THE MATH AND THE MATTER: AN ETHICAL EXAMINATION OF HUMAN RELIANCE ON ALGORITHMS

by
Christian Farls

a thesis submitted to
Central Catholic High School
in partial fulfillment of the requirements for completing
The Brother David S. Baginski, FSC Scholars Program

May 2021

Dr. Patrizia Costa
Director, The Brother David S. Baginski, FSC Scholars Program

---

Mr. Todd Rooney
Teacher, English Department

Mr. John Allen
Director, The Brother David S. Baginski, FSC Scholars Program

PLAGIARISM POLICY STATEMENT

This thesis is the original work of the student. No data, knowledge, interpretations, or phraseology has been taken from another without full acknowledgement. All ideas and data that are not common knowledge have been appropriately cited.

ABSTRACT

THE MATH AND THE MATTER: AN ETHICAL EXAMINATION OF HUMAN RELIANCE ON ALGORITHMS

Christian Farls
Central Catholic High School, 2021

One of the largest issues surrounding the use of software-based algorithms, specifically those used in decision-making processes, is their ability to become biased. Aside from unconscious flaws, some argue that there are distinct human qualities important to completing tasks—qualities that algorithms cannot replicate. So, the question arises: can we fully rely on algorithms to stand in for humans? This thesis proposes a solution to help both users and developers minimize and expose bias in decision-making processes with which algorithms are involved. Research evaluates the severity of consequences resulting from errors of algorithmic biases as well as various perceptions of algorithms and the discourse surrounding their structure and function. My thesis begins by defining the algorithm and understanding how they are developed. The thesis next differentiates between human and algorithmic biases, discussing how algorithms may develop flaws in an attempt to locate steps in the development process where errors can occur. I then speak to how an algorithm can be considered "biased" and why it is important that we use algorithms to recognize biases. Finally, I introduce the Farls Factor—a protocol stating that the amount of human impact should be proportional to the amount of human involvement when implementing algorithms in decision making processes—to determine the most effective way to minimize bias and come to holistic conclusions.

TABLE OF CONTENTS

Dedicated to all of the mentors who have helped me
develop a love for the sciences, humanities,
and everything in between.

INTRODUCTION

Decisions are the foundation for human progression. Regardless of how insignificant they may seem, each and every one of our choices helps to direct our lives in one way or another. Deciding what to have for dinner, for example, may determine how much money someone spends, where they travel, or what culinary skills they may learn that evening. According to researchers at Cornell University, adults make about 35,000 *conscious* decisions each day; 226.7 of these decisions are on food alone.[1] While the study of decision-making has been the subject of a number of intellectual disciplines, it continues to become all-the-more interesting as technology evolves. Even the simplest of tools have led individuals to make decisions more efficiently and effectively. If a person needs to examine a dark space, they will most likely reach for a flashlight before thinking about grabbing a candle since the flashlight is brighter, much more difficult to burn out, and its light can be more easily directed.

Our biases help to shape our decisions, both consciously and unconsciously assisting humans in coming to conclusions. It is important for individuals to recognize that *all biases are not bad*. While the word tends to have a negative connotation, a bias is simply the preference of one factor over another. Humans typically either develop biases from the results of past experiences or inherit them by accepting others' preferences. At their core, decisions are biases in themselves—favoring one option over others. What many do not realize, though, is that most of our biases work unconsciously in the background. If an individual is cautious about spending money, they may not even consider going out to eat

[1] Williams, Ray. "How Neuroscience Can Help Us Make Better Decisions." Ray Williams, December 13, 2018. https://raywilliams.ca/neuroscience-can-help-us-make-better-decisions/.

since they know that they can spend less money cooking a meal for themselves; driving to a restaurant may never even cross their mind.

In recent years, algorithms have become essential tools in the lives of many. They are often used to help humans make decisions as they can be much more efficient and effective in completing what tend to be tedious processes. From sorting data to performing physical labor, their application to various frameworks has resulted in humans growing very reliant on their assistance. These math-based systems may give the illusion of being objective, though they can develop biases as well. Whether it be in the early stages of their development or well-into their commercial use, algorithmic bias has grown in scope as systems continue to become more complex. Algorithmic biases may be intentional depending on its role in a process, though they often go unnoticed until an individual raises a concern.

In order for humans to make more well-rounded decisions, they must attempt to understand the biases that may influence decision-making processes. To better explain the role biases play in decisions where algorithms are involved, an example can be used in which a company is looking to hire ten new employees. If one hundred individuals apply for these positions, the statistical data from their application (age, race, gender, home address, whether or not they have a college degree, etc.) is entered into an algorithm. From here, whoever is in charge of the hiring process sets parameters by which each applicant will be sorted. When the algorithm is finished doing its job, the user is able to more easily view which applicants live closest to the workplace/would have the shortest commute, which are educated in the specific field in which they will be working, and which have the most work experience in this specific field. From this point forward, the top thirty applicants who best fit these chosen requirements are invited to be interviewed, and the top ten most-likable candidates are offered the position.

This may seem like a completely fair process at first, but much more is revealed when the ten employees are chosen. Out of the final candidates, nine are males, eight of those males are white, and one is a white female. It must also be understood that the applicant pool was full of diversity, yet this outcome still occurred. Multiple biases could have been present in this decision if more closely examined; distance from the workplace, for example, may have had a very heavy impact on the decision, yet most of the African American candidates live far from the office due to the result of redlining. In addition, the developer may have assumed that the company wanted to hire the candidates with the most work experience, so they set a threshold to sort by those who worked ten or more years in the specific field. At first glance, if bias is recognized, it cannot be assumed that the neither human nor the algorithm was fully responsible for the error. An examination of the process as a whole is required for an explicit answer on what exactly went wrong.

Approaching decisions, it is important that individuals understand which biases can help them make more effective decisions and which are unnecessary in the context of the process—which should be activated and dismissed. The only way that biases can be fully examined is through a thorough analysis of the decision as a whole and how each factor plays a role in coming to a conclusion. Since human and algorithmic biases differ in both their influence and their origin, a “separation of variables” is required to target how each entity’s biases may influence the other’s. Various methods have been proposed in an attempt to require that algorithmic bias is eliminated before the algorithm can be used commercially, though there has been very little discourse surrounding the ideal balance between human and algorithmic involvement in decision-making processes. By finding this balance, the space for both algorithmic and human error is minimized, as each entity is only required to contribute when their assistance is most important.

I begin by defining the algorithm and understanding the history behind the term. A more thorough explanation is given as to how they differentiate from humans when making decisions. I then discuss the various ways in which algorithms can develop biases and the different types of biases that may be present in programs. Next, an example is presented of a developer that discovered a bias in his own algorithm and a definition is given as to what “algorithmic fairness” should entail. I then speak to how algorithms can be considered “biased” and why it is important that we use algorithms to recognize biases. I then discuss the concept of implicit biases and give an explanation as to why recognizing and monitoring biases in decision-making processes is important. Finally, I introduce the Farls Factor, my proposed solution to a major issue in the current framework for algorithmic bias control. An understanding of how to minimize negative (unintentional or harmful) biases in decision-making processes can only be achieved through examining and understanding the human-algorithm relationship and their influences in processes, themselves.

CHAPTER ONE

From Algebra to Algorithm: A Brief History and Definition

The term “algorithm” can refer to anything from a sequence of numbers to a complete process with a given outcome. Since the ambiguity of the term can lead to confusion in certain cases, its use in the context of this paper must be identified. In the purpose of this thesis, “algorithm” will be defined as a software-based “procedural framework for accomplishing a particular task,” specifically one that “describes a sequence of steps that will lead to an answer.”[2]

“Humans have been devising, altering, and sharing algorithms for thousands of years; they needn’t involve graduate school math or even math at all.”[3] Though altered over the years through public discourse to have a STEM-focused[4] connotation, the term has a very simple meaning: “an algorithm is a set of instructions to be carried out perfunctorily to achieve an ideal result.”[5] Individuals perform algorithms throughout their daily lives, often forming routines around similar courses of action.

Each person’s unique processes to complete tasks may follow the same set of steps or slightly differ; whether steps be added, removed, or reversed, humans structure their lives around the algorithms they know and create new ones based on past experiences. Take washing a car, for example. One person may decide to wash with soap, rinse with water, and dry with a towel. Another may decide to wax their car, though, adding one or two more steps

---

[2] Brock, Kevin. "Rhetoric and the Algorithm." In Rhetorical Code Studies: Discovering Arguments in and around Code, 33-70. ANN ARBOR: University of Michigan Press, 2019. Accessed October 20, 2020. http://www.jstor.org/stable/j.ctvndv9pc.9.

[3] Steiner, Christopher. *Automate This: How Algorithms Took Over Our Markets, Our Jobs, and the World*. New York: Portfolio/Penguin, 2013.

[4] STEM – Science, Technology, Engineering, Math

[5] Ibid.

to the procedure. While processes can be identified throughout human life, algorithms still have their own, unique meaning—one that can only be discovered by looking back at the origin of the term.

The name “algorithm” is derived from the name of ninth-century mathematician Abu Abdullah Mohammed ibn Musa al-Khwarizmi. Introducing Hindu-Arabic numerals to the West, he quickly replaced the standard of what used to be Roman numerals. This new number system, including the decimal point, was the foundation for our current number system as the Arabic characters were slightly altered to resemble our familiar digits (1, 2, 3, 4, etc.).

