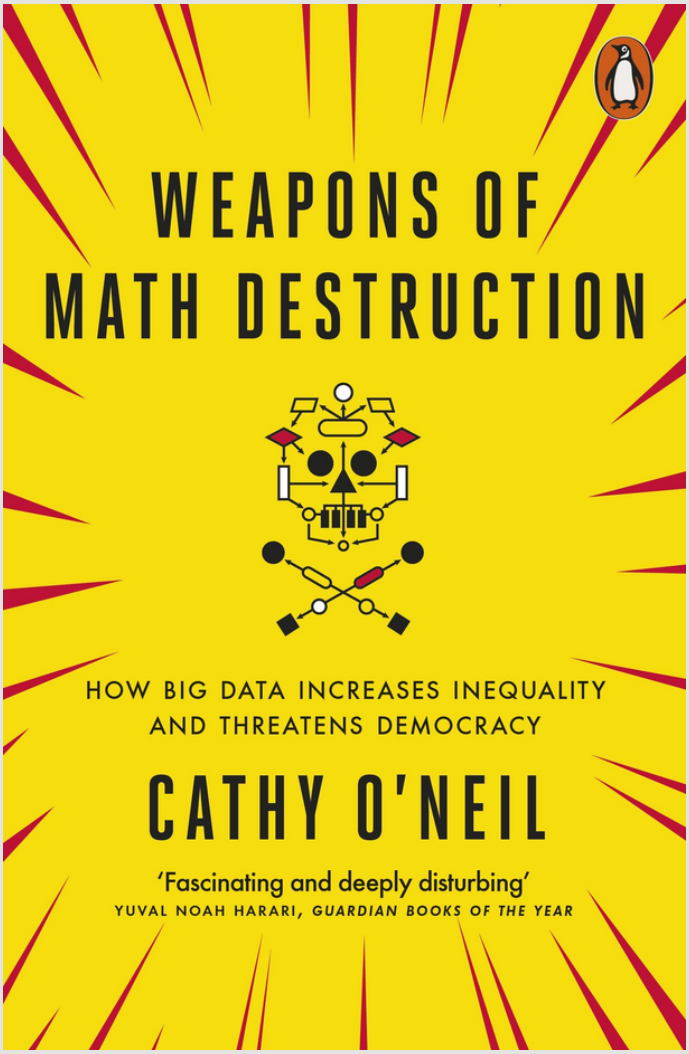


STUDY GUIDE

PREPARED BY

DATA FEMINISM
NETWORK

www.datafeminismnetwork.org
@DataFemNetwork

TABLE OF CONTENTS

INTRODUCTION:
THE POWER OF ALGORITHMS

CATHY O'NEIL

@mathbabedotorg

Ever since Cathy O'Neil was a little girl, she had a passion for math. She went on to earn a Ph.D. in mathematics and became a tenure-track professor at Barnard University. She made a career switch to the corporate sector as a quant for a leading hedge fund. In 2011, O'Neil quit her job at the hedge fund and rebranded herself as a data scientist. She joined an e-commerce startup and launched a blog, Mathbabe to mobilize fellow mathematicians and speak out against "sloppy statistics and biased models that created their own toxic feedback loops" (11). Cathy is currently the CEO of ORCAA, an algorithmic auditing company, and is a member of the Public Interest Tech Lab at the Harvard Kennedy School. Her new book The Shame Machine: who profits in the new age of humiliation came out in March 2022."
Source: https://mathbabe.org/contact/

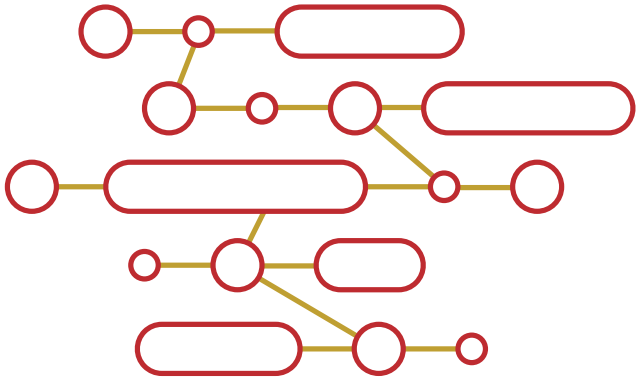
NOTE: This Study Guide is only a quick glimpse into the amazing content of this book. We encourage you to support Cathy O'Neil and purchase Weapons of Math Destruction, and use this guide as supplementary material.

