5-28-2017

Algorithmic Discrimination White Paper

Vicky Wei

Teresa Stephenson

MEMO

May 28, 2017

To: Shankar Narayan, Technology and Liberty Director; Professor William Covington

From: Vicky Wei and Teresa Stephenson, UW: Technology Law and Public Policy Clinic

Re: Algorithmic Discrimination White Paper

Introduction

Technological innovation has led to the prevalent use of algorithms in everyday decision making. So ubiquitous is the application of algorithms that many may not recognize its impact on their daily lives. From online shopping to applying for a home loan, algorithms are at play in categorizing and filtering individuals to serve the goal of providing more accurate and efficient results than human decisionmaking would. At the basic level, algorithms are nothing more than a series of step-by-step instructions compiled by a computer, which then analyzes swaths of data based on those instructions.[1] However, when algorithms use incorrect variables to filter results - such as certain stereotypes about minorities - or, more imperceptibly, learn bad habits from how humans behave online, our absolute reliance on their results can cause disparate harm to minority communities.[2]

The pervasive use of algorithms by both corporate and government organizations for the purposes of efficiency and pattern analysis in the collection of Big Data has brought questions to light as to (1) whether these algorithms are fair across the board and (2) whether they contribute to disparate outcomes resulting in discriminatory practices. The inquiry then

[1] Luke Dormehl, *The Formula: How Algorithms Solve All Our Problems...and Create More* (Perigee, 2014).
[2] *Id*. at 150.

ultimately turns to the legal methods to regulate algorithms in order to combat their negative influence while still maintaining all the technological success and convenience society enjoys.

Defining Algorithms

I. What Are Algorithms?

"Algorithm" is a generic term that, at the most basic level, simply means "a sequence of instructions telling a computer what to do."[3] It is created from a string of computer codes that analyzes, filters, and select information from large data sets about individuals and trends. Algorithms can also track behavior over time and reverse engineer data to determine the individual characteristics that make up a particular result. For example, banks use algorithms to detect money laundering by tracking every transaction the bank makes on a daily basis to find suspicious activity at the root of the money laundering.[4] An algorithm is also the formula that search engines like Google use to show tailored results for our searches or suggestions of products that Amazon predicts and individual will purchase based on previous search history product purchases. Algorithms are also responsible for recommending the television programs one watches, books they read, and people they date. While these seem like benefits to maximize our preferences, algorithms can also be used to determine whether an individual is likely to be a hardworking employee, commit a crime, or be a bad driver.[5]

The advantage of algorithms is that they are useful to navigate large amounts of data (analyzing about 2.5 quintillion bytes of data that are generated each day)[6] that humans are incapable of synthesizing, and subsequently draw conclusions and identify patterns from all the data. They are also cost and time-efficient while drawing objective, logical conclusions of the

[3] Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake our World*, 1 (Basic Books 2015).
[4] Dormehl, *supra* note 1.
[5] *Id*. at 18.
[6] *Id*. at 2.

chaotic social order humans have created. However, society has begun to recognize the dangers brought by the tailored efficiency of algorithms. By sifting through an individual's preferences and providing the best search results to suit the preference, a filter bubble emerges in which individuals are unable to see opposing or alternative viewpoints or products in social media or online shopping activities. For example, Google searches will provide very different results to individuals who the algorithm deems to be a liberal from online activity than to individuals who are deemed to be conservatives. Results may also differ between those the algorithm deems to be male and those deemed female, again by tracking the individual's' online behavior over time.[7] And while some may say this is, again, benign and positive preference individualization, consider the dangers when such differences are used by the government to determine criminal threats. Police departments across the country have used algorithms to detect whether a particular neighborhood in the city is more likely to attract crime based on past police activity, as well as what types of individuals are more likely to reoffend based on one's age, childhood factors, and resident zip code.[8]

The danger of such reliance on algorithms is that, despite the benefits and assumption that algorithms are efficient, logical, and data-driven and therefore unbiased, algorithms are not infallible and oftentimes carry biases of their own or of their creators.[9] Algorithms are not only sometimes wrong in predicting shopping preferences, but even more dangerous, are noted to turn out false positives and false negatives in forecasting individuals likely to commit a crime.[10] Such mistakes based on incorrect assumptions or stereotypes about an individual's characteristics (age, race, gender, zip code, etc.) results in unfortunate consequences for those of

[7] *Id*. at 47.
[8] *Id*. at 110-113.
[9] *Id*. at 232.
[10] *Id*. at 123. Predictive tools on criminal activity are only accurate 75% of the time.

differing socioeconomic and racial backgrounds when used in policymaking, police strategies, or employment hiring practices.

II. How Do Algorithms Work?

Big data is the term given to the large amounts of data gathered by companies about individuals through the use of devices, computers, and the internet.[11] Big data is necessary for algorithms to function and crunch numbers. In order for algorithms to work initially, there needs to be a starting set of data to be analyzed.[12] The more data the better for learning algorithms because it allows the algorithms to find patterns or answer questions posed to it by combing through enormous amounts of data.[13] Learning algorithms are generally various types of sub-algorithms that take in differing types of data and produce different answers based on what information was fed into the algorithm.[14] Every sub-algorithm that produces a desired result, such as an individual clicking on a certain ad or website, increases the amount of times the sub-algorithm is used to make that decision. Algorithms and data sets, at their core, are just numbers.