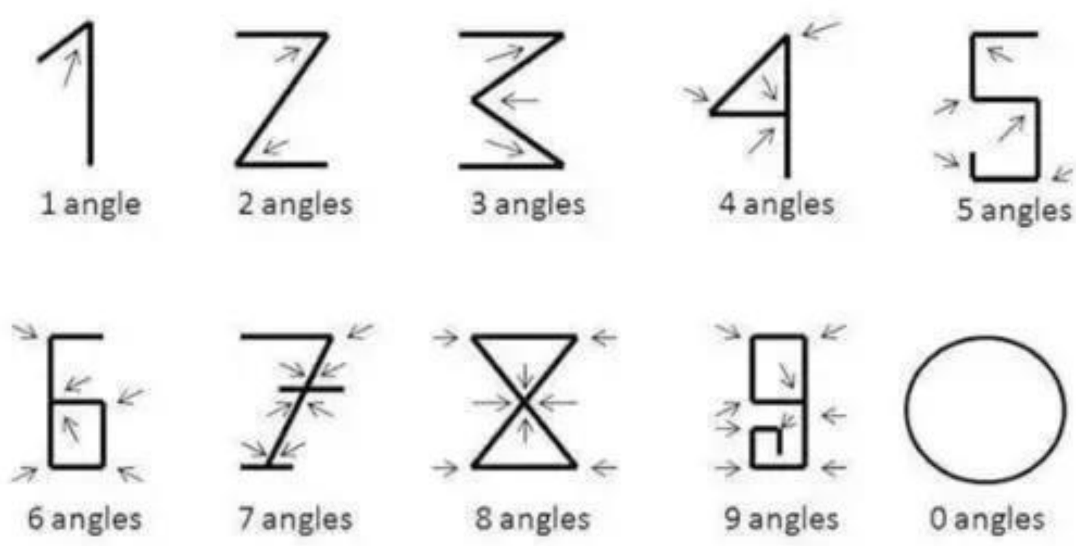

Figure 1: Numerals that resemble those found in al-Khwarizmi’s writings. Their values are assigned by the number of angles each possesses. This style of numeral is the foundation for those used today. (Daniel E. Otero, “The Birth of Algebra”)

Famous for his book *Al-Kitab al-Mukhtasar fi Hisab al-Jabr wa l-Muqabala*[6] (The Compendious Book on Calculation by Completion and Balancing), “it is through al-

[6] Published in about 830.

Khwarizmi's writing that the algorithm becomes codified as a procedural framework whose functionality is articulated through a specific grammar."[7] According to Daniel Otero, this book represents "the first true work of algebra ever written," describing rules for solving problems with unknown quantities.[8] When Latin scholars translated the writings, al-Khwarizmi's name was altered to become *algorismi*; it is from this term that the word "algorithm" was born.

Algorithms are inherently unique from other processes; "by constructing a framework to which a mathematician could adhere in order to solve discrete problems, the capabilities of symbolic systems to reflect logical procedures were clearly articulated."[9] Commonly viewed as being mathematics, algorithms utilize a specific framework through which they complete tasks. At the end of the day, though, "the specific field in which a scholar or practitioner works has some influence upon how the scholar is likely to approach algorithms and their potential for certain tasks."[10] Differentiation in framework can lead to varied algorithmic potential, altering the overall effectiveness of procedures in the process.

It is also important to understand how these systems are constructed. It is obvious that algorithms are made up of steps. In computer programs, these steps are visualized as lines of code. Many modern-day computers operate on what is known as Boolean logic, primarily basing algorithmic grammar on statements of "true" and "false." These operations permit mathematical calculations to act as the foundation of computer-based algorithmic procedure, "allowing a software program to execute particular computational tasks so as to express a

[7] Kevin Brock, "Rhetoric and the Algorithm", 34.

[8] Otero, Daniel E. The Birth of Algebra, 2000. http://cerebro.xu.edu/math/math147/02f/algebra/algebra.html.

[9] Kevin Brock, "Rhetoric and the Algorithm", 34.

[10] Ibid, 37.

relevant "output" body of data."[11] In his book *The Pattern on the Stone*, Daniel Hillis describes algorithmic procedure as being "all about performing tasks that seem to be complex (like winning a game of tic-tac-toe) by breaking them down into simple operations (like closing a switch)."[12] By paralleling the opening and closing of circuits through Boolean logic, algorithms "waver on this line between the complex and the simple, and crossing that line is what computation is all about."[13]

Similar to humans, algorithms use both deductive and inductive reasoning to come to conclusions. Deduction is "a process of reasoning that starts with a general truth, applies the truth to a specific case, and from those two pieces of evidence, draws a specific conclusion about the specific case."[14] For example, if all men are mortal and Jim is a man, it can be assumed that Jim is mortal. Induction "infers a general conclusion based on individual cases, examples, specific bits of evidence, and other specific types of premises."[15] An example of induction would be that if you break out when you eat peanuts and that is a symptom of being allergic, you can conclude that you are allergic to peanuts. Humans also use these two methods that algorithms attempt to replicate; what primarily sets them apart, though, is the algorithm's avoidance of external pressures—emotions. Love, hate, joy, and resentment are only a few of these forces that may influence human decisions.

Algorithms are assigned to follow sets of steps though they cannot be perceived as being fully objective. At their core, algorithms are frameworks created by developers to complete processes—they are tools, not truths. Both deductive and inductive reasoning are

---

[11] Ibid, 37.

[12] Hillis, W. Daniel. "The Pattern on the Stone: The Simple Ideas That Make Computers Work." New York: Basic Books, 2015.

[13] Daniel W. Hillis, "The Pattern on the Stone: The Simple Ideas That Make Computers Work."

[14] University, Montana State. "Induction vs. Deduction." Montana State University Billings. Montana State University, n.d.

[15] Ibid.

not foolproof. Algorithms can understand that certain conclusions can be made based on specific parameters, but the data that they output must be interpreted in the context of the program. If too much trust is given to algorithmic objectivity, their possible flaws in computation may be overlooked. Not only could errors be disregarded, but the tool becomes much less useful in discovering how an answer is constructed. While algorithms have the potential to outperform humans in some cases (facial recognition, for example), they assist in decisions as a means to an end—they do not define the end, itself.

Algorithmic complexity does not ensure a more truthful result. In his writing "What is an Answer?," Professor Herbert S. Wilf takes note that "sometimes the "answer" is presented as a formula that is so messy and long, and so full of factorials and sign alternations and whatnot, that we may feel that the disease was preferable to the cure."[16] Algorithms are written to make processes more time-efficient, allowing technology to compute large quantities of data in an attempt to eliminate possible human error. Wilf argued that "the quantitative criterion is the computational complexity: the amount of work required to get an answer," understanding that more complex programs will lead to more well-rounded and efficient procedures.[17] This is crucial to understanding algorithmic potential, as their primary purpose of shortening a procedure's length of time can lead to certain human qualities being dropped in the process; "as faster algorithms for listing objects become available, a proposed formula for counting the objects will have to be comparably faster to evaluate."[18]

By understanding the term as a series of steps, it is much easier to determine where an issue may occur. Holistic views of algorithms are important; it is crucial that the purpose of a

---

[16] Wilf, Herbert S. "What Is an Answer?" The American Mathematical Monthly 89, no. 5 (1982): 289-92. Accessed October 21, 2020.

[17] Ibid, 289.

[18] Herbert S. Wilf, "What is an Answer?", 289.

procedure is understood before it is thoroughly examined. What must also be understood is an algorithm's approach to processing data in relation to that of a human; the processes are vaguely similar, though cannot be directly evaluated. While both loosely follow a type of true/false logic, the differences in decision-making frameworks can lead to mistakes in programs, resulting in later-formed biases. This paper will primarily focus on computer-based algorithms working on the foundation of Boolean-based logic. Understanding the framework and limitations of this system can more clearly and effectively explain why, and how, algorithmic biases exist.

CHAPTER TWO

Biased Backgrounds: An Overview of Algorithmic Biases

In a Forbes article written in March of 2019, the Forbes Insights Team clearly outlined what developers believed may lead to algorithms being biased. Ruchir Puri, a chief scientist at IBM Research, says that "'up to 94% of AI projects end up sitting on the shelf because there are issues of trust and transparency around them.'"[19] This shows that, even though many public computer programs have occasional biases and flaws, many programs are never even completed or released due to the worry that they may have the same (if not worse). The team also went on to interview Bruno Masonnier, CEO of AI company AnotherBrain; this company created an emotion-detecting robot named Pepper which has allowed AnotherBrain to see certain biases arise first-hand during the development process. Through his research, Masonnier has identified the three main sources of bias in AI algorithms with which many scientists also agree: the training data set, constraints given by the developer, and the principle of the AI algorithms themselves.

Method 1: Data Bias

The Forbes article, along with many others, agree that "[d]ata bias is, by far, the biggest problem AI developers face," since much of the training data can lead an algorithm to favor one result over another.[20] Since underfunded projects can lead to smaller sample sizes when training machine learning software, data bias can lead to inconsistencies in the

[19] Team, Insights. "Forbes Insights: Managing The Ethics Of Algorithms." Forbes. Forbes Magazine, June 5, 2019. http://www.forbes.com/sites/insights-intelai/2019/03/27/managing-the-ethics-of-algorithms/.
[20] Ibid.

algorithm as time passes. Most of the time, developers are not aware of data bias until something goes wrong. Different types of data bias can lead to varied issues throughout different programs. There are seven main types of data bias: sample/selection, exclusion, measurement, recall, observer, social, and association bias.

1) Sample, or selection, bias occurs "when a dataset does not reflect the realities of the environment in which a model will run."[21] Creating programs requires more than just typing code; it is important for the individual developer to understand the context in which their algorithm will be used, regardless of whether it is a solo or team project. The role of the programmer requires research to be performed before anything is put into action, for an understanding of context will lead to more reliable content. If an algorithm is created to predict health insurance costs though is tested using data from people with similar occupations, for example, the system may assume that every individual has the same portion deducted from their payroll (given that insurance is provided by their employer). Developers must be aware of how and by whom their algorithms are being used or the software can incorrectly favor certain aspects over others.

2) Exclusion bias is most common as it deletes data thought to be unimportant.[22] Similar to selection bias, the developer must be familiar with which aspects of the algorithm are most important so that proper variables can be emphasized. When creating complex programs, hundreds if not thousands of variables are often used to store specific data points. If they aren't properly understood in relation to the use of the program, some important data may be forgotten or get lost in the rest of the code. Though variables can't be "deleted," their

[21] Lim, Hengtee. "7 Types of Data Bias in Machine Learning." Lionbridge AI, August 5, 2020. https://lionbridge.ai/articles/7-types-of-data-bias-in-machine-learning/.

[22] Ibid.

importance can be misunderstood by the programmer, leading to unintentional flaws in computing. Models can inaccurately portray information if exclusion bias is present based on unrepresented groups having much smaller amounts of data existing in certain programs.