INTRODUCTION:
THE POWER OF ALGORITHMS

Cathy's career in the finance industry was going well until the 2008 economic crash which "made it all too clear that **mathematics** [...] **was not only deeply entangled in the world's problems but also fueling them**" (2). 2008's unemployment rates, housing crisis, and downfall of major financial institutions were all "**aided and abetted by mathematical formulas**" (2). Mathematics was being used to make life-altering decisions and predictions about the future. Marketed as a fair and objective tool, mathematical formulas were thought to save time and reduce human bias. Many of these models, however, "**encoded human prejudice, misunderstanding, and bias into the software systems that increasingly managed our lives**" (3). O'Neil coined the term Weapons of Math Destruction (WMDs) to describe these harmful kinds of models. WMDs are "opaque, unquestioned, and unaccountable and operate at a scale to sort, target, or "optimize" millions of people" (12). We will go through several examples of WMDs in the following sections.

O'Neil's book focuses on the damages inflicted by WMDs and the inequalities they amplify. From college admissions to bank loans to prison sentencing, "all of these life domains are increasingly controlled by secret models wielding arbitrary punishments" (13).

"Welcome to the dark side of Big Data." (13)



CHAPTER 1:
WHAT IS A MODEL?

KEY INDICATORS:

"IS IT A WEAPON OF MATH DESTRUCTION?"

Opacity: Is the model opaque or even invisible? (28)

Damage: Is the model Unfair? Does it damage or destroy lives? (29)

Scale: Does the model have the capacity to grow exponentially? (29)

Chapter 1 explores three different examples of statistical models. The first model uses data from past baseball games to predict the outcomes of future baseball games. Baseball models, for the most part, are considered fair because they are transparent. It is no secret what data are fed into the model because they are available to everyone. Baseball models also use data that are relevant to the outcomes they aim to predict. This seems like a given, but as we will see later on, WMDs often use proxies, such as a person's zip code, to make a prediction about someone's likelihood to recommit a crime or pay back a loan. In a country like the U.S. where many neighborhoods were historically segregated, **zip codes can be a proxy for race**. The most important thing to note about the baseball model example, is that there are data constantly feeding into it. This provides the ability to update and edit the model to be more accurate. After all, "**conditions change, and so must the model**" (18). In comparison to the other two models we will describe, the baseball model is the most effective and ethical.

The second example is O'Neil's hypothetical family dinner model that inputs each member of her family's dietary preferences and aversions, previous dinner successes and failures, and all available food options. In theory, this model would maximize success and save O'Neil time in deciding what to cook for her family. This model, however, would not leave O'Neil's household and could never be adopted at a large scale.

One model that has become very popular is the Level of Service Inventory (LSI- R). This model is used to predict recidivism rates. LSI-R includes a series of personal questions for each person in prison to answer, such as "how many prior convictions have you had?" or "what part did others play in the offense?" (25). It's crucial to consider how a person from a privileged background might answer a question about their previous run-ins with police in comparison to a person from a poor inner-city background. A white man from a wealthy suburb has maybe never had an encounter with the police whereas "**young black males, by contrast, are likely to have been stopped by police dozens of times**, even when they're doing nothing wrong" (25). A 2013 New York Civil Liberties Union study found that "while black and Latino males between the ages of fourteen and twenty-four made up only 4.7% of the city's population, they accounted for 40.6% of the stop-and-frisk checks by the police", 90% of which were innocent (25). **According to the LSI-R model, these early encounters with the police, though most of the time innocent, make poor people and racial minorities appear far 'riskier'** (26). This model is an example of a pernicious feedback loop, a toxic cycle that sustains itself (27). Consider a person who is from a poor neighborhood that scores "high-risk," this leads to a longer prison sentence and therefore more time surrounded by fellow criminals. When they are released from prison they return to the poor neighborhood but this time with a criminal record, making it harder to find a job. If this person commits another crime, the model is considered to be 'successful'. Yet, (to put it simply), the model is confirming what it believes to be true: that poor people are more likely to recommit a crime. "Sentencing models that profile a person by [their] circumstances help to create the environment that justifies their assumptions. This destructive loop goes round and round, and in the process, the model becomes more and more unfair" (29).