Furthermore, the data that is entered into the algorithm has the ability to affect the outcome based on the parameters set by the algorithm creator. At its very core an algorithm is a model, a tool, to simplify and make combing through data more efficient. Unfortunately in the process of creation the model (algorithm) is oversimplified and the creators "make choices about what's important enough to include, simplifying the world into a toy version that can be

[11] Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion?* (January 2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.
[12] Domingos, *supra* note 3.
[13] *Id*. at 9.
[14] *Id*.

easily understood and from which we can infer important facts and actions."[15] From this desire to over simplify, often factors that should or should not be included in the data collection process are added to the mix resulting in discriminatory outcomes for individuals in different sectors.

Identification of Problems with Algorithm Use

Although an algorithm may not be discriminatory in nature, the use of the algorithm's results can lead to disparate effects. The collection of more data about individuals as they use the internet leads to possible new methods of discrimination. Companies and data brokers that collect data about individuals as they surf the web and use products from different vendors collect and aggregate all the data about individuals in order to create a profile. The problems associated with such aggregation of data is beyond the scope of this paper, but the use of the profile is important to consider when looking at how discrimination happens at the hands of algorithms.

Two common discriminatory issues arise with the reliance on seemingly benign algorithms. First, the input of data that allows algorithms to identify these trends is biased to begin with, and then ultimately wielded by humans, results in discriminatory ends. Second, and the more difficult problem to identify, is learner algorithms who learn from bad habits of humans. Both theses types of algorithm use reinforce pre-existing discrimination or biases that humans suffer from even without the use of technology.

I. Discriminatory Input

The use of algorithms has become commonplace as technology advances because it removes much of the pressure of decision-making off human beings. This is not only

[15] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 20 (Crown 2016).

convenient, but it is intended to remove bias inherent in human decision-making. "The attraction of [an algorithm] isn't hard to understand. It promoted 'automation bias,' an assumption that a machine-driven, software-enabled system is going to offer better results than human judgment."[16] Unfortunately, the data entered into the neutral algorithm is flawed from the beginning. Typically, the initial data entered into the system is based off previous year's record keeping and data or factors that a particular algorithm creator deems important to consider (race, gender, age, weight, etc.). Depending on the sector that data has inherent biases of countless reviewers which is all fed into the algorithm. "[T]he reality is that humans craft those algorithms and can embed in them all sorts of biases and perspectives."[17] Therefore, the danger with this type of algorithm is that even with the most "neutral" analysis of data trends aggregated over time, humans are still capable of using the results to draw discriminatory conclusions and create disparate results.

II. Machine Learning Algorithms

Machine-learning algorithms adapt to new information it acquires and fine-tunes its targeting and analysis based on evolving input. For example, in advertisement targeting, the learning mechanism of algorithms have identified a click-through trend from some women on certain websites or social media sites clicking on advertisements for lower wage employment opportunities and therefore maximized on this "success" rate by increasing the amount of such advertisements seen by women on social media pages, user profile pages, and websites geared toward women.[18] Unfortunately, this ultimately creates a self-perpetuating bias as women are

[16] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, 107 (Harvard University Press, 2015).
[17] Dormehl, *supra* note 1, at 150.
[18] Amit Datta, Michael Carl Tschantz, and Anupam Datta, Automated Experiments on *Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, 1 Proceedings on Privacy Enhancing Technologies 92, 92-112 (2015).

potentially less exposed to other employment options with higher wages on social media platforms.

In learning how to perfect results, algorithms can perpetuate their own bias and create the bias. Machine learning algorithms are consider a black box

> a term used by cybernetician whenever a piece of machinery or else a set of commands is too complex. In its place, the black box stands in as an opaque substitute for technology in which nothing needs to be known other than inputs and outputs. Once, opened, it makes both creators and the users confront the subjective biases and processes that have resulted in a certain answer.[19]

This closed, black box system, presents the need for transparency, for both consumers and those that use the algorithm. Unless a creator of an algorithm can say without a doubt that the algorithm is only using appropriate data in determining its answers auditors should be able to examine the code. Machine-learning algorithms present a challenge that need to be met with accountability.

State of the Law: Examining Algorithms

I. Barriers to Examining Algorithms

A. Trade Secret Laws

A large problem is the fact that most algorithms and the way they work are ensconced in a black-box where the coding and technical details of how algorithms track and collect data is mainly unknown to anyone other than the computer designers who create the algorithms.[20] To make things more complicated, the machine learning and adaptive quality of the complex algorithms used today that constantly change to provide more tailored results evolve to a point

[19] Dormehl, *supra* note 1, at 235.
[20] O'Neil, *supra* note 15, at 29.

where its computer engineer creators may not even recognize or have the ability to control.[21] Unfortunately, with today's technology competition of search engines and web platforms attempting to provide the most accurate results for customers, the likelihood of divulging such information to the public would violate companies' trade secrets self interest. The problem with the inaccessibility of black box trade secrets with algorithms is the reliance on data conclusions that human decisionmakers do not understand.