3) Measurement bias occurs when "the data collected for training differs from that collected in the real world, or when faulty measurements result in data distortion."[23] It is important for developers to use accurate data when testing their algorithms. While most simple computing programs do not require test data to teach the system, artificial intelligence heavily relies on off of the differentiation of human inputting to properly produce a desired outcome. If training a voice-recognition software such as Siri, for example, mumbled speech must be tested along with sentences spoken with perfect pronunciation. If only one or the other is used to train the program, it may often fail to perform simple tasks. Testing algorithms with real-world simulated data is important; if not given the correct resources, programs could result in acting much different than intended.

4) Recall bias is a kind of measurement bias, arising when similar types of data are labeled inconsistently.[24] Returning to the voice-recognition example, if trying to translate language through speech, languages with similar vocabularies and grammar rules—Spanish and Portuguese, for example—must be distinguished in one way or another. By programming an algorithm to search for key words that differentiate the two languages, a program would be able to successfully follow through with its translation. When similarities in varying entities are ignored, though, problems can arise and biases may occur. If two similar pieces of data are sorted into two separate categories by the developer, the algorithm could get confused and begin to inaccurately label certain inputs.

[23] Hengtee Lim, "7 Types of Data Bias in Machine Learning".
[24] Ibid.

5) Observer bias is “the effect of seeing what you expect to see or what to see in the data.”[25] Human biases exist right beside those present in algorithms, so an individual’s view on a system may lead them to create a procedure in which *their* desired outcome is reflected. When creating a system of any kind, it is important for a person to evaluate their own views in relation to those visible through its steps. Since the developer has control over the way the program functions, they may feel tempted to process data in a certain way. If a researcher is expecting a certain output, they may intentionally or unintentionally alter the algorithm to favor that output. This is not uncommon and could lead to various major issues if the program is publicly released.

6) Social bias is not traditional data bias, though it has occurred often in recent AI technology. While Lim’s article states that the sixth type of data bias is “racial bias,” a more overarching title must be used to include algorithmic bias towards other social constructs. Social bias occurs “when data skews in favor of particular demographics.”[26] This is often seen in facial and speech recognition technologies which fail to recognize people of color due to various factors. Social bias differs from the rest as it can be a product of the other types of data bias—consistency in data across social groups may lead to this issue, not the groups, themselves. Whether an algorithm is trained with data particular to a certain demographic, has no way of differentiating between similar, yet distinctly different demographics, or being altered due to the developer’s views, social bias has led to various issues in algorithmic use. If social bias is present in a hiring algorithm, for example, an entire company’s societal view can be portrayed by a faulty program. Social bias can be eliminated by a thorough

[25] Hengtee Lim, “7 Types of Data Bias in Machine Learning”.
[26] Ibid.

examination of the algorithm and how it is favoring certain parameters in the context of the process.

7) Association bias occurs "when the data for a machine learning model reinforces and/or multiplies a cultural bias."[27] If input data suggests that all men are doctors and all women are nurses, for example, machine learning may assume that male nurses and female doctors do not exist. Association bias is most notably seen as having gender bias due to factors such as the example above.[28] It can occur surrounding many different scenarios, though; its occurrence into algorithms is similar to that of racial bias whereas its inclusion cannot be traced back to a single, generalized error. By eliminating association bias throughout algorithms, cultural stereotypes could possibly be eliminated or reversed based on the space in which the algorithm is being used.

While some of these favoring techniques may not be important to the algorithm's purpose, they may go overlooked, only to reappear later in its use. It is important for developers to understand and investigate the appearance of biases in their programs. Whether intentional or unintentional, small errors in computing data due to a misunderstanding of the algorithm's purpose can lead to massive shifts in expected outputted data. As data biases are often revealed later in an algorithm's use, it is crucial that developers understand the importance in attempting to eliminate them early in the development procedure.

[27] Hengtee Lim, "7 Types of Data Bias in Machine Learning".
[28] Ibid.

Method 2: Constraints Around Data

The second source of bias is in constraints around data, which "can also lead an AI to make a decision incorrectly, causing decisions to be based on irrelevant information."[29] This does not only have to deal with bad data, though, but also acceptable and well-thought-out boundaries. The example used to understand the constraint scenario is one in which a network designed to detect a fish frequently became confused with pictures of apples since it was programmed to find a hand holding an object in the photo (as many love to take pictures holding fish); this algorithm also failed to recognize pictures of fish in tanks due to the lack of a hand present in the photo. This issue paired with that of initial training data leads to issues such as the fish dilemma—if only pictures of people holding fish are used to train the algorithm to find one, other photos of fish will go unnoticed due to the lack of certain features.

Creating too many guidelines for a program to follow can be dangerous, but the simplicity of an algorithm can also lead to certain factors being left out. Boolean logic is very straightforward—"trues" and "falses" represent human binary decision-making by understanding that two choices are either the same or different based on their more complex factors. When programming a software to follow a certain set of steps, it is not self-aware of what it's tasked to do. Constraints that are too general may lead to issues in sorting data, for example, putting the fish and the apple (as used in the previous example) in the same category even if the hand constraint is not present. There must be a healthy balance between what an algorithm must consider and what it must not, as extremes to either side can lead to inherently problematic biases.

[29] Insights Team, "Forbes Insights: Managing the Ethics of Algorithms".

A prominent reason for increases in data constraints is the human's desire to give the algorithm more than it can handle. One example of an increase in algorithmic intelligence is the evolution of CAPTCHA.[30] CAPTCHA tests are primarily used as an extra layer of security when attempting to log into websites and have been continually increasing in difficulty over time. While the software was originally developed to prevent "bots" of spammers from quickly accessing websites, "the bad guys' computers have been getting smarter and, well, people have not."[31] CAPTCHAs must continue to include data that would easily trick an algorithm into failing for the test to succeed, leading spammers to add even stricter constraints to their bots; more explicit parameters will help algorithms identify more specific features in photographs, sequences of letters, etc. The CAPTCHA-spammer relationship is important, though, as it acts as a visual representation of the continual drive for algorithmic perfection. The more specific algorithms become, they will most likely favor certain features/methods over others. Developers must only use necessary constraints when creating decision-making processes in order to avoid an overabundance of user error.

In "Iterate Analysis of Competing Hypotheses to Overcome Cognitive Biases in Cyber Decision-Making," A. Lemay and S.P. Leblanc—cyberattack detectors from Montréal and Kingston, Canada, respectively—discuss the effect of human cognitive biases on the performance of cyber incident response.[32] Cyberattacks have greatly increased in both frequency and scope in recent years. An effort has been made to not only eliminate these attacks, but appropriately improve the tactics and techniques around incident response.[33]

---

[30] Completely Automated Public Turing test to tell Computers and Humans Apart. Developed in 2000 at Carnegie Mellon University.

[31] Burling, Stacey. "CAPTCHA: The Story behind Those Squiggly Computer Letters." Phys.org, Phys.org, June 15, 2012. https://phys.org/news/2012-06-captcha-story-squiggly-letters.html.

[32] Lemay, A., and SP Leblanc. "Iterative Analysis of Competing Hypotheses to Overcome Cognitive Biases in Cyber Decision-Making." *Journal of Information Warfare* 17, no. 2 (2018): 42-53. Accessed December 18, 2020.

[33] Ibid.

With attention given to these improvements, they continue to fail time and time again as cyberattack methods constantly change: “In the M-Trends Report, detailing trends from 2017, Mandiant (2018) reveals that 38% of victims still require notification of a third party to know that they have suffered a breach.”[34]

The first bias examined is the error in an algorithm’s “base rate”—the rate by which an algorithm detects if an event is an incident or not. They strictly define an ‘incident’ as “a cyber event that requires incident response: the observation of an unusual packet traversing the network, the reception of an email with an anomalous attachment, or the generation of an alarm by a security sensor.”[35] In their study, Lemay and Leblanc mention that an algorithm can have a 99% success rate yet fail if the base rate is set too low. The proper base rate for cyber-detection algorithms cannot be uniform, as different conditions may lead to different results. Through extensive testing, scientists can attempt to understand the trends of these errors, but it is up to the developer to decide how to make the process the most efficient without being too constrained. Errors in data constraints are typically seen in cybersecurity algorithms in which precision is of the utmost importance. By creating algorithms that do not try to exceed their limits, constraints can be regulated and processes can become more efficient.

Method 3: Principles of Algorithms

The principles of algorithms can lead to biases, for “the code for a neural network is unlike that of traditional computer code, and algorithmic bias can be intentional or (more

---

[34] Ibid.

[35] A. Lemay, "Iterative Analysis of Competing Hypotheses to Overcome Cognitive Biases in Cyber Decision-Making."

commonly) completely accidental."[36] Since humans are inherently biased with regards to certain aspects of life (not only negatively speaking), unnoticed biases can be put into algorithms without the developer thinking twice about them. The common "AI gone wrong" scenario is often a product of this source, for some programs "were simply designed to look for the wrong things."[37] Human biases are typically the beginnings for algorithmic biases—especially if a single developer is employed to create a program—since their own personal thoughts and motives are often unconsciously added into software.

It is through intended algorithmic principles that the largest issue regarding algorithmic biases arises. Since algorithms deductively and inductively reason based on set guidelines, they strictly do what they are told without any external influence during the decision-making process. As stated early, people can be emotionally pressured—whether preceding or during the steps of a procedure—which can lead to an alteration in their decision. For example, if a college admissions officer has a son who plays tennis and comes across an application of a student who also plays tennis, they may be more favorable toward that candidate. While algorithms can be programmed to simulate these inherent human qualities, they would never be able to fully parallel human decision-making since some factors cannot yet be replicated.[38]

---

[36] Insights Team, "Forbes Insights: Managing the Ethics of Algorithms".

[37] Ibid.

[38] There are ongoing discussions around the concept of how the human mind works and if algorithms can be created to replicate it. Since there are still very few biological answers as to what thoughts are and how we process them, a fully effective "algorithmic mind" will most likely not be developed in the near future.