To sum it up, the **three elements of a WMD are opacity, scale, and damage.** Some of these algorithms, you may notice, are not harmful to everyone.

They may help someone get a good job, go to college, or reduce their jail sentence. For others, however, "models powered by algorithms slam doors in the face of millions of people [...] and offer no appeal. They're unfair" (31).

CHAPTER 2
SHELL SHOCKED:
MY JOURNEY OF DISILLUSIONMENT

In this chapter, O'Neil discusses her experience working in the finance industry and her account of the WMD that led to the U.S. housing crisis of 2008. Mortgages, typically issued to people who would be able to pay them back and bring steady revenue, were being issued at insane rates. This is because credit rating agencies were studying and scoring the risk levels of the mortgages, giving a false sense of security. The strategy was to draft unsustainable mortgages, exacerbate the fees, and then release the resulting securities into the booming market (40). **Mortgage companies sought out people in poor and minority neighborhoods and offered them housing they couldn't afford**. In one federal lawsuit, **Wells Fargo was targeting black neighborhoods with subprime loans** carrying the highest interest rates (40). Even borrowers with rock-solid credit who would otherwise have qualified for fair loans were victims of Wells Fargo's scam. By the time the lawsuit was filed in 2009, over **half of the properties with loans from Wells Fargo were empty and 71% were in predominantly African American neighborhoods** (40). The mortgages themselves were not WMDs, rather they were mere financial instruments. The true destructive force was the model that rated risk.

Similar to the world of finance, O'Neil noticed a growing crisis in Big Data. She saw a "growing dystopia with inequality rising" (48). In the age of Big Data, the lucky few continue to thrive and convince themselves they deserve it, **while those at the bottom are programmed to stay there** (48).

CHAPTER 3
ARMS RACE: GOING TO COLLEGE

Imagine what would happen if all of a sudden veganism became the national standard. Our entire food economy would turn upside down. There would be massive food shortages as well as the waste of all of the products we no longer need. **Diets on their own do not have this impact, but when adopted on a national scale would create a "distorted and dystopian economy"** (51). It's the same for WMDs. A formula may be innocuous in theory, but if it grows to be a national standard, it can be detrimental (51). This is exactly what happened with the U.S. higher education system.

In 1983, *U.S. News* took on the ambitious project of ranking and evaluating 1800 universities. In a society that rewards alumni from top universities with high social status and fantastic job opportunities, this ranking was taken very seriously. To create this list, "all the magazine had to do was look at data" (51). The issue was deciding which data were important, or in other words, which criteria make a university 'excellent'. A few examples of the criteria were SAT scores, student-teacher ratios, acceptance rates, and campus facilities. One crucial component for many that were omitted from the evaluation was tuition cost. By leaving this out, "it was as if U.S. News had handed college presidents a gilded checkbook" (60). **They had 15 areas to boost their performance on the ranking, and maintaining low tuition fees was not one of them**. Universities hiked up their tuition to improve their facilities, services, and other important components according to U.S. New's formula. As a result, **tuition rose as much as 500% between 1985 and 2013** (60). **Wealthier people could take advantage of this system** and get ahead while everyone else was at an extreme disadvantage. Some resources that privilege wealthy students include SAT tutors, the money and time to be able to participate in extracurricular activities, and college admission tutors that help with applications and essays. There are also certain pieces of key information that only the students with insider connections would have access to. Many universities in the U.S. not only require you to pay insane tuition amounts (up to $70,000 a year) and/or to be deep in student debt for life but privilege people with the resources listed above that result in more 'impressive' applications.

This leads to an education system that amplifies wealth gaps, as the rich and upper-middle-class are able to attend top schools and earn good jobs whereas the lower-middle and lower class do not have the resources to go to these universities and, in turn, cannot get as high paying jobs. Of course, there are exceptions to this, but people with fewer resources attending university and getting a good job should not have to be an exception. Though today there are many community colleges around the U.S. that offer tuition for a fraction of the price, **top universities are more competitive and expensive than ever**. University rankings are not only a national standard but a "vicious loop materialized" (53).