While most trade secret law is statutory varying from state to state, federal and international regulations like the Uniform Trade Secrets Act and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) provide commonalities in trade secret law. The general goals of trade secret laws are to protect commercially valuable proprietary information that gives competitive advantage, including business strategies, management, design concepts, manufacturing techniques, and formulas. According to the Congressional Research Services (CRS) Report:

> Whether information qualifies as a "trade secret" under federal or state law is a question of fact that may be determined by a jury. A jury may consider several factors in assessing whether certain material is a trade secret, including the following: the extent to which the information is known outside of the company; the extent to which it is known by employees and others involved in the company; the extent of measures taken by the company to guard the secrecy of the information; the value of the information to the company and to its competitors; the amount of effort or money expended by the company in developing the

[21] Dormehl, *supra* note 1, at 236.

> information; and the ease or difficulty with which the information could be properly acquired or duplicated by others.[22]

Companies will continue to be protected by trade secret law, as unlike patents or copyrights, trade secrets never expire and therefore do not ever have to be revealed or released to the public.[23]

At the state level, Washington's trade secret law is codified in RCW 19.108.010 which defines trade secret to mean information including a formula or method that:

> (a) Derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use; and (b) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.[24]

The only ways in which trade secret protection can be lost are if there is even a single unprotected disclosure to the public of the trade secret (either by photograph, technical, paper, or any other form of disclosure); the product is legally reverse engineered to discover the trade secret; or if a competitor independently develops around it.[25] The crux of trade secret protections revolve around reasonable measures that a company has taken to maintain secrecy.[26]

Algorithms used to enhance search engine results or provide predictions on rates of recidivism of criminally convicted individuals are certainly information that has substantial economic value, engendering fierce competition, incentivizing its secrecy. An oft-referred to

---

[22] Congressional Research Services, 7-5700, R43714, Protection of Trade Secrets: Overview of Current Law and Legislation, 2 (Apr. 22, 2016).
[23] Pasquale, *supra* note 16, at 83.
[24] RCW 19.108.010
[25] Office of Policy and External Affairs, *Trade Secrets Protection in the U.S.*, United States Patent and Traditional Trademark Office (2012), https://www.nist.gov/sites/default/files/documents/mep/marinaslides.pdf
[26] *See Nationwide Mut. Ins. Co. v. Mortensen*, 606 F.3d 22 (2010) (holding that no reasonable measures taken when client lists were left with Nationwide agents without further protections).

example is Google's use of multiple algorithms that optimize search results for users, known in the industry as RankBrain and Page Rank, which has deep machine learning components that are not only complex but heavily guarded in secrecy.[27] Google and other commercial service providers maintain strict rules to prevent competitors from analyzing and reverse engineering their algorithms, protected by trade secret and patent law.[28] Thus, whether at a federal or state level, the formula that computer engineers use to create such systems are difficult, if not impossible, to obtain in order to challenge or analyze even if a lawmaker, attorney, or judge was adept at computer engineering. Further, the incentive to maintain the innovative nature of technology via the use of algorithms counsels against blanket legislation directed at undoing or diminishing the role of trade secret laws.[29]

B. Deep Learning: Questions That Simply Cannot Be Answered by Data Scientists.

The goal for learner algorithms is to theoretically become self sufficient. Data scientists strive to create the "Master Algorithm," one in which the algorithm self corrects and correlates data without feedback from the data scientists.[30] As algorithms continue to process data and create parameters based on patterns apparent in the data the outcomes may move away from the original intention but still fall within the goal of the program. This creates several problems as some algorithms have advanced to the point that the creators are no longer sure as to how the algorithm comes to its final answer. If the data scientists and algorithm creators are no longer able to identify the process which the algorithm used to determine its answer, how is anyone able to identify the path the algorithm took? The profiles created by data brokers are put into an algorithm and an answer is given depending on the sector: whether it's this individual is going

[27] Danny Sullivan, *Sneak Peak into Black Box of Google's Search Algorithm*, Search Engine Land (June 23, 2016, 1:01 PM), http://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440.
[28] Darren Stevenson, *Locating Discrimination in Data Based Systems*, Open Technology Institute, 16-19 (2014) https://www.ftc.gov/system/files/documents/public_comments/2014/10/00078-92938.pdf
[29] *Id*.
[30] Domingos, *supra* note 3, at 25.

to be a great employee to this individual has a high potential recidivism rate. Leaner algorithms inhibits the transparency needed to ensure only appropriate categories of data are used (not race, sex, ethnic background, etc.) to make determinations.

II. Laws Which Enable the Examination of Algorithms

The unregulated nature of algorithms has begun to change as society begins to recognize the dangers of leaving algorithms primarily in charge of human decisionmaking. A good sign is the European Union (EU), for example, which has incorporated a new routine regarding the use of machine learning algorithms. The EU's General Data Protection Regulation (GDPR), effective in May 2018, will legally provide a "right to explanation" for users to challenge the ways in which algorithms are used especially in decisions made about them.[31] Additionally, Article 22 ("*Automated individual decision-making, including profiling*") lists prohibitions against the use of algorithms in systems like credit and insurance risk assessments and other programs which may use machine learning techniques.[32] While the effects of the new law cannot be fully understood before 2018, the outlook is positive in understanding the need to regulate and understand the tools that underlie so much of human decisionmaking.