CHAPTER THREE

What's in the Math?: Determining Bias

It is understood that algorithms can become biased in one way or another—this is not something widely debated. What may seem to be a problem, though, is how people consider an algorithm to be biased or not. What given variables are there that decides if it is? How can biases be compared between different algorithms? It is not effective to generalize and list a certain set of "rules" a program must pass before being declared "bias-free," as different programs often have different issues that arise. In his paper *The ethics of algorithms: Mapping the debate*, Brent Mittelstadt writes that "determining whether a particular problematic decision is merely a one-off 'bug' or evidence of a systemic failure or bias may be impossible (or at least highly difficult) with poorly interpretable and predictable learning algorithms."[39] It must be understood that, through this quote, Mittelstadt is claiming that it is difficult to fix biases in poorly *developed* algorithms—not that it is difficult to fix *all* biased algorithms. This should not stop researchers and developers from identifying and helping to eliminate possible errors in their programs, though. To further understand how algorithmic biases can be identified, it is important to examine an example of a developer who ran into this very problem.

Nicholas T. Young, a writer for *Scientific American*, was assigned to create an algorithm as a first-year graduate student at Michigan State University. The program was supposed to be "a machine-learning algorithm to analyze a survey sent to United States

[39] Mittelstadt, Daniel Brent, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society*, 2016.

physics instructors about teaching computer programming in their courses."[40] Young decided to send a survey to physics teachers across the country and use an algorithm to predict whether or not they taught programming in their respective courses. While this was by no means a difficult task, he created a program in Python (a popular computer language) to identify "whether the instructors taught programming and which questions on the survey were most useful in making that prediction."[41] Once he began to run the program, he noticed that the multiple-choice answers did not play into the predictions—only the answers to the free response questions; the algorithm favored one part of the process over another. This initially stumped Young, so he "analyzed those questions using a different technique and didn't find any differences between the instructors who taught and did not teach programming!"[42] He knew that there was an error because he "visualized [his] data and saw that [his] algorithm's predictions were not aligned with what [his] data or previous research said."[43]

In his article, Young brings up another example, explaining a more serious algorithmic bias present in a healthcare program. When using an algorithm to "find patients who may be good fits in a 'high-risk care management' program, white and black patients [were] identified as having equal risk [even though] the black patient was sicker than the white patient."[44] Since the algorithm was predicting heath care costs rather than illness severity, inaccurate results were outputted. The bias was only recognized because of the severity of the error in the results. Both the developer and the user of the algorithm could

[40] Young, Nicholas T. "I Know Some Algorithms Are Biased-Because I Created One." Scientific American Blog Network. Scientific American, January 31, 2020. https://blogs.scientificamerican.com/voices/i-know-some-algorithms-are-biased-because-i-created-one/.

[41] Nicholas T. Young, "I Know Some Algorithms Are Biased-Because I Created One."

[42] Ibid.

[43] Ibid.

[44] Ibid.

comfortably say that it was biased and could see the error when looking through the source code. In his own experience, if Young would not have compared the predictions to his own research, he never would have come across this issue. As with most discrimination issues, bias may not be obvious until someone involved in a process expresses concern about its presence.

Sometimes errors like Young's are not as obvious; it is very difficult to identify biases in some cases, or even come to the conclusion that an algorithm is biased if an output is not as expected. Developers must check, then, if what is thought to be the error is not only systematic but also repeatable. When the program is run multiple times with similar, unfair outcomes, then it may be safe to say that the algorithm is considered biased. Even though favoring towards a single group of data may be small, it is important to run the algorithm multiple times to identify consistency.

Various interpretations can be developed regarding what a negative bias truly is, though a static framework must be set up in order to determine its presence in a process. In their article "On the Legal Compatibility of Fairness Definitions," Alice Xiang and Inioluwa Deborah Raji discuss the current status of "fairness" in machine learning discourse; they state that "[machine learning] fairness definitions must align with their corresponding legal definitions."[45] Their article covers common terms from anti-discrimination law and how, in machine learning fairness literature, they are typically oversimplified. "Discrimination," for example, "is often presented as an unjust correlation between protected class variables and some metric of interest," though consideration of The Civil Rights Act of 1964, Fair Housing Act, etc., show that these terms must have further specified definitions depending on the

[45] Xiang, Alice, and Inioluwa Deborah Raji. "On the Legal Compatibility of Fairness Definitions." *Partnership on AI*, November 25, 2019, 1–6. https://arxiv.org/pdf/1912.00761.pdf.

context of the issue.[46] Since negative algorithmic biases ultimately lead to procedural errors similar to negative human biases, algorithms must be held accountable to the same legal definitions already put in place.

Many debates surround how to eliminate biases from programs, though some algorithms are created with them in mind. The typical facial recognition software does not need access to past records in order to make a decision—it uses on-site collected data to answer the binary "yes, this matches" or "no, it does not"—but algorithms created for tasks such as hiring, advertising, or credit reporting need to have a collection of pre-collected and determined data. Since companies and other institutions have access of past records and can collect information from various data sources, an understanding of biases is crucial in making these types of decisions. To fully comprehend the different factors of a decision, the more data that is examined will lead to a more holistic understanding of the decision, itself.

"When sensitive information is used responsibly and proactively, ongoing discrimination can be made transparent through data-checking processes that can ultimately improve outcomes for discriminated-against groups."[47] A lot of the security surrounding these situations leads many to ponder certain questions: when is it appropriate to collect and use sensitive data, when does it cause harm, and when does it prevent it?[48] Since "many people and institutions [in the United States] discriminate due to prejudice, and still more make decisions tainted by past prejudices," an understanding of the origin of negative biases is important in helping to eliminate them.[49] When Amazon's hiring algorithm consistently

[46] Alice Xiang and Inioluwa Deborah Raji, "On the Legal Compatibility of Fairness Definitions," 2.

[47] Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications." Journal of Information Policy 8 (2018): 78-115. Accessed October 21, 2020.

[48] Ibid.

[49] Betsy Ann Williams, "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications".

favored men over women, not only did the program's issue become apparent, but the company itself gained a reputation of being "sexist" due to the traits it searched for in an employee.[50]

Algorithmic biases cannot be directly compared to human biases when initially discovered. Due to the various different methods of how these prejudices can arise in computer programs, the algorithm may have not been programmed with the intent of simulating human thought. It truly depends on the way in which the algorithm is considered "biased" before placing blame on the developer, for machine-learning can develop these favoritisms on their own.[51] It is not always obvious when humans are being biased. During processes in which individuals do not share their work with others, personal prejudices could be the cause of skews in data. While sorting through information, a person may favor one aspect over another, "unfairly" eliminating what could be important data points. When someone receives the result of their work, they may realize that the given conclusion is slightly inaccurate but has no way of proving the researcher's thought process throughout the procedure. With an algorithm, though, the given choices can be seen throughout the program, making it easier to identify changes in selection. It can be seen as a map, for the developer is able to much more easily pinpoint where a bias could have appeared. Since human biases can develop through a number of factors, they sometimes cannot be easily identified if not stated, differentiating them from those that algorithms develop.

---

[50] Dastin, Jeffery. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women." Reuters. Thomson Reuters, October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. Amazon's failed experiment to create an automated hiring algorithm was filled with implicit biases against women without any intention to be sexist.

[51] As covered in Chapter 2, machine-learning algorithms can unintentionally develop biases through inappropriate and uniform training data or other accidental methods. When creating a program that changes its thresholds over time based on the trends that it sees, it is possible that it may begin to ignore data that is either consistent or nonexistent in relation to the trend.

Decision-making processes can be broken up into three steps: input, assortment, and selection. In a human-algorithmic hybrid procedure, it is clear which entity controls each of these three steps—humans are in charge of inputting data, algorithms are created to help sort this data based on the parameters set by the user, and, once the data is correctly organized, the human reenters the process to make the final decision(s). By strategically examining human and algorithmic involvement in these stages, the biases that arise in each can be traced back to the entity from which they originated. As stated earlier, developers are typically responsible for algorithmic biases, though different approaches must be taken to identify the biases of algorithms and humans.

When determining bias in the "input" stage of a decision-making process, favorability is more likely to be seen through set parameters rather than the data, itself. If there are five factors by which a user can choose how data is organized, the human will take advantage of these parameters in an attempt to receive the information that they believe is necessary in making a decision. While some parameters that the user sets may be crucial to obtaining a desired outcome, others may be completely unnecessary. The algorithm can be taken advantage of (consciously or unconsciously) in an attempt for the human to receive a certain outcome or disregard data without certain attributes. While bias cannot be determined until the data is sorted, a late-stage examination of the process may reveal biases from the very first stage.

To understand algorithmic biases within data assortment, the definition of the algorithm given in Chapter 1 must be revisited: a software-based procedural framework for accomplishing a particular task, specifically one that describes a sequence of steps that will lead to an answer. By breaking up software into its individual lines of code, developers should be able to identify algorithmic biases (if they are present). Sendhil Mullainathan, a writer for the New York Times, stated that "discrimination by algorithm can be more readily discovered and more easily fixed [compared to human biases]" after having performed extensive research on algorithmic assistance in reducing discrimination.[52] Whether it be an offset conditional, an altered threshold, or another analytical issue, trained programmers should have the skills to be able to detect these issues when they occur. To understand if the issue is or is not a bias, though, developers must look for repeated patterns in data assortment that are not directly related to the data that is being computed; not all algorithmic errors are biases. Data biases are not always the algorithm's fault either, as the developer may accidentally insert certain restrictions based on *their* perception of the situation the algorithm is being used for. A multitude of these issues may arise even in the late stages of the development process, but, if given a proper framework to examine the program, most algorithmic biases can be detected and resolved with the help of a good developer.

Finally, bias may occur once the data is sorted and the human makes the final decision. If the entire process is fair up to this point, trends may be seen across which data points the user chooses, signifying the action of either an explicit or implicit bias based on the severity of the association. Again, the suspected bias must be repetitive through various choices before it can be determined as a truthful favorability or not. When a decision-making process is revealed to be biased in one way or another, blame can be immediately shifted onto neither the human nor the algorithm. Through each individual stage, various errors may

[52] Mullainathan, Sendhil. "Biased Algorithms Are Easier to Fix Than Biased People." The New York Times. The New York Times, December 6, 2019. https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

occur due to a number of factors, and the ability to return to the medium by which the decision is made and examine the system as a whole allows for a much easier method of correctly determining bias in decision-making processes.