CHAPTER 4
PROPAGANDA MACHINE: ONLINE ADVERTISING

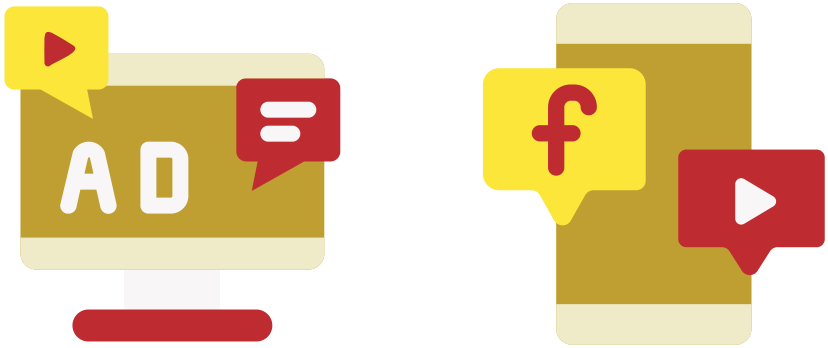
In 2014, Corinthia College, a for-profit university, declared bankruptcy after their access to federal student loan funding was put on hold (71). One year prior, Corinthia College was "busted by the attorney general of California for lying about job placement rates, overcharging students, and using unofficial military seals in predatory ads to reel in vulnerable people" (71). Their targeted students are described, in their words, as "'isolated', 'impatient' individuals with 'low self esteem' who have 'few people in their lives who care about them' and who are 'stuck' and 'unable to see and plan well for future'" (71). Despite the vulnerable student population, their tuition fees appear to be far from affordable. In one of Corinthia College's divisions, a bachelor's degree in paralegal is $68,800 USD, while other traditional colleges charge less than $10,000 USD for equivalent courses (71).

These universities know that "**vulnerability is worth gold**" (72), so their recruiters **lure in potential students with "low self-esteem**". How do these recruiters know so much about one's vulnerability? Prospective students are identified through targeted advertising, in-person visits, and even through their internet search history (73, 74). **Targets of for-profit universities are often in the dark** about how their information has been used by these "so-called diploma mills" (69).

The big question is, do students get a good degree that prepares them for the workplace, in return for the price tag? Investigators at CALDER/American Institutes for Research say no (80). In submitting nearly nine thousand fake resumes with various types of education, "[they] found that **diplomas from for-profit colleges were worth less in the workplace** than those from community colleges and about the same as a high school diploma" (80).

According to a ProPublica report, "**between 20 and 30 percent of the promotional budgets at for-profit colleges go to lead generation**" (77, 78). This cost is usually higher than the instruction cost in the for-profit education industry (79). For-profit universities constantly test and refine their marketing programs in order to "recruit the greatest number of students who can land government loans to pay most of their tuition and fees" (73, 74). The majority of the budget is spent on advertising on online platforms like Google and Facebook, which receive and relay data on viewers' interactions with the ads (75). This massive pool of data accelerates the fine-tuning of these campaigns within hours (75).

Predatory advertising is a form of WMDs (70). In the case of for-profit universities, predatory ads "**zero in on the most desperate among us at enormous scale**" (70). What seems to be a promise for upward social mobility, good job prospects, and better life outcomes **plunges vulnerable populations deeper into debt** (81).



CHAPTER 5
CIVILIAN CASUALTIES: JUSTICE IN THE AGE OF BIG DATA

Following the market crash of 2008, the small city of Reading, Pennsylvania had to cut forty-five police officers despite growing crime levels (84). Police chief William Heim needed to find a way to maintain or improve policing efforts with a smaller force. In other words, to do less with more. In 2013, Heim invested in PredPol, a predictive crime model that helps target and track criminal acts. The model does not target individuals, but rather geographic locations. PredPol then directs police to 'high-risk zones' which are more often than not, impoverished neighborhoods (85). It is crucial to acknowledge that **impoverished neighborhoods in America are disproportionately populated by Black and Hispanic people**. In America's widely segregated cities, "**geography is a highly effective proxy for race**" (87).

A major flaw of this model is that it includes nuisance crimes such as vagrancy (homelessness), aggressive panhandling, and the selling and consuming of small amounts of drugs (86). Nuisance crimes, according to O'Neil's research, are "endemic to many impoverished neighborhoods" (86). Including nuisance crimes in the model **skews the analysis and leads to disproportionate policing in impoverished neighborhoods**. Higher surveillance means more and more nuisance crimes will be reported, making it seem as though these neighborhoods are more crime-ridden and therefore dangerous. If the model was fair, police would be reporting jaywalking incidents in wealthy neighborhoods or underage drinking at frat houses. Yet, only in impoverished neighborhoods are people criminalized for these second-tier crimes and forced to face the consequences of prison time and a criminal record.