Other methods have also emerged to combat the issues raised by the secretive and complicated nature of algorithms perpetuating discrimination. Among them are methods like (1) algorithmic auditing, in which test variables are placed into search engines or systems run by algorithms to find discrepancies due to racial barriers; (2) using consumer protection laws like the Computer Fraud and Abuse Act to provide for equal opportunity protections affected by algorithmic tailoring in a commercial setting; (3) antitrust laws in lieu of overcoming trade secret laws to perhaps regulate the use and creation of algorithms in ways that may engender

[31] Bryce Goodman and Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, OXFORD INTERNET INSTITUTE (2016), https://arxiv.org/pdf/1606.08813v3.pdf.
[32] *Id*. at 2.

open competition of search optimization corporations, for example; and (4) constitutional law claims against government agencies from using algorithms that result in disparate effects based on complex, unverified, and biased data assumptions.

A. Algorithmic Auditing

In order to protect consumers and individuals freely giving information online for services there needs to be an auditing mechanism which allows independent third parties to evaluate the algorithms in use to search for and determine whether biases are present in the system. In order for this to be plausible, "data brokers need to fess up about the data they are hoarding, trading and selling."[33] Similar to the laws being established in the E.U., governments in the U.S. need to create policies to keep internet companies accountable.

B. Consumer Protection Laws

Many Federal regulations have already been implemented in the field of consumer protection in different sectors by Federal Agencies. These laws could possibly be applied to the realm of algorithms and collection of data in order to protect consumers from the potentially discriminatory outcomes. Data brokers create profiles of individuals by accumulating data from several different companies. These profiles are created unknowingly by monitoring everyday online interactions. Acts likes the Computer Fraud and Abuse Act; The FTC Act; Fair Credit Reporting Act; and Equal Employment Opportunity Act were all created to protect consumers and individuals from bigger (more sophisticated) parties.

i. <u>The Computer Fraud and Abuse Act[34]</u>

The Federal Trade Commission's (FTC) mission is to protect consumers by preventing business practices that are unfair or deceptive.[35] The FTC has taken upon itself the role of

[33] Pasquale, *supra* note 16, at 145.
[34] 18 U.S.C. § 1030

enforcer under its Section 5 powers to ensure online business is conducted fairly.[36] In regards to the use of big data and algorithms the FTC has published a report seeking the creation of fair guidelines to ensure the privacy of individuals and best business practice.[37] FTC Chair Edith Ramirez stated we must "ensure that by using big data algorithms [firms] are not accidentally classifying people based on categories that society has decided- by law or ethics - not to use, such as race, ethnic background, gender, and sexual orientation."[38]

The Computer Fraud and Abuse act provides punishment for individuals that have had accessed and violated another's privacy by obtaining information through the individual's computer without permission or caused that individual harm by using that information with intent to defraud.[39] If the act were to be applicable corporations, the entities would most likely be found to have violated the act by using individuals information to determine a "proxy credit score." The information often used to create "proxy credit" scores, is information that an individual might have given to a website for another purpose. The information was not intended to be used for the purposes of determining a credit score and therefore, the information was obtained without the consent of the individual.

ii. Fair Credit Reporting Act (FCRA)[40]

The purpose of the FCRA requires credit reporting agencies to " adopt reasonable procedures for meeting the needs of commerce for consumer credit, personnel, insurance, and other information in a manner which is fair and equitable to the consumer, with regard to the

---

[35] Federal Trade Commission, *About the FTC*, https://www.ftc.gov/about-ftc.
[36] 15 U.S.C. § 45(a).
[37] FTC, *Big Data*, *supra*, note 11.
[38] Pasquale, *supra* note 16, at 40 (citing Edith Ramirez, "Privacy Challenges in the Era of Big Data: The View from the Lifeguard's Chair," Keynote Address at the Technology Policy Institute Aspen Forum (Aug. 19, 2013), https://www.ftc.gov/sites/default/files/documents/public_statements/privacy-challenges-big-data-view-lifeguard%E2%80%99s-chair/130819bigdataaspen.pdf).
[39] 18 U.S.C § 1030.
[40] 15 U.S.C. § 1681.

confidentiality, accuracy, relevancy, and proper utilization of such information."[41] The law outlaws the use of certain information about individuals to used in determining credit worthiness, which relates to an individual's credit score (e.g. the neighborhood in which one lives (redlining), race, color, religion, national origin, sex, etc.)[42]. The act should theoretically limit the usage of data obtained online to prevent the use of information that goes into a credit score, however, the act is specifically tailored to apply only to "credit reporting agencies." The FTC has called for data brokers to be included in the scope of the act,.[43] "Credit Reporting Agencies," are defined in FCRA to means any individual for money, or fees, that "engages in whole or in part in the practice of assembling or evaluating consumer credit information or other information on consumers for the purpose of furnishing consumer reports to third parties."[44] Applying this definition to Data Brokers, Data Brokers sell information about individuals to third parties in order for the third parties to develop "proxy credit scores" based on the information that has been received. Therefore, as the FTC claims, Data Brokers should be considered "Credit Reporting Agencies" under FCRA.