CHAPTER FOUR

What's the Matter?: How to Approach Bias in Decisions

Algorithms are set to run strictly through the set of methods the developer provides it with; unless programmed to do so, they cannot tell if they favor one set of data over another when assigned to choose. Discourse surrounding human biases typically holds the connotation that a biased individual is fully aware of their favoritisms. In the early twenty first century, though, resulting thoughts from the Freudian revolution had challenged the stereotype that human behavior is largely under conscious control—there is another actor at work. In July of 2006, Anthony G. Greenwald and Linda Hamilton Krieger submitted an article to the California Law Review that outlines the scientific foundations of the newly proposed idea of "implicit bias." Their definition states that "the science of implicit cognition suggests that actors do not always have conscious, intentional control over the processes of social perception, impression formation, and judgement that motivate their actions."[53] Algorithmic involvement in decision-making processes can help both developers and users understand what conscious and implicit biases may exist within programs and, furthermore, regulate which ones are needed to make certain decisions.

As an example of implicit cognition, Greenwald and Krieger explain a memory experiment in which subjects were asked to pronounce a list of names; while some of the names were recognizably famous, others were not. On the second day of the experiment, the subjects were asked to judge whether names on a different list were famous or not. The catch: half of day one's non-famous names were on day two's list. The data collected on day

---

[53] Greenwald, Anthony G., and Linda Hamilton Krieger. "Implicit Bias: Scientific Foundations." *California Law Review* 94, no. 4 (2006): 945-67. Accessed January 31, 2021.

two showed that more of the non-famous names were incorrectly judged as being novel, simply due to the fact that the subjects had heard and remembered these names from the day before.[54] To explain this outcome, Greenwald and Krieger state that "[s]ubjects presumably go through a mental process resembling the following: 'This name seems familiar. Why is it familiar? Perhaps it's famous.'"[55] Implicit cognition can lead to implicit attitudes and implicit stereotypes—two incredibly relevant issues surrounding biases.

Implicit attitude indicators are actions that "indicate favor or disfavor toward some object but is not understood by the actor expressing that attitude."[56] Someone may vote for a candidate knowing nothing but their name, simply because the candidate's initials are the same as the voter's. While this example may seem impractical and farfetched, an observer can understand that the voter is self-favorable—by seeing similarities between an unfamiliar entity and a favorable one, the unfamiliar is indirectly favored. When an algorithm is given parameters to sort data, implicit attitude indicators can be examined to understand the user's specific interest in the data. Parameters that are offered but left untouched are seen as least important to the user and their motivations, while set parameters expose their explicit attitudes.

In this article, a stereotype is defined as "a mental association between a social group or category and a trait."[57] Stereotypes can sometimes reflect a reality, but it is not always the case. When the correlation between association and expression is completely false, though, stereotypes become interesting. Take, for example, the relationship between age and driving. If, hypothetically, 10-15% of people over the age of seventy drive at slower speeds on the

---

[54] Anthony G. Greenwald and Linda Hamilton Krieger, "Implicit Bias: Scientific Foundations", 947.
[55] Ibid, 948.
[56] Ibid, 948.
[57] Ibid, 949.

highway, but only 5% of younger individuals do so as well.[58] If these statistics were exact, there would be a direct association between age and driving speeds, but the stereotype that "elderly individuals drive slower" would only account to a small portion of the population. Implicit stereotypes can be seen through the fame experiment explained earlier. As they collected the data, Greenwald found that "the false-fame effect was substantial when the pronounced names were male, but was noticeably weaker when the names were female."[59] The experiment was not set up with any influence for subjects to associate fame and gender, though the results showed that this seemed to be true; at the end of the day, the subjects had no intention of making this connection.

Implicit attitudes and implicit stereotypes are indicative of the fact that biases are the foundation of decision-making processes. The only possible way to fully eliminate bias is through a random selection which, in most situations, is fairly impractical. In order for negative bias to be analyzed and eliminated, though, the user must have the intention of creating a fair decision-making process. Laws exist which attempt to ensure that corporations are as unbiased as possible, helping to regulate entities which may be *too* grounded in their discriminatory conscious biases. There exists a much greater importance in the issue at hand, though, especially when algorithms get involved.

The initiative towards understanding our biases and attempting to minimize those that create unfair decisions is grounded in preserving identity. Allowing biases to get in the way of complex decisions leads to an oversight of many factors that play a role in choice. Creating a hierarchy of factors is important in making decisions; in most decision-making processes, some aspects are going to be much more important than others. Each factor needs

[58] Anthony G. Greenwald and Linda Hamilton Krieger, "Implicit Bias: Scientific Foundations", 949.
[59] Ibid, 949-950.

to be understood in its context, though, for a decision to be as “fair” as possible. As discussed in Chapter 1, false objectivity is one of the most deceiving influences in conscious thought. Scales are often created to measure different attributes. The IQ test is a common measure of intelligence, but an analysis of its creation and its methods of examination show that it has very little relevance in comparing individuals’ thinking skills.

The implementation of algorithms as tools in helping to make decisions has allowed for a certain level of objectivity that humans can use to their advantage. As mentioned in Chapter 3, algorithms can be viewed as “maps” of decisions, themselves. The various parameters needed to create and sort data points can be implemented and viewed by developers in order to set up a “physical decision system.” Even the act of creating an algorithm specific to a certain framework requires consideration of the decision as a whole. Algorithmic involvement can also help to rethink a faulty system. If the hiring algorithm used in the introduction was a process that used to be performed strictly by humans, the creation or adoption of an algorithm forced the user to determine what they wanted to include or remove from the application process—they had to examine their biases towards the system in order to examine the biases within the system.

Understanding the roles that various biases play in making decisions also allows for a promotion of diversity. By eliminating negative biases from decisions, diversity of thought is much easier to accomplish. Each factor that plays a role in a decision has something to say about the entities being decided upon. If the company that commissioned the hiring algorithm was a local advertising company, maybe location from the workplace was actually more important than years of experience. In the context of the decision, it can be better understood that the company would benefit by employing people who lived in different areas of the city—people who have had different experiences in the region where the advertisements would be posted.

In the next chapter, I take a much deeper look into a decision-making process and how each factor influences what may seem like a fairly simple decision. Examining both human and algorithmic biases leads to processes becoming more fair and decisions being more diverse in thought, though, when entities are reduced strictly to data points for algorithms to sort, it is of the utmost importance that their identities must be preserved. In order to create an effective decision-making process, the person in the deciding role must understand their choices as best as possible. Some factors may seem more important than others, though a consideration of each individual factor, regardless of how many there are, will help to arrive at a much more fulfilling conclusion. Examining and minimizing *how* and *why* we make decisions alongside algorithms will not only benefit us, but also the decision as a whole

CHAPTER FIVE

The Farls Factor: A Hybrid Approach to Decision-Making

On April 10th, 2019, the Algorithmic Accountability Act of 2019 (H. R. 2231) was introduced by Ms. Yvette Clarke in the House of Representatives. The bill was proposed "[t]o direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments."[60] The lack of regulation on algorithmic processes had finally been given much-needed attention in an attempt to avoid detrimental impacts before they could occur. Similar to environmental impact statements required for construction projects, many believed that algorithms needed statements to "document the goals of the program, data quality control for input sources, and expected outputs so that deviations can be detected."[61] By taking a step back and examining a process as a whole, possible biases in algorithmic decision-making would be much easier to identify.

One of the largest flaws in the Algorithmic Accountability Act, though, is the fact that algorithms are considered to be privately owned by the consumer. Since owners of algorithms would not be required to share information on these tools, the bill suggests that consumers must conduct their own testing to determine whether automated decision systems are free of algorithmic bias. Soon after the bill was introduced, Andy Sanchez, Kevin McDermott, and Alicia Cintora of Cornell University published an article suggesting that an "independent nonprofit accreditation board, the Forum for Artificial Intelligence

[60] Clarke, Yvette D. "Text - H.R.2231 - 116th Congress (2019-2020): Algorithmic Accountability Act of 2019." Congress.gov, April 11, 2019. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

[61] Shneiderman, Ben. "The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight." *Proceedings of the National Academy of Sciences of the United States of America*113, no. 48 (2016): 13538-3540. Accessed March 15, 2021. https://www.jstor.org/stable/26472631.

Regularization (FAIR)[, must be established.]"[62] By requiring an outside expert to, with the algorithm's impact statement, assess the system and determine whether it is acceptable for commercial use, algorithms can be better evaluated due to the lack of influence from company insiders.

Outlining the goals and possible faults algorithms may possess is a step forward toward algorithmic fairness, though not much has been established as to *when it is necessary* that algorithms are developed/used. It is evident that humans should remain involved in many decision-making processes—the issue is how to control the amount of human influence needed for the process to most effectively and efficiently minimize bias. In order to regulate this issue, I have developed the Farls Factor, a concept stating that: **in a decision-making process, the amount of human impact should be proportional to the amount of human involvement**. Due to the rise in their complexity, algorithms must be evaluated to determine whether they are trying to accomplish too much or whether their use contradicts the objective of the process. The Farls Factor can be better understood through examples of fully algorithmic procedures, fully human decisions, and the area that lies between.

Even though algorithms can develop flaws in a multitude of ways, there are certain processes in which no human involvement is needed. Imagine a warehouse filled with pallets of goods; there's a loading dock where 50 pallets have recently been delivered, and they need to be taken and stored in the rightmost corner of the building. With the appropriate technology, an automated forklift can perform this task while no one is present at the warehouse; since humans are not directly impacted in the situation, any algorithmic bias

[62] Sanchez, Andy, Kevin McDermott, and Alicia Cintora. "A Policy Memorandum to the Partnership on AI: Accreditation and Educational Programs to Ensure Fairness in AI." *Journal of Science Policy & Governance* 15, no. 1 (October 2019).

(choosing to move one pallet before another, for example) is completely disregarded. The lift may drop a pallet, minimizing the company's overall profit, but errors like these are not due to algorithmic bias. Some may consider a bias to be the decision of whether the robot decides to do its job or not based on a number of factors; while they would be correct, the issue could be easily identified and solved through an accessible change of code. Situations such as this that have little to no ethical relevance can be left completely up to algorithms since any possible algorithmic bias would have no effect on the outcome.