This example of the PredPol model illustrates that **poverty and race are criminalized while the model is considered fair** and to be keeping communities safe. People who are impoverished and not white, "are punished more for being who they are and living where they live" (97). The long-term impacts of models like this are detrimental: the more they amplify crime rates in impoverished neighborhoods, the more impoverished people of color are criminalized, and the harder it is for these people to find jobs which makes it nearly impossible to escape poverty. This is not a justice system, it is a system that reflects history's biases towards people of color and poverty.

CHAPTER 6
INELIGIBLE TO SERVE: GETTING A JOB

Kyle Behn studied at Vanderbilt University, a prestigious school in the United States. He took some time off to get treatment for his bipolar disorder and after a year and a half resumed his studies at a different school. Kyle was looking for a part-time job and applied to a grocery store that his friend worked at. Thinking his application was a sure thing, Kyle was rejected. Thanks to his friend's insider knowledge, Kyle found out that he was rejected because of his results from a personality test. Kyle did not match the 'ideal' personality of an employee. He applied to several more companies and was faced with the same result: rejection. Finish Line, Home Depot, Kroger, Lowe's, PetSmart, Walgreen Co., and Yum Brands **were all using similar personality tests that unlawfully screened out people with mental disabilities**. People who are rejected from jobs rarely know the reason nor have the resources to contact a lawyer. Opacity, as we discussed in 1, is one of the key indicators of a WMD.

Today, **machines are increasingly deciding who gets hired, and more importantly, who doesn't**. In an effort to save money and ironically enough, reduce human bias, algorithms are employed to screen applicants according to specific criteria. These algorithms can't predict how a person would actually perform in the role, so they use proxies. This is not only unfair but also inaccurate. Research suggests that personality tests are in fact poor indicators of future job performance (108). What if Kyle landed a job at Red Robin and performed amazingly? It's unlikely that the company that runs the personality test is going to tweak its algorithm. They have no incentive to update their algorithm. **The company is saving money and that's what matters to them**, more than losing out on a potential star (111).

Racial and ethnic minorities as well as women have historically faced discrimination when applying for jobs. Even if their applications aren't automatically screened out by an algorithm, it is a challenge to make it to the top of an applicant pool. In 2001 and 2002, researchers from the University of Chicago and MIT submitted five thousand fake resumes for various corporate positions.

Half of the resumes had stereotypically white names like Emily Walsh and Brendan Baker while the other half had similar qualifications but more African American sounding names like Lakisha Washington and Jamaal Jones (113). White applicants were invited to 50% more interviews than black applicants (113). Research also showed that white applicants with high-quality resumes had more callbacks than white applicants with low-quality resumes, but for black applicants, the quality of the resume didn't make much of a difference. This means that decision-makers were paying much more critical attention to white applicants than black applicants.

The key to these job sorting algorithms is to **know what the algorithms are looking for**. Similar to college admissions, those with privilege and insider knowledge will have the tools to succeed, and everyone else, will not.

CHAPTER 7:
SWEATING BULLETS: ON THE JOB

Clopening

Verb

1. When an employee works late one night to close the store or restaurant and then returns a few hours later, before dawn, to open it (123)

Last minute and irregular schedules are becoming more and more common for low wage workers at major companies like Starbucks and Mcdonalds. Sometimes employees only learn two days in advance that they'll be "clopening," a.k.a. getting no sleep. Not only does this wreak havoc on their lives but it also disproportionately impacts women who do the majority of infant and elderly care work. How are people supposed to determine their childcare plans when they don't even know when their next shift is? **These irregular schedules are a result of the data economy (124).**

Efficiency-focused WMDs **treat workers like cogs in a machine, prioritizing profits above human well-being**. Software scheduling programs process new streams of ever-changing data like weather or special events to determine staffing needs for each individual day (125). Conditions change hour by hour, so the workforce must accommodate the fluctuating demands. With this system, no hour is left overstaffed or understaffed and every second is busy.

