Societal Implications From Racial Bias In Machine Learning and Artificial Intelligence

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfilment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Everett Patterson Jr
Spring, 2020

On my honor as a University Student, I have neither given nor received
Unauthorized aid on this assignment as defined by the Honor Guidelines
For Thesis-Related Assignments

Signature ____ Date __________
Everett Patterson Jr

Approved ____ Date __________
Sean Ferguson, Department of Engineering and Society

Introduction

Artificial intelligence brings disproportionate risk to certain communities, a result of bias in data and those who program algorithms into these machines and software tools. Improper regulation or bias mitigation efforts, before integrating these technologies within society, lead to both direct and unforeseen harm done to already marginalized groups. "We are now at a tipping point where if this bias is left unrecognized and unchecked, it will result in serious negative consequences that impact populations at scale" (Lloyd, 2018, p.1). According to NYU professor, Kate Crawford, there exists two types of harm caused by AI applications: representative harm, and allocative harm (Crawford, 2017). Allocative harm occurs when "a system allocates or withholds a certain opportunity or resource" (Crawford, 2017). In other words, allocative harm results from an AI system unjustly withholding a group of people from a resource or opportunity. Conversely, representative harms, "occur when systems reinforce the subordination of some groups along the lines of identity," (Crawford, 2017) which spawn when AI systems reinforce negative stereotypes of a certain demographic of individuals.

The criminal justice system has begun to invest more into AI. Most notably, police departments across the country use AI agents for a strategy known as predictive policing, known as, "the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions" (Perry et al., 2013). From rural counties such as Hanover, Virginia (Hanover County Sheriff, n.d.) to major urban hubs like New York City and Chicago (Collins, 2018) police departments nationwide are adapting these new innovative strategies. Using a predictive approach is certainly enticing for law enforcement. The ability for police to anticipate crime rather than to just react to crime would theoretically allow police to maximize their limited resources while also reducing crime. However, even though police have been quick to integrate predictive policing tools into their policing strategies, a major concern is that the improper use of AI tools could instead result in allocative harm to the same communities the police are supposed to protect and serve. Instead of aiding police to optimize their resources and reduce crime, AI agents could instead magnify current human biases and harmful policing tactics.

As the third biggest police department, with over 9,000 officers sworn in to protect the civilians of the second biggest metro area in the country (Gascón, n.d.), I analyzed the Los Angeles Police Department (LAPD) in an extensive case study. The LAPD is considered to be one of the most innovative and data-driven police departments in the country that also is located in one of the most racially diverse cities in the world. However, over the years the LAPD has been engulfed in controversies regarding racial bias and poor policing strategies. The intention of this research is to focus on this notion of allocative harm induced by AI systems. Specifically in law enforcement, using STS frameworks in order to highlight the limitations of modern day AI systems and explore how these limitations lead to biases to percolate through AI applications, by example of the LAPD.

Literature Review

Osaba & Welser IV illustrates the phenomenon known as the paradox of artificial agency (2017). Defined as the "paradoxical effect that artificial agents, learning autonomously from human-derived data, will often learn human biases—both good and bad" (Osaba & Welser IV, 2017, p17). The significance of this paradox can be further exacerbated by the concept of dirty data, known as: missing data, wrong data, and non-standard representations of the same data (Kim, 2013, p.81). Richardson et al. (2019) expands this definition of dirty data in how it pertains to the use of AI software tools and machines in the criminal justice system. They argue that data from policing records can be “derived from or influenced by corrupt, biased, and unlawful practices, including data that has been intentionally manipulated or ‘juked,’ as well as data that is distorted by individual and societal biases” (Richardson et al, 2019, p.4).

There is certainly no shortage of research conducted revealing the prevalence of racial bias within policing systems nationwide. From stop and search (Warren et al, 2006), to speeding tickets (Anbarci & Lee, 2014), to drug arrests (Becket et al, 2005), to the use of force (Legewie, 2016), and to even when making a decision to shoot black or white criminal suspects in a training simulator (Plant & Peruche, 2005). This history of racial bias exercised by police departments around the nation certainly brings validity to the claims made by Richardson et al. when questioning the both the quality and the integrity of data fed into the AI agents used by the criminal justice system. Even without the direct and intentional manipulation of data suggested by Richardson et al., Brantingham (2017), details the role that implicit bias has in every pathway of predictive policing. From the initial reporting of the crime, to the way police respond, and to the manner in which the police handle the reported crime after arriving at the scene are all individual pathways that can potentially be influenced by implicit bias (Brantingham, 2017).