C. Antitrust Laws

As data collection and the internet have become dominated by relatively few companies, especially in the realm of search engines, there have been many claims against large internet companies in regard to antitrust violations, specifically Section 2 of the Sherman Act.[45] These antitrust claims are not only popping up in the United States but in the European Union as well.[46] In relation to Google specifically, Joshua Hazan, has found that the company's method

[41] 15 U.S.C § 1681(b).
[42] Federal Trade Commission, *Your Equal Credit Opportunity Rights*, Consumer Information, (Jan. 2013), https://www.consumer.ftc.gov/articles/0347-your-equal-credit-opportunity-rights.
[43] FTC, *Big Data*, *supra*, note 11.
[44] 15 U.S.C. § 1681a(f).
[45] Pasquale, *supra* note 16, at 162.
[46] *Id*. at 163.

of sorting search results favors its own subsidiaries or related companies over third parties.[47] Section 2 of the Sherman Antitrust Act makes it unlawful for any person to "monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations."[48] In order to meet this standard a company must (1) have "dangerously" close to or has already achieved a monopoly power and (2) participate in anticompetitive conduct.[49]

The Federal Trade Commission has attempted to bring action against Google under the Sherman Act to challenge its use of search ranking algorithms to benefit its own products over that of their competitors. The FTC claim against Google regarded the algorithms hindering competitive behavior by manipulating its search options for consumers. However, the success has been limited, and the FTC concluded its investigation without major changes to Google's search ranking practices.[50] However, the European Commission has brought a more anti-competitive action against Google under the same claims of inherent algorithmic unfairness it is search ranking results relating to comparison shopping services which were preferred in Google search results over Google's competitors. Google's search engine dominates 90% of the European market share and therefore its influence on the market in the EU may make the European Commission's claim stronger than the FTC's investigation under the Sherman Act.[51]

i. The Fourth Amendment

[47] Joshua Hazan, *Stop Being Evil: A Proposal for Unbiased Google Search*, 111 Michigan L. Rev. 789, 798 (2013).
[48] 15 U.S.C § 2
[49] Hazan, *supra* note 47, at 799.
[50] Eric Savitz, G*oogle and the FTC's Investigation: A Cautionary Tale*, Forbes, (Nov. 4, 2012, 07:48 PM), http://www.forbes.com/sites/ciocentral/2012/11/04/google-and-the-ftcs-investigation-a-cautionary-tale/#263a228f3434
[51] James F. Peltz and Tiffany Hsu, *Google Faces Long Battle in EU Antitrust Case*, Los Angeles Times (Apr. 17, 2015), http://www.latimes.com/business/la-fi-eu-google-strategy-20150417-story.html

The Fourth Amendment of the U.S. Constitution is meant to protect the right of "people to be secure in their persons, houses, papers, and effects,"[52] this right has been extended to protect that which people assume or reasonably suspect to be private in terms of information that is posted given free online.[53]

Case Study

I. Use of Algorithms in Hiring and Firing Practices

A case study in the use of algorithms to decide hiring and firing practices is instructive because the consequences of its use is not straightforwardly positive or negative. In criminal sentencing, predictive policing, lending and credit score ratings, and insurance coverage, the results of using algorithms have been clearly negative on a wide scale and often a clear proxy for race. However, the results of using algorithms in hiring and firing practices is not absolute. There is still space for change especially when the disparate impact is usually not a result of malicious intent in the creation of the algorithm. As long as blind reliance is avoided and transparency in the algorithm used is ensured, algorithms may be successful in the human resources world.

The use of algorithms in predictive policing contexts has been evidenced to result in negative and discriminatory consequences, as well as the use of algorithms in credit scoring coding particular individuals as high risk or lazy simply because of their name or their zip code.[54] However, it is undeniable that using algorithms to analyze numerous resumes filed by over-optimistic candidates is a positive and efficient use of the tool. Further, directing algorithms to look for whether candidates meet a specific education requirement, for example, is not necessarily illegal while other hiring benchmarks like proxy intelligence tests are.

[52] U.S. CONST. AMEND. IV, §3.
[53] *Katz v. United States*, 88 S.Ct. 507 (1967).
[54] Pasquale, *supra* note 16, at 38

However, one cannot ignore the disparate impact that the use of algorithms have on minority groups in many areas of the work force. Furthermore, the secrecy and complexity of algorithms in making these determinations are dangerously ensconced in a black box that no one, not even employers carrying out the hiring and firing decisions based on these algorithms, can explain the reasons behind the choice. Therefore, while there are many arguments for the positive effect and use of algorithms in hiring practices, it is still important to be cognizant of ways that these tools can mask discrimination.