In 2014, the New York Times published an article about a single mother named Janette Navarro who was working her way through college as a barista at Starbucks. Janette's uncertain and ever-changing schedule strained her ability to care for her four-year-old, causing her to put her studies on pause. This story is common. According to U.S. government data, **two-thirds of food service workers and over half of retail workers learn their schedules with notice of a week or less** (126). Scheduling software creates a pernicious feedback loop. In Janette's case, she had to drop out of school which decreased her employment prospects and financial opportunities and therefore "kept her in the oversupplied pool of low-wage workers" (129). Since the software is designed to maximize labor efficiency (and save money), employees often work less than 30 hours/per week which makes them ineligible for company health insurance. With an uncertain schedule, people also find it difficult to get a second job. It seems as though "the software was designed to expressly punish low-wage workers and to keep them down" (129).

In 1983, the Regan administration warned that a "rising tide of mediocrity in the schools threatened [America's] very future as a Nation and a people" (134). The main contributor to this 'failure' was a drop in SAT scores from 1963 to 1980. Teachers, according to the Regan administration, were to blame for this national catastrophe. The call to action was performance testing on students that was thought to measure the effectiveness of teachers. It is called the value-added model and it is well-entrenched within forty states (140). Tim Clifford, a middle school English teacher in New York City with twenty-six years of teaching experience, bombed a teacher evaluation with an abysmal score of 6/100 (135). Tim was given a failing grade and absolutely no insights on how to improve it. Continuing with business as usual, Tim scored 96/100 the following year. According to Tim, he "knew that [his] low score was bogus, so [he] could hardly rejoice at getting a high score using the same flawed formula.

The 90 percent difference in scores only made [him] realize how ridiculous the entire value-added model is when it comes to education" (136). Unsurprisingly, Tim's scores were "based almost entirely on approximations that were so weak they were essentially random" (137). Once we begin to understand WMDs and their statistical flaws, we can advocate for alternatives that are fair and just.

CHAPTER 8
COLLATERAL DAMAGE: LANDING CREDIT

It was a thing of the past when local bankers had the mighty power to decide who would get a loan and who would not (141). Consciously or subconsciously, these bankers, knowing your church-going habits and family stories, would correlate your religion and network with your trustworthiness for a loan (141, 142). Under this system, "**outsiders, including minorities and women, were routinely locked out**" (142). Then came FICO – an algorithm developed by Earl Isaac and Bill Fair "to evaluate the risk that an individual would default on a loan" (142). FICO "was fed by a formula that looked only at a borrower's finances—mostly her debt load and bill-paying record" (142). FICO is considered respectable and a non-WMD because:

1) FICO knows when its model makes mistakes and can re-evaluate it accordingly
2) FICO provides transparent credit scores
3) The credit-scoring industry is regulated for privacy and misinformation (142)

The regulated credit-scoring industry prevented companies from using credit scores for marketing purposes. To target potential customers, they needed an alternative (144). Then came along e-scores, a FICO stand-in, and a WMD. Instead of looking at a person's financial data, which is regulated information, e-scores used data, from "our zip codes and internet surfing patterns to our recent purchases," as **proxies for our creditworthiness** (143).

E-scores are like a backward innovation from FICO. A banker deciding whether a person from a poor area of the village deserves a loan is similar to e-scores considering zip codes in their credit rating. Looking at zip codes, "they are expressing the opinion that the history of human behavior in that patch of real estate should determine [...] what kind of loan a person who lives there should get" (146). In other words, the modelers for e-scores have to make do with trying to answer the question "how have people like you behaved in the past?" when ideally they should ask, "how have you behaved in the past?" (146).

Although **credit reports** are regulated under laws, they are **often used as proxies** for things like hiring or even on some dating apps**.** While it only takes one injury for a laid-off employee in a recession with no health insurance to miss a bill, some **companies have decided that credit reports are an accurate signal of trustworthiness and reliability** (147, 149). According to a Society for Human Resource Management survey, "**nearly half of America's employers screen potential hires by looking at their credit reports**" – some employees even have to go through the same process for promotion (148). Candidates are allowed to refuse the disclosure of their credit report, but in doing so they risk being disqualified (148). This becomes a feedback loop: "[I]f you can't get a job because of your credit record, that record will likely get worse, making it even harder to land work" (148). In an attempt to innovate and save resources for business, this hiring practice creates a poverty cycle for "the unlucky people caught up in it" (148).

We need humans...