The abundance of ways that racial bias seeps into the criminal justice system is only compounded by the rather aggressive implementation and progressing dependence of predictive policing across major cities and urban areas throughout America. In 1973 urban planners, Rittel and Weber, theorized the term wicked problem, now a concept widely discussed within STS circles. “There is a whole realm of social and organisational planning problems that cannot be successfully treated with traditional linear, analytical (systems-engineering-like) approaches” (Rittel & Webber, 1973). As described by Ritchey (2005), problems are “wicked” in the sense that they are devious and can lead to unintended consequences when trying to resolve them. On the other hand, tame problems are well defined and stable in scope and have solutions with definite stopping points that can be evaluated as right or wrong (Ritchet, 2005, p.2).

Criminal justice policy is a wicked problem; it is a culmination of public policy, social structure, complex human interactions, conscious decision making and much more. Thus, a question we must ask is, is the reliance of AI agents in making critical decision outcomes in law enforcement a tame solution to a wicked problem? We as a society already understand about the existence of both intentional and implicit bias within the criminal justice system, and thus need to retarget the conversation to discuss how these same biases persist in the data used for predictive policing methods. Without properly mitigating bias in the data and algorithms used to train AI tools, we risk further inflaming the preexisting social disparities suffered by already marginalized groups. We must analyze the current methods used by predictive policing in order to assess the source of its shortcomings and investigate the extent of the role it plays in police departments across the country. By these means, we can better address the issue of allocative harm inflicted by the prevalence of AI systems in the criminal justice system.

The Paradox of Artificial Agency in the Criminal Justice System

To discuss the ways AI machines mirror human bias, I will refer back to the paradox of artificial agency (Osaba & Welser IV, 2017). In order to produce substantive outcomes, AI machines rely on data. The power of AI is its ability to quickly compare large sums of data in intricate ways and use that as a baseline to realize correlations and patterns (Dickson, 2018). Essentially, AI uses previous outcomes as a method to predict future ones. And, while the recent rise of the use of AI systems is a testament to its potential, the reality is that any given AI agent's capabilities are severely limited by the data it's trained against. Despite the importance of quality data, a study conducted by Harvard revealed that only 3% of companies' data met basic quality standards within their tested sample size of 75 companies (Neagle et al, 2017). The results of this study demonstrate widespread neglect of data cleansing and improper practices of data collection.

For entities such as the criminal justice system, having poor quality data could ultimately risk the wellbeing of the communities they are meant to protect and serve. While the criminal justice system has widely adopted AI agents as a means to more efficiently allocate their limited resource, there is serious risk associated with failing to eliminate biased data. "The problem isn't the algorithm, but the dirty data fed into it. That data is, in turn, the product of conscious or unconscious biases in our society" (Compton, 2019). The reality however, is current data collection practices within the criminal justice system leave it highly susceptible to bias. As pointed out by Friedman and Ponomarenko (2015), the lack of democratic policy making within policing results in law enforcement being the both the least regulated and least transparent governmental agencies. "In many jurisdictions, the rules governing policing are not even available to the public" (Friedman & Ponomarenko).

The lack of systematic uniformity and accountability across police departments suggests "police data is a reflection of the department's practices and priorities; local, state or federal interests; and institutional and individual biases" (Richardson et al., 2019, p.199). Instances such as the New Orleans Police Department's backdoor multi million deal with Palantir, a data mining firm, in 2012 exemplify this issue (Winston, 2012). The city's former mayor, Mitch Landrieu, established the partnership as a philanthropic partnership, however, the public and even city council members were never made aware of this decision and the program was able to continue for five years without their knowledge (Winston, 2012). The ability of police departments to collect data and self govern their own policies and standards under minimal oversight demonstrates just how vulnerable their data is to containing bias.