A. Defining the Problem

An example of algorithms used in the field of human resources is told by Cathy O'Neil in her book, Weapons of Math Destruction.[55] Ms. O'Neil identifies Sarah Wysocki, a fifth grade teacher who had routinely received positive scores on her yearly performance reviews, yet was abruptly given abysmal scores the year an algorithm was used to evaluate teachers rather than human based performance reviews. Ms. Wysocki was later terminated because of district-wide policy to cut low performing teachers. Ms. O'Neil and Ms. Wysocki both argue that algorithms ignore the wide range of soft factors that go into human resources and evaluations of performance like teaching, which is not inherently data-driven.[56] Student to teacher relationship, self-esteem, and parent comfort levels with the teacher are all factors incapable of being evaluated by algorithms.

Another problem with algorithms that Ms. Wysocki's story reveals is the dangerous lack of transparency inherent in how algorithms arrive at their conclusion about job performance. Even if advocates of algorithms argue that these tools remove human error and bias which might have given Wysocki soft and positive evaluations in the past, there is no denying that

[55] O'Neil, *supra* note 15.
[56] *Id.* at 6.

school administrators relying on these algorithms do not understand how the algorithm found Wysocki inadequate. When Ms. Wysocki sought explanations as to why she received such a low score, how she could improve, and what she could work on, her school principal and the district simply did not know.[57] They had blindly trusted a private company program that had promised efficient and accurate identification of problem teachers, without "peeking under the hood" of the algorithm to know what it was that the tool was actually evaluating. Even if they had been given full access to the code and data used by computer engineers, the likelihood that a school administrator could understand its significance is low.

This lack of transparency is frustrating for someone in Ms. Wysocki's position, as well as many other teachers who are faced with technology-based evaluations and performance reviews based on standardized test results of their students. However, it is even more insidious and dangerous when discriminatory impact results for the use of these algorithms, with no one to explain why or how this discrimination is occurring. When companies and government agencies begin to use algorithms for hiring and firing purposes, and see decreased minority representation in their workforce, for example, without transparency behind what is really going on in the algorithm, such blind reliance on these technology tools can have harmful social effects.

When talking about the use of algorithms to replace humans in the recruitment and retention of employees decisionmaking process, there are diametrically opposing viewpoints - some believe that this is the only neutral way in ensuring diversity and equality in the working field,[58] while others think that algorithms serve as a mask for inherent discrimination[59]. The

[57] *Id.* at 5.
[58] Bourree Lam, *For More Workplace Diversity, Should Algorithms Make Hiring Decisions?*, THE ATLANTIC, (Jun. 22, 2016), https://www.theatlantic.com/business/archive/2015/06/algorithm-hiring-diversity-HR/396374/

truth is somewhere in between: while humans are notorious for implicit bias, a total reliance on algorithms can continue the trends that humans have already perpetuated. For example, in the context of employment, the hiring of employees can be greatly streamlined if an employer receives thousands of resumes a day for any given position, an algorithm that codes for specific requirements can eliminate non-starter candidates and save the employer time and money from sifting through piles of applications by hand. However, if algorithms are written to look for terms in the resume which end up being a proxy for race or gender. Especially when great weight is given to the reputation of a university, the systemic inequities that may have prevented an otherwise well-qualified applicant from enrolling in a prestigious institute of higher learning would then be perpetuated in the work field, further causing wealth and social disparities. Further, algorithms that look for certain terms or specific education requirements may ignore "soft" factors which could perpetuate systemic discrepancies in education and social economics, leading to an even widened minority gap.

Even more involved than sifting through large piles of resumes for keywords, some algorithms mine for data about a particular candidate to find attributes about a person that would suggest a good employee (punctuality, charisma, detail-oriented translating to being a good salesperson, longer tenure, or high production value).[60] Data-mining is a complicated practice, but in the employment context, it includes algorithms combing through the internet's vast networking and informational system about a candidate and aggregate information to form conclusions given trends or patterns in every piece of information about them available - from

---

[59] Gideon Mann and Cathy O'Neil, *Hiring Algorithms Are Not Neutral*, HARVARD BUSINESS REVIEW (Dec. 9, 2016), https://hbr.org/2016/12/hiring-algorithms-are-not-neutral; Patrick Marshall, *Algorithms Can Mask Bias*, SAGE BUSINESS RESEARCHER (February 15, 2016), http://businessresearcher.sagepub.com/sbr-1775-98200-2717795/20160215/algorithms-can-mask-biases-in-hiring.

[60] Solon Barocas and Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 679 (2016).

zip code to social media behavior.[61] This way, data-mining algorithms get a more holistic image of what a candidate is truly like and therefore arrive at more "accurate" predictions about their likelihood of success at a given position. However, because data on racial and gender minorities already reflect a trend in inequalities that result in lower employment rates, higher conviction rates, more time off from work to care for family, etc., the algorithms blinding churning through this data will only learn from this system of inequalities and make predictions about minority candidates based on patterns in the minority group. [62] Therefore, minority candidates will still be labeled by algorithms as less favorable for employment because of existing societal bias. Algorithms do not solve the problem of bias, but further hides the problem under a veil of neutrality.

Finally, another issue presented with algorithmic hiring that emphasizes data mining and pattern recognition is the populations on the fringes of technology who are left out of the employment process if they do not have access to technology. For example, if geographically or economically, a certain group of individuals are less likely to engage in technology like social media or reference the correct coded terms in their job applications and resumes for algorithms to identify, their lack of technological footprint may preclude them from otherwise viable employment opportunities.[63] This discrimination by ommission further marginalizes those who are economically vulnerable through barriers in technology.