Catherine Taylor learned that machines had been confusing her with Chantel Taylor and two other Catherine Taylors, all of whom had criminal charges (152, 153). This machine error has cost her multiple job offers and her application for federal housing assistance (153). Lucky for Catherine Taylor, "one conscientious human being had cleared up the confusion generated by web-crawling data-gathering programs" in one meeting (153).

Unfortunately, having a human being correcting the system doesn't happen all too often because (a) systems are built to run on their own, so "humans in the data economy are outliers and throwbacks" (153), and (b) WMDs don't receive feedback on their errors, so "injustice persists" (155).

"There's a paradox here. If we return one last time to that '50s-era banker, we see that his mind was occupied with human distortions—desires, prejudice, distrust of outsiders. To carry out the job more fairly and efficiently, he and the rest of his industry handed the work over to an algorithm.

Sixty years later, the world is dominated by automatic systems chomping away on our error-ridden dossiers. They urgently require the context, common sense, and fairness that only humans can provide. However, if we leave this issue to the marketplace, which prizes efficiency, growth, and cash flow (while tolerating a certain degree of errors), meddling humans will be instructed to stand clear of the machinery" (155).

While problems with the credit rating models in the lending industry are emerging, **newcomers are already pouring into the industry**. Facebook (now Meta) has patented **a credit rating system based on its social network**. A college graduate with no credit score, who is Facebook-friends with PhDs and other successful graduates, is considered more creditworthy than a hardworking cleaner, who is Facebook-friends with unemployed people or people with criminal charges (155, 156).

ZestFinance, a start-up founded in 2009, strives to **offer loans at a lower interest rate than payday loaners** (158). To measure an applicant's credit rating, they process **thousands of data points** from "whether applicants have kept up with their cell phone bills'' (158) to "whether applicants use proper spelling and capitalization on their application form, how long it takes them to read it, and whether they bother to look at the terms and condition" (158). Because of their limited language skills, **immigrants and/or poor people might have higher fees**, making it harder for them to repay their loans and then causing them lower their credit scores (158).

."**When new ventures are built on WMDs, troubles are bound to follow, even when the players have the best intentions**" (158). We, as borrowers, are "batched and bucketed according to secret formulas" (159). These algorithms believe that if you live in a high-crime area, if enough people around you are unemployed or criminals, or if you signal any signs that you are struggling or poorly educated, you and "people like you" cannot be trusted with a loan or a job, they decide (145).

CHAPTER 9
NO SAFE ZONE: GETTING INSURANCE

In the late nineteenth century, renowned statistician Frederick Hoffman created a potent WMD. He published a 330 page report "that set back the cause of racial equality in the United states and reinforced the status of millions as second-class citizens" (161). Confusing causation with correlation, he used statistics to reinforce his thesis that race was a powerful predictor of life expectancy (161). Hoffman never considered whether factors like poverty and injustice may influence the death rate of African Americans. He never thought about the lack of decent schools, modern plumbing, safe workplaces or inaccessible health care that could be killing African Americans at a younger age (162). These ideas permeated into housing and insurance industries and were used to justify the redlining of black people. Half a century later, redlining is "coded into the latest generation of WMDs" that similarly confuse correlation with causation (162). They punish the poor as well as racial and ethnic minorities.

Beginning in the seventeenth century, mathematicians were able to use math to predict the prevalence of accidents within large groups of people. Today, with more advanced technology and the ever-growing access to personal information, insurers are capitalizing on their ability to calculate risk for individuals as opposed to large groups.

The problem, according to O'Neil, is that "the models place us into groups we cannot see, whose behavior appears to resemble ours. This opacity can lead to gouging" (164). In short, we are charged varying insurance rates based on our behaviour, even when behaviour is not a direct indicator of our risk.

Risk scores based on proxies set by the insurer can take advantage of the poor and less educated. Take auto insurance, for example, which puts a larger weight on one's credit score than driving record when determining their risk score. This becomes a feedback loop, where "poor drivers who can least afford outrageous premiums are squeezed for every penny they have" (167).