We now see the significance of the paradigm of artificial agency as it applies to law enforcement. Because data collection and policing practices vary from each policing department a given AI agent could perform differently in two given localities. This potential disparity exists as a result of both inconsistency of data quality and inconsistency in policing practices by each police department. In other words, an AI agent will mirror the biases or motives contained within a particular department. Thus, a department that is corrupt, exercises poor policing practices, and does not adhere to adequate data collection standards will make their AI systems more susceptible to bias. A police department that has good practices will not be completely free of bias, but will have more precise AI agents. The performance of these AI agents is a direct reflection of human behavior. The only way an AI agent would not cause allocative harm is if those within the criminal justice system do not create allocative harm to the communities they serve. More importantly, AI tools cannot be viewed as a means to cover up the current shortcomings of our criminal justice system, and any attempt to do so will lead to unforeseen circumstances often resulting in allocative harm inflicted upon already marginalized communities.

LAPD Case Study: Racial Bias

The Los Angeles Police Department (LAPD) is certainly no stranger to facing controversy when it comes to race relations and were most recently scrutinized for racially profiling black and Latino drivers. Despite blacks accounting for 9% of the population they accounted for 27% of traffic stops, and 47% of traffic stops were Latino drivers which is consistent with their population share of the city. Both white and Asian drivers were pulled over at a rate underrepresented of their population share, with white people accounting for just 18% of stops and Asians at only 4%, despite both representing 28% and 11% of the population respectively (The Guardian, 2019). Furthermore, black drivers were four times more likely to be searched during a traffic stop than white drivers and passengers and Latino drivers and passengers were three times more likely to be searched than white drivers and passengers. This is in spite of the fact that white drivers had a slightly higher chance (20%) of being caught with drugs or another contraband when searched, compared to a 17% and 16% chance for blacks and Latinos (The Guardian, 2019).

This is certainly not the first or the most high profile case of the LAPD being accused of racial bias. The infamous 1992 LA riots that left much of the city in flames after a video surfaced of LAPD officers brutally beating a black motorist, Rodney King for allegedly speeding, shed light on the deep distrust between law enforcement and civilians in the city of Los Angeles. Certain communities within the city developed this distrust from a history of LAPD officers accused of engaging in brutality and over-policing. The consequences of over-policing certain communities leads to selection bias. Meaning that actual crime patterns are not accurate to the crimes reported to and/or discovered by law enforcement. Reverting back to the LAPD traffic stops example, if there were no selection bias at play, that would mean the proportion of black drivers pulled over by the police would mirror the ratio of black people that make up the overall population of Los Angeles. Instead, black people were overrepresented in the number of people pulled over by a factor of three. And, within the given sample of drivers who were pulled over, black drivers and passengers were four times more likely to be searched than whites and Latino drivers and passengers were three times more likely to be searched. There are two pathways of selection bias in this particular scenario. Within the already biased sample of drivers who were pulled over, existed another pathway of selection bias for black drivers and Latino drivers who faced a higher chance of being searched.

The significance of this selection bias means that the reported offenses of drug and other contraband offences observed by the LAPD does not reflect the actual pattern of offences committed by the full population. By definition, selection bias "occurs when individuals or groups in a study differ systematically from the population of interest leading to a systematic error in an association or outcome" (Nunan et al., 2017). In Los Angeles, due to the selection bias in traffic stops and searches, blacks and Latinos were overrepresented in the actual crime data and conversely whites were underrepresented. The disproportionate number of black and Latino drivers who were caught leads to the false assumption that this kind of criminal behavior is more prevalent among blacks and Latinos even though they were both less likely to be caught with drugs than white drivers while being searched. Whether intentional or unintentional, the selection bias resulted in black and Latino drivers being more heavily targeted by law enforcement. Conversely for white drivers, even though they were more likely to be caught with drugs and other contrabands while being searched, the number of offences committed by white drivers went underreported because they were targeted less.