Some look to the technology industry as an indicator of where overreliance on algorithms in hiring practices have gone wrong, where as little as two percent of Google's vast work force, for example, was African American.[64]As Frank Pasquale states, "algorithms just dry

[61] *Id*. at 680.
[62] *Id*. at 691.
[63] *Id*. at 685.
[64] Pasquale, *supra* note 16.

up discrimination further up."[65] The problem with algorithms in aiding the human decisionmaking process in employment practices is has more to do with the assumption that algorithms are unbiased and neutral without understanding the mechanisms that go into developing the algorithm or how the algorithm arrives at a particular result. The "black box" of algorithms - whether it be an complexity of computer coding or the trade secret protections of data brokers or companies who develop algorithms making the understanding of algorithms opaque - hampers a reasonable approach to human resources to justify hiring and firing decisions.[66]

Even more dangerously, this black box of secrecy can allow discrimination based on race or gender to proliferate insidiously and go undetected until disparate impacts on minority populations are revealed. Perhaps the European Union's solution in the Right to Know and challenge algorithmic decisions can pave a way to better accountability and oversight which could reveal the dangers of relying on secret formulas that no one can explain.

B. Where Does the Bias Enter the Algorithm?

Bias enters the algorithm mostly at the early stages of creation. Discrimination can result in primarily two ways: First, algorithms are learning tools, which mimic human behavior. For example, if a particular company has hired more men than women or has retained mostly white employees, an algorithm may associate the majority gender or race as qualities of successful employees, and therefore select applicants who are male, or white, or have white-sounding names.[67] In this way, because of institutionalized discrimination already prevalent by human machinations, algorithms mimic these patterns to perpetuate prior prejudices. The

[65] *Id*. at 21
[66] *Id*.
[67] Mann, *supra* note 59.

following is an example of what Solon Barocas and Andrew D. Selbst describe as the "garbage in, garbage out" problem of data science:

> If LinkedIn determines which candidates to recommend based on the demonstrated interest of employers in certain types of candidates, Talent Match will offer recommendations that reflect whatever biases employers happen to exhibit. In particular, if LinkedIn's algorithm observes that employers disfavor certain candidates who are members of a protected class, Talent Match may decrease the rate at which it recommends these candidates to employers. The recommendation engine would learn to cater to the prejudicial preferences of employers.[68]

The second way that algorithms intuit bias is through the creation of the code itself. If the code looks for certain terms that are characteristic to one particular protected class of minorities, then the resulting candidate pool shrinks to those who have those desired traits. Similar to proxy credit scoring, if coded terms are simply a proxy for race, for example, like zip codes, the effect would be a homogenous set of individuals. Further, if there are flaws in the underlying data being collected which rely on historical prejudices or systemic inequalities, such trends will be baked into the underlying algorithm and further exacerbate the discrimination, however unintentional.

Both of these problems extend in some way from faulty data. For example, if a company or agency uses an algorithm based on grades or assessments given by human generated performance evaluations, whatever bias from previous human input would be perpetuated and expanded by the algorithm with no corrective oversight.[69] Solon Barocas and Andrew D. Selbst have deemed both of these algorithmic errors as false or inaccurate target

[68] Barocas, *supra* note 60, at 683
[69] *Id*. at 671.

classifications.[70] Algorithms work by classifying and labelling traits and qualities in order to come to a conclusion about an individual with predictive ability. Either through mining data about a job applicant or decoding language on an applicant's resume and application, by identifying patterns regarding a person with a subset of certain traits and the patterns associated with each of those traits, algorithms can then generate an easy conclusion on which employers base their hiring and firing decisions. Yet because these errors are common, innocent, opaque, and its impact pervasive over reliance on stereotypical pattern-matching can lead to vastly disparate impact for minority populations. For example, if turnover rates for minority groups have a pattern of being high, an employer who engages in algorithmic hiring that codes for likelihood of tenure status would be less likely to find minority applicants identified as qualified candidates.[71]

C. What Laws Might Be Applicable?

While protected minority classes like race, gender, sexual orientation, and veteran status are covered from employment discrimination under Title VII of the Civil Rights Act, the legal hurdle with combatting discrimination with the use of algorithms in hiring and firing practices is the lack of clear evidence of knowledge or intent to discriminate. The only avenue that lawyers can seek is through disparate impact litigation - by looking at the negative results that algorithmic hiring produces even without malicious intent. A successful attempt at this strategy was *Griggs v. Duke Power Company*,[72] where the Supreme Court ruled that it was a violation of Title VII of the Civil Rights Act for the power company to use intelligence test scores and the receipt of a high school diploma as the sole basis of hiring decisions where the impact is its exclusion of racial minorities. However, the reach of this case's success is limited in that