Some procedures can be completely left to humans due to the urgency or intensity of the decision or, more specifically: binary decisions in which each choice is equally important to the other. The most important factor in these decisions, though, is the opportunity for humans to provide the decision with a perspective that an algorithm cannot comprehend (compassion, love, distaste, etc.). If an individual is found unconscious and it is determined that they are having a stroke, an algorithm should not be the one to decide whether the patient is given IV tPA[63] or not. While the decision could be made through a random choice, there is an element of faith involved; either the doctor or the patient's family/friend/rescuer will have a gut feeling on what to do. This does not mean a positive outcome is guaranteed, but the choice loses its nature of pure randomness. If the time since the onset of the stroke is unknown, each choice could either mean life or death—it is up to chance. Algorithms may not even be considered for these types of decisions, and it is understandable why they would be left strictly up to humans; algorithmic involvement would only take up needed time, leaving the process worse off in the end.

[63] IV tPA, or intravenous tissue plasminogen activator, is a protein that the body makes to break up clots. It can be injected into a stroke patient within four-and-a-half hours of the onset; if given any later than this period, the patient could die.

When an algorithm is created to choose between real-world entities reduced to quantifiable data points, though, neither human nor algorithm should dictate the entirety of the process. Due to the various biases that can arise within each method, a hybrid system would, theoretically, help to better expose implicit biases while activating necessary ones. By examining the examples of fully algorithmic and fully human procedures, the responsibilities become much more clear: humans set parameters for algorithms to sort data and later make their decisions based on what the algorithm has revealed. To better explain how this method could be considered when developing and using algorithms, I have created a theoretical college admissions algorithm in Java that sorts students by GPA and SAT Score.

```java

public class Student {

    private double gpa;
    private int sat;

    public static boolean gpaStatus(double studentGPA){

        gpa = studentGPA;

        if(gpa < 3.0)
            return false;

        else
            return true;

    }

    public static boolean satScore(int studentSAT) {

        sat = studentSAT;

        if(sat < 1250)
            return false;

        else
            return true;
    }

}
```

Figure 2: Two basic methods used to determine whether a student's GPA and SAT score are high enough for this school's requirement.

These two methods simply receive their respective inputs—the student's GPA and SAT score—and determine whether or not they fall below a certain, set threshold. Returning to the discussion on binary processes, these methods are prime examples; they may only act as extremely small parts of a complete college admissions program, though the rest of the algorithm would be made up of these simple, yes-or-no decisions. Students are not only considered based on their GPA and standardized testing score, though; a multitude of factors play into accepting a student to a university. By analyzing each element of a student's application, it is easier to understand the complexity behind these decisions. For this example, I will be covering GPA, Course Rigor, SAT/ACT Score, Race/Gender, Economics, Activities/Work Experience, Legacy Status, Essays, Letters of Recommendation, and Interviews.

One of the largest signifiers of academic merit is the grade point average (GPA). Based on the difficulty level of classes a student takes as well as their performance in those courses, a number is calculated to reflect the individual's caliber. Students submit their secondary school GPA to a college as part of their application—these figures cannot be immediately compared, though; different schools use different scales which may result in one student's 6.0 being another's 4.0. It is understood that universities are given information to determine the weighting of each high school's system, so an algorithm could be used to sort this data. By converting each applicant's GPA to a standard scale, the method in Figure 2 could be appropriately used, though this still does not set a baseline for academic excellence to be compared.

Many selective are known to understand a student's intellectual experience through the courses they take throughout secondary school. They are often evaluated based on how many Advanced Placement (AP) or Honors courses they completed throughout their four years as well as their performance in those respective classes. While two students from different high schools may have received the same grade in AP Chemistry, for example, the difficulty of each school's respective courses must be considered; one student may have had hours of daily homework while another's grade may have been completely determined off of exams. Universities typically have regional admissions counselors who supposedly understand the intensity each secondary school in that area, but it is extremely difficult for them to comprehend the differences between institutions without experiencing the first-hand. A school's difficulty being entirely subjective on the part of the regional director may result in various biases based on the presentation of the school as well as trends seen through past applicants' experiences.

The Scholastic Aptitude Test (SAT) and American College Testing (ACT) exam are taken by most prospective college students in order for universities to better understand and compare students' intelligence through a *standardized* test. This system may seem promising in theory, but many aspects surrounding these tests, themselves, leave some at disadvantages. Each test costs around fifty dollars, making it difficult for low-income students who do not receive exam waivers. In addition, while the questions vary from test to test, they all follow a very similar format; SAT/ACT tutors are extremely popular and, if a student can afford one, they will be able to clearly understand the exam format and types of questions well before test day. When it comes to the testing experience, different environments can serve different challenges based on the testing location.[64] Universities must also consider the difficulty of each specific test, as these exams are curved to represent a normal distribution across the

[64] When I took the exam in September of 2020, a car alarm was active outside of my testing room for almost the entirety of the exam. Since I was testing at a school in a city, distractions like these (construction, traffic, etc.) are much more likely to occur than at a suburban high school campus.

scores. Various factors play into an individual's SAT/ACT score, though algorithms can be used to easily sort students based on their results.

Aside from academics, some individual traits of students are out of their control—specifically their gender and race. While some believe that these factors should not influence decisions, it is important that they are considered, for they help form an individual's identity. These societal frameworks allow admissions officers to better understand the context surrounding applicants' lives—how these attributes may have influenced their experiences. As the world continues to shift into a more progressive landscape, the concept of affirmative action has been accepted by many universities.[65] The line is very blurred as to what is considered a "fair" bias towards minority groups, though it is still important that race and gender are emphasized in order to preserve an applicant's identity. It is simply unethical to sort students based on these qualities, for their specificity may lead to admissions officers to promote applicants for a sole, inherent trait; whether students are favored or not based on their race and/or gender is for the college to decide, but these factors must be recognized in order for university admissions to better understand their applicants.

Most universities claim to practice need-blind admissions, though economic status is still included on students' college applications. In an article on Fastweb.com, a website that shares information about colleges, financial aid, etc., The Fastweb Team reports that "very few colleges are completely need-blind, as financial need often affects the wait-listed,

[65] Affirmative action is the effort toward promoting racial diversity and, more specifically, favoring individuals of historically discriminated communities to help them succeed in higher education, the workforce, etc.

international and transfer students."[66] If a student does not apply for financial aid, the school may be tempted to accept the individual since they will be paying full tuition. This information also allows admissions officers to better understand a student's situation in general; if their list of activities is short and their parents' income is low, it may be an indicator that the student was not able to afford participating in clubs, sports, etc.. On the contrary, students from low-income backgrounds who perform exceptionally well academically may look more appealing in the university's eye.

Activities and work experience allow admissions officers to better understand what students are involved in outside of the classroom. There is a stigma surrounding the idea that universities favor applicants who are involved in their communities, often driving individuals to fill their applications with each and every club they have been involved in throughout high school. If an applicant lists *too many* activities, admissions officers may be inclined to think that the application is not a true representation of the student. A student who is interested in the same activities as the admissions officer reading their application may also seem more appealing due to implicit bias. Work experience is also important, showing an applicant's commitment to something other than school. Sometimes students may have to work in order to help provide for their families, so the addition of their job on their application allows them to better tell their story.

Legacy status plays a large role in the college admissions process, often giving students large advantages if a family member had attended the university at which they are

[66] Team, The Fastweb. "Does Applying for Financial Aid Affect Your Chances of Admission?," March 14, 2021. https://www.fastweb.com/financial-aid/articles/does-applying-for-financial-aid-affect-your-chances-of-admission.

applying. In their article “The Origins of Legacy Admissions: A Sociological Explanation,” Deborah L. Coe and James D. Davidson claim that “a number of students are accepted on the basis of admission policies that have more to do with applicants’ family lineage than their high achievements.”[67] Schools often state that legacy status only benefits students when the admissions office is making a decision between a legacy and an equally-qualified candidate, but the statistics show that this is not always the case: “Legacy admissions make up 10% to 15% of most Ivy League schools’ freshman classes each year.”[68] It is known that most universities strive to obtain a high yield rate, so offering admission to students who are connected to the school may result in a higher probability of the applicant accepting the offer.[69] While legacy status may give an indication of whether or not a student would attend a university, it would be more beneficial to hear from the students, themselves.

Most selective universities require that applicants write one or more supplemental essays as part of their applications. Prompts may vary from “Why [insert university here]?” to “What makes you unique?” as schools attempt to better understand students’ way of thinking and their desire to attend the university of choice. This is the least-reductionist *required* aspect of the application, though word limits often leave applicants struggling to pick and choose what information is important enough to include. It is typical for students to write about a traumatic experience they may have had in an attempt to emotionally capture the admissions officer. While this may give more insight on the applicant’s life and how their academics had been affected by external experiences, it is up to the admissions team to

---

[67] Coe, Deborah L., and James D. Davidson. "The Origins of Legacy Admissions: A Sociological Explanation." *Review of Religious Research* 52, no. 3 (2011): 233-47. Accessed March 15, 2021. http://www.jstor.org/stable/23055549.

[68] Deborah Coe and James Davidson, “The Origins of Legacy Admissions: A Sociological Explanation”, 233.

[69] A university’s yield rate is the ratio of students accepted to students who attend/accept their offer.

determine who they feel “belongs” at their university.[70] Thousands of students apply to universities every year and many are often rejected before their essays are even read. Whether it be due to their GPA, SAT/ACT score, or other statistic, colleges may deny students if they fall below the school’s threshold. For those whose essays are considered, though, an acceptance may come down to how an admissions officer is feeling on the day they read an application.

Rather than trusting students to give a personal account of themselves, letters of recommendation are often helpful as they allow admissions to understand an applicant from an outsider’s perspective. Universities typically require letters of recommendation from teachers so that student participation/achievement in the classroom can be better understood. As it is understood that course rigor may vary from school to school, teachers have the opportunity to better explain their own curriculum and the applicant’s respective success. Students are often given the option to submit additional letters of recommendation from employers, coaches, or other mentors. While these seem to be practical accounts of an applicant’s identity, students may request that all negative information is eliminated from their letters; what a recommender writes is up to their discretion, but they are more likely to give a positive account of a student. Admissions officers may not fully trust letters of recommendation based on previous experience, so this is not a foolproof method of student evaluation.