This example highlights that police data reflects the practices, motives, and biases of a specific police department. Due to these factors, criminal activity may be more likely to be observed in some communities and more likely to be overlooked in others. A survey of 2,000 residents suggests that there exists a considerable disparity in the public perception of the LAPD. Only 40 percent of black citizens responded yes to the question, do LAPD officers, “treat people of all races and ethnicities fairly,” compared to 70 percent of white citizens who answered yes in the survey (Hersher, 2016). Additionally, a report that analyzed prior arrests in the Los Angeles Unified School District reported that 93 percent of arrests and tickets were given to black and Latino students, as Latino students were 2.6 times more likely to be arrested or ticketed and black students were 4.5 times more likely than white students (Community Rights Campaign, 2013, p.11). The Community Rights Campaign explains this disparity as: “evidence strongly suggests that these racialized punishments are not caused by differences in student behavior, but rather by differences in adult responses to that behavior” (Community Rights Campaign, 2013, p.11).

Even if the LAPD has reasonable evidence to support that certain areas do have higher crime rates than others, that doesn’t justify having selection bias in policing strategies. The purpose of crime data is to accurately model actual criminal patterns, which is only established by having the likelihood of a crime being reported and how police respond to it, independent of the offender's physical appearance and the location of where the crime occured. Two crimes of the same magnitude should be handled equally regardless if the offender is a white driver or is a black driver, and similarly, two crimes of the same magnitude should be handled equally regardless if the offender is a high school student in Beverly Hills or is a high school student in South Central LA.

LAPD Case Study: Low Quality Data

The LAPD's racial targeting of civilians during traffic stops also brought awareness to another issue. In January 2020, at least twenty LAPD officers were suspended which opened up a criminal probe over California's gang database named CalGang (Davis, 2020). The investigation came about after it was discovered that officers within the LAPD, "falsely identified innocent Californians in traffic stops as gang members in an effort to make their police work appear more successful" (Davis, 2020). This misconduct relating to the CalGang database was not just an isolated incident as revealed by a state audit, "CalGang's weak leadership structure has been ineffective at ensuring that the information the user agencies enter is accurate and appropriate, thus lessening CalGang's effectiveness as a tool for fighting gang‑related crimes" (California State Auditor, 2016, p.1). The CalGang database's lack of proper oversight, as revealed by the state audit, lead to instances of faulty or incomplete data entries. The police departments of: Los Angeles, Santa Ana, and Sonoma were only able to demonstrate that one out of the total nine gangs reviewed in the audit met the requirements of CalGang policy (California State Auditor, 2016, p.1). Additionally, the four law agencies examined could not provide sufficient evidence for 13 out of 100 people reviewed in the audit for being included in the database, and 42 individuals were found who had birth dates indicating they were less than one years old with 28 of those individuals admitting to being gang members. (California State Auditor, 2016).

The mismanagement of the CalGang database further led to recurring inaccurate and improper data entries. The audit reported that the law enforcement agencies involved failed to verify that the database records were removed, added, or shared in a manner that maintains proper accuracy of the system or protects the rights of individuals' rights (California State Auditor, 2016). None of the law enforcement agencies conducted required supervisory reviews or had any other mechanisms in place that could mitigate any improper data entries and any internal audits conducted were both comprehensive and lacked transparency. Furthermore, the lack of oversight also lead to violations of state and federal laws (California State Auditor, 2016):

> "Further, the user agencies we reviewed have not fully implemented a state law that took effect in January 2014 that generally requires law enforcement agencies to notify juveniles and their parents or guardians (parents) before adding the juveniles to a shared gang database such as CalGang. As a result, many juveniles and their parents were not afforded the right to contest the juveniles' gang designations." (p.3)

