liptak, Andrew. 2019. The NYPD is using a new pattern recognition system to help solve crimes. (March 2019). Retrieved April 20, 2021 from https://www.theverge.com/2019/3/10/18259060/new-york-city-police-department-patternizer-data-analysis-crime

Maina, I. W., Belton, T. D., Ginzberg, S. and Aingh, A. 2017. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science & Medicine*, 199, (2018), 219-229.

National Research Council. 2004. Understanding Racial and Ethnic Differences in Health in Late Life: A Research Agenda. (2004). Retrieved April 25, 2021 from https://www.ncbi.nlm.nih.gov/books/NBK24693/

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366 (2019), 447-453.

O'Neil, Kathy. 2016. *Weapons of Math Destruction.* Crown Publishing Group, New York, NY, 29.

PredPol. 2020. Predictive Policing Blog. Retrieved April 29, 2021 from https://www.predpol.com/

Schuetz, Peter N.K.. 2021. Fly in the Face of Bias: Algorithmic Bias in Law Enforcement's Facial Recognition technology and the Need for an Adaptive Legal Framework. *Minnesota Journal of Law & Inequality* 39, 1, Article 8 (February 2021), 225-237.

Simonite, Tom. 2020. Google offers to help others with the tricky ethics of AI. (August 2020). Retrieved April 29, 2021 from https://arstechnica.com/tech-policy/2020/08/google-offers-to-help-others-with-the-tricky-ethics-of-ai/

Singh, Moninder and Ramamurthey, Karthikeyan Natesan. 2019. Understanding racial bias in health using the Medical Expenditure Panel Survey data. arXiv: 1911.01509v1. Retrieved from https://arxiv.org/pdf/1911.01509.pdf

Snow, Jacob. 2018. Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots. (July 2018). Retrieved April 13, 2021 from https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

The Sentencing Project. 2020. Criminal Justice Facts. (2020). Retrieved April 21, 2021 from https://www.sentencingproject.org/criminal-justice-facts/

Vincent, James. 2016. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. (March 2016). Retrieved April 9, 2021 from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Wilson, B., Hoffman, J. and Morgenstern, J. 2019. Predictive Inequity in Object Detection. arXiv: 1902.11097v1. Retrieved from https://arxiv.org/pdf/1902.11097.pdf