Interviews are the most effective measure of understanding a student, though their weight in the admissions process is often undervalued. Since they are not able to be given to

[70] I am writing “belongs” in quotes simply because of the selectivity of the phrase. There are typically more applicants that would be fit to attend a university than accepted but, due to the low acceptance rates of some schools, many are often denied admission.

every applicant (at most schools), it would be unfair to favor interviewed students over those who did not receive the opportunity. Because of this, "[t]he interview actually counts for around 5% of [the] total application."[71] In an ideal world every applicant would have the chance to be interviewed and, now that Zoom has opened the door for virtual meetings, it is no longer necessary that they must be conducted in-person. Another issue with the interview process is the fact that they are typically conducted by alumni of the university, not the admissions officers, themselves. Since admissions only receives a report on the student, the opportunity simply boils down to another letter of recommendation—it may be more credible, but the disjointed nature leaves the experience fairly insignificant.

Decisions are complex. While the college admissions process is extremely involved on the surface, a deeper analysis reveals that not everything can be taken at face value. A student's statistics are affected by multiple factors, many of which are out of their control. As discussed in Chapter 2, any introduction of algorithms into decision-making processes lead to a condensation of identity in some way. Algorithms can act as incredibly helpful tools in order to increase efficiency, but when they are assigned in areas that contradict the purpose of the process, there is more room for implicit negative biases. Consideration of the Farls Factor as part of the creation of an algorithm's impact statement will help both developers and users understand where biases may occur. A separation of variables is needed for a closer examination of how and why decisions are made.

The Farls Factor was not created to be implemented in every single step of a process; it is a concept that should be considered on a macro scale. It has already been determined

[71] Calcagnini, Lily. "How Much Do College Interviews Matter?" CollegeVine, March 12, 2020. https://blog.collegevine.com/how-much-do-interviews-matter/.

that, due to the nature of the situation, there should be a hybrid of both human and algorithmic involvement in the college admissions scenario. Breaking it down even further, GPA and SAT/ACT Score are the only two categories in which the use of algorithms is fully appropriate. While tons of factors must be considered when evaluating students' GPA and standardized testing scores, universities typically have a threshold that allows them to gauge who has the academic capability to succeed in the college's general curriculum.

Activities/Work Experience, Essays, Letters of Recommendation, and Interviews would all fall on the fully human end of the process, as there is no practical reason for algorithms to dictate any of these categories. These four categories are included on the application so that admissions officers can get a better understanding of the individual as a whole, not only their academic standing. While they may influence academic success, the implementation of algorithmic involvement when analyzing any of these categories would be inappropriate for the situation. It is possible to develop an algorithm, for example, that searches for key words in a student's essay; if the goal of the supplemental essay is to allow the admissions officer to understand who the student is by trying to connect to them through text, an algorithm would only lead to a further loss of identity.

Course Rigor, Race/Gender, Economics, and Legacy Status all lie in a grey area—they should be included on the applicant's application, though their status should not dictate the applicant's final decision. These four categories help to further shape an applicant's identity and may help to fill in holes that may appear in their application. While algorithms can be created to sort by any of these areas, they are not necessarily needed since they do not act as primary factors that influence applicant admission.[72]

---

[72] Categories would have to be created for course rigor, generalized by the overall difficulty/competitiveness of a secondary school.

By developing algorithms and their impact statements with the Farls Factor in mind, negative biases must be considered. Even though algorithmic impact statements only cover the negative biases that may appear in the algorithm, itself, development of the statement requires the contractee to consider both sides of the process—when algorithms are used, if they are being used appropriately, and what actions are left to the human. A true symbiotic relationship cannot be achieved if each party does not understand why and how the other is a part of the process. By holding algorithms to a non-objective status and making proper use of their assistance, decisions become much more transparent, and each factor can be considered in a holistic way.

CONCLUSION

The human-algorithm relationship will only continue to strengthen from here-on-out as computers become more accessible to people across the globe. Individuals are able to learn much more about the decisions they make with the help of algorithms, and their assistance can serve extremely helpful in coming to more holistic conclusions. While systems may seem to work at first glance, a deeper look at their structure and function may reveal issues that can unintentionally alter results. Algorithmic impact statements will allow for a reexamination of decision-making processes as a whole which may help to uncover various underlying biases. Through an extensive analysis of algorithms, their potential biases, relationship with humans, and why a minimization of negative biases is important, I was able to create a solution to fill the ever-present gap in the current framework surrounding algorithmic bias mitigation.

The question remains: where do we go from here? The Farls Factor can be proposed to companies or individuals required to create algorithmic impact statements. By considering this suggestion, developers can determine whether or not algorithms are fit to assist in certain areas of a process. While algorithms have the potential to complete a multitude of tasks, sometimes their implementation can damage processes if they begin to alter the purpose of a process. FAIR, the theoretical accreditation board Andy Sanchez and his colleagues wrote about, was proposed to be run by the Partnership on AI—a nonprofit built by heads of big-tech companies and civil rights groups that is dedicated to educating the masses on algorithms (specifically artificial intelligence). As algorithms are currently considered private property, companies must create their own impact statements, though the ability for outside accreditation boards, such as FAIR, to perform routine examinations of algorithms will allow for better transparency of the system as a whole.

Algorithmic bias is also a topic that has grown in size, though is not yet a crucial part of computer science education. It was not until recently that Partnership for AI began educating on algorithmic biases, showing that its presence in computer science discourse is relatively new. Algorithmic bias should be a crucial topic taught to computer science students at the university level in order for developers to understand the importance of creating reliable algorithms from the ground up. The Farls Factor could also be implemented into computer science education so that developers consider their algorithm's relationship with a system before they begin development. This topic can be further researched or reimagined at a technical level in order for developers to better understand how its implementation may affect machine learning algorithms. As the number of developers continues to grow, whether self-taught or educated at the university level, it is important that algorithmic bias and the Farls Factor are crucial parts of computer science discourse.

Even though these suggestions are primarily for developers and contractees, their implementation does not have to take place before we laypeople can use algorithms to learn about our current biases. When I get into my car to start my drive to school every morning, my iPhone typically suggests the music I should listen to. It is typically much too early for me to consciously think about why I was shown the songs that the algorithm predicted, though looking back I can better understand what I may choose to listen to at that time of day or when I begin a commute. The simplest of our biases can be better examined when reflected onto algorithms, and the purification of these systems can ultimately teach us more about ourselves. Humanity's relationship with algorithms has only begun; as it continues to grow, we must always remember who created these tools and what they are meant to accomplish.

BIBLIOGRAPHY

Brock, Kevin. "Rhetoric and the Algorithm." In Rhetorical Code Studies: Discovering Arguments in and around Code, 33-70. ANN ARBOR: University of Michigan Press, 2019. Accessed October 20, 2020. http://www.jstor.org/stable/j.ctvndv9pc.9.

Burling, Stacey. “CAPTCHA: The Story behind Those Squiggly Computer Letters.” Phys.org, Phys.org, June 15, 2012. https://phys.org/news/2012-06-captcha-story-squiggly-letters.html.

Calcagnini, Lily. “How Much Do College Interviews Matter?” CollegeVine, March 12, 2020. https://blog.collegevine.com/how-much-do-interviews-matter/.

Clarke, Yvette D. “Text - H.R.2231 - 116th Congress (2019-2020): Algorithmic Accountability Act of 2019.” Congress.gov, April 11, 2019. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

Coe, Deborah L., and James D. Davidson. "The Origins of Legacy Admissions: A Sociological Explanation." *Review of Religious Research* 52, no. 3 (2011): 233-47. Accessed March 15, 2021. http://www.jstor.org/stable/23055549.

Dastin, Jeffery. “Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women.” Reuters. Thomson Reuters, October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Greenwald, Anthony G., and Linda Hamilton Krieger. “Implicit Bias: Scientific Foundations.” *California Law Review* 94, no. 4 (2006): 945-67. Accessed January 31, 2021. Doi:10.2307/20439056.

Hillis, W. Daniel. *The Pattern on the Stone: The Simple Ideas That Make Computers Work*. New York: Basic Books, 2015.

Kleinberg, John, et al. “Algorithmic Fairness.” *American Economic Association*, 2018, www.aeaweb.org/articles?id=10.1257/pandp.20181018.

Lemay, A., and SP Leblanc. "Iterative Analysis of Competing Hypotheses to Overcome Cognitive Biases in Cyber Decision-Making." *Journal of Information Warfare* 17, no. 2 (2018): 42-53. Accessed December 18, 2020. doi:10.2307/26633153.

Lim, Hengtee. “7 Types of Data Bias in Machine Learning.” Lionbridge AI, August 5, 2020. https://lionbridge.ai/articles/7-types-of-data-bias-in-machine-learning/.

London, Alex John, director. *From Automation to Autonomy and the Ubiquity of Moral Decision-making*. *YouTube*, 31 Apr. 2018, www.youtube.com/watch?v=ZFrzCZQsoNY&feature=youtu.be.

Martin, K. Ethical Implications and Accountability of Algorithms. J Bus Ethics 160, 835–850 (2019). https://doi.org/10.1007/s10551-018-3921-3

Mittelstadt, Daniel Brent, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. “The Ethics of Algorithms: Mapping the Debate.” *Big Data & Society*, 2016.

Mullainathan, Sendhil. “Biased Algorithms Are Easier to Fix Than Biased People.” The New York Times. The New York Times, December 6, 2019. https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

Otero, Daniel E. The Birth of Algebra, 2000. http://cerebro.xu.edu/math/math147/02f/algebra/algebra.html.

Pearson, Steven D., et al. “Is Consensus Reproducible? A Study of an Algorithmic Guidelines Development Process.” Medical Care, vol. 33, no. 6, 1995, pp. 643â660. JSTOR, www.jstor.org/stable/3766517. Accessed 1 Sept. 2020.

Sanchez, Andy, Kevin McDermott, and Alicia Cintora. “A Policy Memorandum to the Partnership on AI: Accreditation and Educational Programs to Ensure Fairness in AI.” *Journal of Science Policy & Governance* 15, no. 1 (October 2019).