At the height of World War II when American armies needed to round up as many people as possible, the Internal Revenue Service (IRS), tweaked the tax code to offer employer-based health insurance (173). At this time, only 9% of American workers had health care as an employee benefit. 10 years later, this number grew to 65%. Within one decade, companies exerted immense control not only over their employees' finances but also over their bodies (174). To save money from "high-risk" employees with more expensive health insurance costs, companies encouraged or even required wellness programs. The fewer emergency room visits, "the less risky the entire pool of workers looks to the insurance company, which in turn brings premiums down" (174). In 2013, CVS announced that it would require employees to either report levels of body fat, blood sugar, blood pressure, and cholesterol or pay $600 a year (176). By shaming their employees into being 'healthier' or making them pay the cost of their 'unhealthiness,' CVS was saving money. As a journalist from Bitch Media put it, "have someone tell you you're overweight or pay a major fine" (176). The practice outwardly promotes fatphobia and false notions of health and humiliates employees. Wellness programs are not yet considered a WMD, as they are not based on mathematical algorithms. However, they are considered "a simple and widespread case of wage theft [...] wrapped in a flowery health rhetoric" (178). If, (or when), surveillance extends into employee health data and becomes a health productivity model, we could have a horrific WMD.

CHAPTER 10
THE TARGETED CITIZEN: CIVIC LIFE

Let's imagine that O'Neil launches a petition on her Facebook page for tougher regulations on WMDs. Which of her friends will actually see the petition on their newsfeeds? We don't actually know. The second O'Neil posts the petition, it belongs to Facebook and their algorithm that decides "how to best use it" (179). For many of O'Neil's friends, the petition will be buried so low in their feed that they'll never see it (180). At the end of the day, **Facebook is a publicly traded corporation**. They determine what we see and learn according to their own interests (180). By programming their algorithm and choosing the news we see, "**can Facebook game the political system?**" (180).

During the 2010 and 2012 elections, Facebook conducted an experiment called the "voter microphone" that put a sticker on someone's profile when they voted. Facebook was encouraging people to vote through peer pressure. If they saw that their close friends had voted, they might be more likely to vote themselves. Researchers found that the campaign had increased voting turnouts by 340,000 people, enough votes to swing an entire election (181). Facebook's power "comes not only from its reach but also from its ability to use its own customers to influence their friends" (181). There's no evidence to believe that the social scientists at Facebook are actively trying to sway the political system, but **what we can prove is Facebook's power to influence what we learn and how we act** (184). Facebook's algorithm is not yet a full-blown WMD, but it is definitely opaque and massive.

Microtargeting developed into another significant political effort. It proved especially effective to sway swing voters with the information they cared about. It was also used to raise money from donors by sending different types of messaging. Microtargeting tactics also extend beyond campaigns and **can infect our civic life** (193). In 2015, an anti-abortion group called the Center for Medical Progress posted videos of what they claimed to be an aborted fetus from a Planned Parenthood clinic. The video falsely asserted that Planned Parenthood was selling aborted baby parts for research, and led to a Republican push to eliminate their funding (193).

Even though the video contained misinformation, the Center for Medical Progress still used microtargeting to build a case against Planned Parenthood. Microtargeting is "vast, opaque, and unaccountable" (196). **It allows propaganda and politicians to be different things to different people**.

What if instead of using algorithms to inflict damage, we used them for humanity's benefit? in a political race, for example, "a microtargeting campaign might tag certain voters for angry messages about affordable rents. But if the political candidate knows these voters are angry about rent, h**ow about using the same technology to identify the ones who will benefit from affordable housing and then help them find it?**" (197). If we were to change the objective from profit maximization to helping people, a "WMD is disarmed" and **could even become "a force for good**" (197).

CONCLUSION

Throughout this book, "we've visited school and college, the courts and the workplace, even the voting booth". We've seen a glimpse of the destruction caused by WMDs that promise "efficiency and fairness" but "distort higher education, drive up debt, spur mass incarceration, pummel the poor at nearly every juncture, and undermine democracy" (199).

"We must come together to police these WMDs, to tame and disarm them. [Oneil's] hope is that they'll be remembered, like the deadly coal mines of a century ago, as relics of the early days of this new revolution, before we learned how to bring fairness and accountability to the age of data. Math deserves much better than WMDs, and democracy does too" (218).

DISCUSSION QUESTIONS

1. Which section stood out to you the most, and why?

2. What makes algorithms so powerful and capable of destruction?

3. How do algorithmic feedback loops deepen societal inequalities?

4. Are there any contexts where you think algorithms can safely replace/supplement human decision-making?

5. How can we leverage the power of algorithms for good?