[70] *Id.*
[71] *Id*. at 680.
[72] *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

disparate impact litigation is only applicable if the plaintiffs can prove that the hiring requirements used are irrelevant to the task required of by the job. If, however, a company or government agency can argue that such requirements are a "business necessity" and directly related to job performance, then, regardless of the discriminatory impact or disparate results, the hiring practice will be deemed valid as long as no malicious discriminatory intent was involved.[73] This business necessity argument holds even further weight if the predictive power of algorithms are accurate in assuring high quality candidates, even though it turns a blind eye to other candidates from minority groups.[74]

Furthermore, after the *Gideon v. Duke Power Company* decision, employers have begun to substitute illegal intelligence tests with proxy personality tests which sometimes mirror the aspects of intelligence tests that would unfairly stigmatize applicants. For example, a class action lawsuit was filed against many large corporations for their use of personality tests in job applications that inherently test for medical and mental disabilities with questions like "Do you find yourself angry often?"[75] Algorithms are then used to detect a pattern or likelihood of individuals with these traits and their tendencies for mental or emotional instability, and therefore discourage employers from choosing these candidates. Such tests, if truly equated to medical questions, would run afoul of the sanctions in place through the Americans with Disabilities Act (ADA) which prohibits employers from discriminating against hiring an otherwise qualified applicant due to a disability.[76] Yet employers and private creators of such personality tests have found ways to circumvent these regulations by asking more nuanced

[73] Barocas, *supra* note 60, at 671.
[74] *Id*. at 672.
[75] O'Neil, *supra* note 15, at 108.
[76] 42 U.S.C. § 12101.

questions that do not reflect medical information on its face, and therefore disqualify disabled applicants from employment.[77]

The secrecy and complexity behind algorithmic decisions is therefore dangerous not only because there is little to no way of telling whether malicious discriminatory intent was purposely coded into the algorithm, but also because most likely algorithms will intuit patterns on their own and inadvertently produce the disparate impact on minority groups. A blind reliance on algorithms does little to challenge the status quo and the preexisting societal biases they perpetuate.

D. Possible Solutions in This Field

Of course transparency and regulation of algorithmic use in the public and private sectors would be ideal. However, this approach is riddled with faults as discussed in the above sections regarding barriers to regulation and the current state of the law. If the European Union is successful with their reform-minded Right to Explanation in the new iteration of the GDPR in 2018, then perhaps there is the possibility of algorithms being made with clarity and employers are fully informed of how the algorithms work (computer science data and all) so as to understand the results. It would be a great step forward, yet there are still barriers to such transparency implemented in the United States even if the GDPR was successful. Again, the state of the law mentioned in above sections like trade secret laws still prevent the much needed regulatory framework around algorithms.

Some have suggested that the disparate impact faced by minorities in the workplace, since it is systemic and cannot be dealt with through individual lawsuits, should instead be

[77] O'Neil, *supra* note 15, at 109.

treated much like affirmative action or disability law.[78] In this way, the burden is on the employer to look at hiring schemes including the use of algorithms through the lens of equality and impact rather than through efficiency or pure profit driven goals. This way, there would not be a prohibition on the use of algorithms or data mining to inform employers, but a more open approach in understanding the effect that employing such algorithms might have and finding ways to counteract the systemic biases it perpetuates.

And similar to the ADA, changes to the workplace to accommodate candidates with traits that algorithms have deemed to be a predictor of a bad employee could change the predicted outcome of the applicant. For example, if algorithms have detected that women with families have lower retention rates and would therefore be undesirable for the employer who wants to minimize tenure, the employer could change the workplace by providing flexible schedules for those with families or make other accommodations to make the workplace more family-friendly.[79] Another example would be to make the workplace more friendly and accepting of other cultures and welcome underrepresented minority groups. Changes to what employers can control in their workplace can also disrupt the trajectory of discrimination against protected classes reflected in society, and break the prophetic models of algorithms.

By recognizing the shortcomings of algorithms and avoiding blind reliance on their seeming infallibility, employer can not only find a whole new subset of employees previously weeded out by algorithms, but also work to change the lack of representation of minorities, women, the disabled community, and other vulnerable classes in the work field. Further, by recognizing these disparities, these algorithms could be retooled to account for the discrimination and bias in society, and data-mining software can look for other societal factors

[78] Christine Jolls, *Antidiscrimination and Accommodation*, 115 HARV. L. REV. 642, 652 (2001) (comparing accommodations law to disparate impact).
[79] Barocas, *supra* note 60, at 731.

that would account for lower employment rates, lower graduation rates, and higher incarceration rates for an applicant.

While we cannot hope for absolute transparency of how these algorithms work, simply being cognizant of algorithms' potential to exclude, discriminate, and skew data can perhaps persuade social-minded employers to add corrective measures into their hiring and firing practices.

Conclusion

Has the use of algorithms gone too far? "A single human showing explicit bias can only ever affect a finite number of people. An algorithm, on the other hand, has the potential to impact the lives of exponentially more."[80] The popularity of algorithms is driven by the need for efficiency and to provide an "unbiased" solutions. The issue is not whether it is better to use algorithms instead of humans - clearly there is a lower chance of blatantly bias decisions being made by the algorithm. Rather the issue is the dangers of assuming infallibility of algorithms - the belief that the algorithm can do no wrong. The over reliance on the "unbiased" efficient algorithm is the true problem.

[80] Dormehl, *supra* note 1, at 152.