Shneiderman, Ben. The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight. Proceedings of the National Academy of Sciences of the United States of America, vol. 113, no. 48, 2016, pp. 13538â13540. JSTOR, www.jstor.org/stable/26472631. Accessed 1 Sept. 2020.

Steiner, Christopher. *Automate This: How Algorithms Took Over Our Markets, Our Jobs, and the World*. New York: Portfolio/Penguin, 2013.

Team, Insights. “Forbes Insights: Managing The Ethics Of Algorithms.” Forbes. Forbes Magazine, June 5, 2019. http://www.forbes.com/sites/insights-intelai/2019/03/27/managing-the-ethics-of-algorithms/.

University, Montana State. “Induction vs. Deduction.” Montana State University Billings. Montana State University, n.d.

Wilf, Herbert S. "What Is an Answer?" The American Mathematical Monthly 89, no. 5 (1982): 289-92. Accessed October 21, 2020. doi:10.2307/2321713.

Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy

Implications." Journal of Information Policy 8 (2018): 78-115. Accessed October 21, 2020. doi:10.5325/jinfopoli.8.2018.0078.

Williams, Ray. “How Neuroscience Can Help Us Make Better Decisions.” Ray Williams, December 13, 2018. https://raywilliams.ca/neuroscience-can-help-us-make-better-decisions/.

Xiang, Alice, and Inioluwa Deborah Raji. “On the Legal Compatibility of Fairness Definitions.” *Partnership on AI*, November 25, 2019, 1–6. https://arxiv.org/pdf/1912.00761.pdf.

Young, Nicholas T. “I Know Some Algorithms Are Biased-Because I Created One.” Scientific American Blog Network. Scientific American, January 31, 2020. https://blogs.scientificamerican.com/voices/i-know-some-algorithms-are-biased-because-i-created-one/.

ANNOTATED BIBLIOGRAPHY

Brock, Kevin. "Rhetoric and the Algorithm." In Rhetorical Code Studies: Discovering Arguments in and around Code, 33-70. ANN ARBOR: University of Michigan Press, 2019. Accessed October 20, 2020. http://www.jstor.org/stable/j.ctvndv9pc.9.

Kevin Brock's article proved to become extremely useful in understanding the makeup and origin of algorithmic thinking. As it gave a brief summary of the birth of the name "algorithm" as well as its evolution, my definition of the term was largely influenced by this work. Without the understanding of algorithms being processes made up of various steps, many arguments throughout this thesis would have failed to land, though the background and explanation surrounding the rhetoric of algorithms in everyday use helped me create a foundation from where I could speak.

Burling, Stacey. "CAPTCHA: The Story behind Those Squiggly Computer Letters." Phys.org, Phys.org, June 15, 2012. https://phys.org/news/2012-06-captcha-story-squiggly-letters.html.

I used this article to explain how CAPTCHA tests work and their significance in examining constraints around data. By using this test as a framework, the benefits and dangers surrounding algorithmic constraints could be better understood.

Clarke, Yvette D. "Text - H.R.2231 - 116th Congress (2019-2020): Algorithmic Accountability Act of 2019." Congress.gov, April 11, 2019. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

This bill, proposed by Yvette Clarke from District 7 of New York, provides a solution to monitoring and eliminating algorithmic biases in commercial systems. Using this bill as background, the Farls Factor's purpose was much easier to explain since this bill does not have any information on determining whether algorithms are fit for processes or not.

Greenwald, Anthony G., and Linda Hamilton Krieger. "Implicit Bias: Scientific Foundations." *California Law Review* 94, no. 4 (2006): 945-67. Accessed January 31, 2021. Doi:10.2307/20439056.

In Chapter 4, Greenwald and Krieger's article allowed me to effectively explain implicit bias and how its involvement in decision-making processes can go unnoticed. While it was easy to explain how accidental algorithmic bias can occur, unconscious human biases are often ignored and labeled as "unimportant." By explaining that there are areas of human thought that originated from outside sources, human bias in algorithmic processes could be better understood and explained. Greenwald and Krieger's article allowed me to define and explain human bias in order to argue why its recognition is important.

Hillis, W. Daniel. *The Pattern on the Stone: The Simple Ideas That Make Computers Work*. New York: Basic Books, 2015.

I use one of Hillis's examples to better explain how algorithmic procedures work in steps. Through the tic-tac-toe and switch analogy, the reader is better able to understand how these systems are set up. The step-based nature of algorithmic processing is important in understanding where errors may occur and how they can be resolved.

Lemay, A., and SP Leblanc. "Iterative Analysis of Competing Hypotheses to Overcome Cognitive Biases in Cyber Decision-Making." *Journal of Information Warfare* 17, no. 2 (2018): 42-53. Accessed December 18, 2020. doi:10.2307/26633153.

Lemay and Leblanc's study on human biases in cyber decision-making helped to outline the beginning of my fifth chapter, showing that human cognitive bias can override any algorithmic bias present. With their example of the different workers' approaches to cybersecurity breach regulation, it was much easier for me to show that not only are humans' interactions with algorithms different, but their individual approaches vary as well.

Lim, Hengtee. "7 Types of Data Bias in Machine Learning." Lionbridge AI, August 5, 2020. https://lionbridge.ai/articles/7-types-of-data-bias-in-machine-learning/.

Lim's article on the various types of data bias gave me the information (and definitions) of each type of data bias in the second chapter. As each slightly varied, I was able to use these definitions to (obviously) define each type and create my own examples for how they could be identified/present.

Mittelstadt, Daniel Brent, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society*, 2016.

Mittelstadt's article was one of the first that I approached in my research, allowing me to develop an understanding of the current discourse surrounding algorithm ethics. While he does not talk specifically about algorithmic biases, he discusses the ethical questions that they ask and how the human relationship with machines will continue to develop. Many of my initial ideas were sparked by this article due to its thorough examination of the worlds of humans, algorithms, and one where they coexist.

Otero, Daniel E. The Birth of Algebra, 2000. http://cerebro.xu.edu/math/math147/02f/algebra/algebra.html.

Otero's article helped to give a great summary of the birth of not only algebra, but the idea of the algorithm that proceeded it. Much of my first chapter is dedicated to the history of the term, from which I received much of the information from this source. Algebra was defined as a form of math that involved various steps to complete problems, and algorithms soon were known as algebra-based processes. Most of this information was retrieved from Otero's article.

Pearson, Steven D., et al. “Is Consensus Reproducible? A Study of an Algorithmic Guidelines Development Process.” Medical Care, vol. 33, no. 6, 1995, pp. 643â660. JSTOR, www.jstor.org/stable/3766517. Accessed 1 Sept. 2020.

When approaching a solution to regulating algorithms in an attempt to minimize bias, the current solutions in-place must be examined. Pearson’s article gives a detailed account of how algorithmic issues are approached and, more specifically, how frameworks are created to help prevent/eliminate problems before they occur.

Sanchez, Andy, Kevin McDermott, and Alicia Cintora. “A Policy Memorandum to the Partnership on AI: Accreditation and Educational Programs to Ensure Fairness in AI.” *Journal of Science Policy & Governance* 15, no. 1 (October 2019).

The proposal made by Cornell students Andy Sanchez, Kevin McDermott, and Alicia Cintora allowed me to better explain how an accreditation board can be created to develop algorithmic impact statements for companies and perform regular testing of their algorithms in order to identify and remove bias if found.

Shneiderman, Ben. The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight. Proceedings of the National Academy of Sciences of the United States of America, vol. 113, no. 48, 2016, pp. 13538â13540. JSTOR, www.jstor.org/stable/26472631. Accessed 1 Sept. 2020.

Shneiderman’s thorough analysis of how to approach the fixes to faulty algorithms helped to support my addition to algorithmic impact statements as it proposes a very similar concept. His article includes many helpful ideas for solutions that I used to help formulate my own concept that would help fill the gap in the current algorithmic accountability act.

Team, Insights. “Forbes Insights: Managing The Ethics Of Algorithms.” Forbes. Forbes Magazine, June 5, 2019. http://www.forbes.com/sites/insights-intelai/2019/03/27/managing-the-ethics-of-algorithms/.

Forbes Insights Team’s article surrounding general algorithmic bias helped to frame my entire second chapter, determining the three major types of algorithmic bias. From these categories I was able to explain what each was and how they could arise in various programs. By using this article, much of my later arguments could be traced back to be supported by the information provided in this article.

University, Montana State. “Induction vs. Deduction.” Montana State University Billings. Montana State University, n.d.

Montana State University’s document on induction and deduction provided me with the appropriate definitions needed to explain the two concepts. Using their simple and understandable definitions, I was better able to explain how these forms of reasoning are used by both humans and algorithms in decision-making processes.

Wilf, Herbert S. "What Is an Answer?" The American Mathematical Monthly 89, no. 5 (1982): 289-92. Accessed October 21, 2020. doi:10.2307/2321713.

This article is an iterative analysis of how we approach getting to answers, strictly through the lens of mathematics. Though I did not use any of the mathematical equations he proposes as examples, Wilf asks questions that must be examined when studying processes. For example, is the number of steps of a process proportional to the complexity of the system? While none of his questions are directly examined in the context of my thesis, his work acts as a foundation

Xiang, Alice, and Inioluwa Deborah Raji. “On the Legal Compatibility of Fairness Definitions.” *Partnership on AI*, November 25, 2019, 1–6. https://arxiv.org/pdf/1912.00761.pdf.

Alice Xiang and Inioluwa Deborah Raji’s paper helped me to draw the conclusion that algorithms should be held to the same standards that are already in place when it comes to fairness. I use this article to frame my third chapter before discussing algorithmic biases.

Young, Nicholas T. “I Know Some Algorithms Are Biased-Because I Created One.” Scientific American Blog Network. Scientific American, January 31, 2020. https://blogs.scientificamerican.com/voices/i-know-some-algorithms-are-biased-because-i-created-one/.

Young’s retelling of his first-hand experience with algorithmic biases helped to give an introduction to my third chapter, for he explains that he wasn’t exactly sure how bias arose in his program. As my third chapter is largely centered around determining bias, Young’s example shows that bias does not have to be associated with certain data points, but also with entire factors of the process, itself. Through the use of Young’s article, I was better able to explain how to determine whether or not an algorithm is biased.