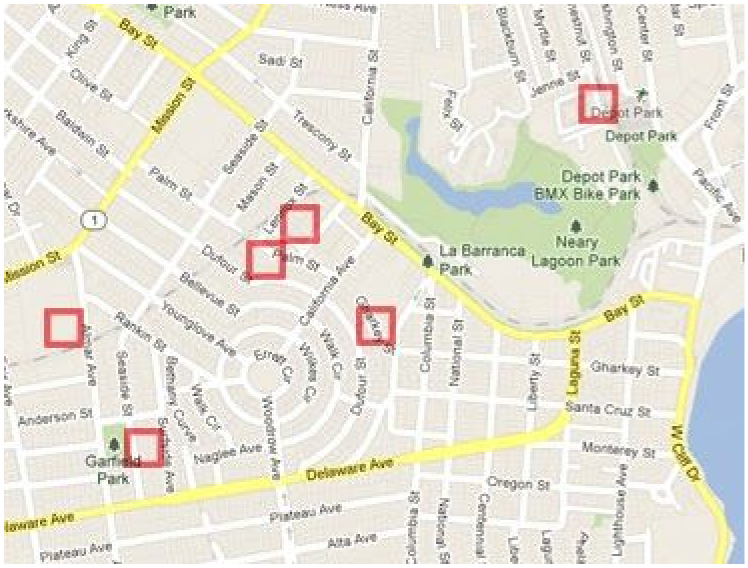
Evident from the state auditor's review, the lack of transparency and accountability of law enforcement’s management of the CalGang database, has led to widespread negligence throughout various police departments in the most populous state in the nation. The audit revealed more misconduct in the management of the CalGang database than what I have discussed; the importance of its findings, however, is it serves an example of the prevalence of database mismanagement in law enforcement. The overall neglect is staggering, especially considering law enforcement data is usually highly sensitive data where any form of database mismanagement could severely damage the livelihoods of innocent individuals. Additionally, without transparent oversight, sensitive data could be used with no proper regulation and undemocratically. In the case of the CalGang database, law enforcement officials stated the database would not threaten the civil liberties of individuals because CalGang only points to source documentation, however, three law enforcement agencies later admitted to using CalGang for employment or military-related screenings (California State Auditor, 2016).

Considering the LAPD was the largest police department involved in the audit, it questions the overall database management of its other systems. Data quality “is a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and whether it's up to date” (Vaughn, 2019), and it's highly relevant in this discussion because of the integral role it plays in the efficiency of AI systems. “In most of the cases, data quality issues explain limited trust in data from corporate users, waste of resources or even poor decisions” (Krasadiskis, 2017). The poor database management practices of the LAPD exposed in the CalGang audit gives reasonable doubt in the department’s overall data quality. CalGang falls embarrassingly short of satisfying some of the main pillars of data quality standards such as accuracy, validity, consistency, completeness among others. The vast majority of private businesses have better database management practices than CalGang and handle much less sensitive information. Any implementation of AI agents by the LAPD would be severely restricted in performance and decision making as a result of their poor data quality standards.

LAPD Case Study: PredPol

The combination of both low quality and biased data leaves any AI agent deployed by the LAPD highly susceptible to producing allocative harm. However, in 2011, the LAPD and mathematicians and behavioral scientists at UCLA developed a software tool used to forecast future crimes based on historical policing data. Today, PredPol is now the market leading predictive policing company used to support law enforcement. Approximately one out of every thirty three people in the United States live in localities whose police departments utilize this software including cities such as Baltimore, Indianapolis, Albany, Atlanta, Long Beach among others (PredPol, 2020). In the words of the company itself, PredPol uses a machine-learning algorithm to calculate its predictions based on historical data (2 - 5 years of data ideally) from a city (PredPol, 2020). Each day, the algorithm forecasts the most high risk areas for crime represented by 500' x 500' boxes on a map as a way to identify locations where police can anticipate crime before it happens

Figure 1: PredPol Map in New Jersey. Map produced from PredPol's algorithm displaying the predicted hotspots.



Note Reprinted from “Defining Predictive Policing”, by Rey, E., 2020

Researchers who developed the algorithm claim to found a similarity between crime and seismic activity (Mohler et al., 2011). Earthquakes are described as self-exciting processes that are triggered over time in consequence of complex processes difficult to predict. At the surface level, earthquakes are abrupt and spontaneous events like a shooting. What PredPol attempts to do is predict the “aftershocks” of a recent crime which is what they say makes a crime self-exciting; once an earthquake happens, the likelihood of another one increases drastically immediately afterwards. Applying this same logic to crime, the moment a crime happens, will spark a series of events that increases the possibility of an after-event. Using this assumption, researchers could then model crime using a mathematical technique called a Hawkes process (Mohler et al., 2011), which is the same class of processes used to model earthquakes.

$$\lambda(t, x, y) = v(t)\mu(x, y) + \sum_{\{k: t_k < t\}} g(t - t_k)(x - x_k)(y - y_k)$$

The self-exciting point process model of burglary (Mohler et al., 2011).

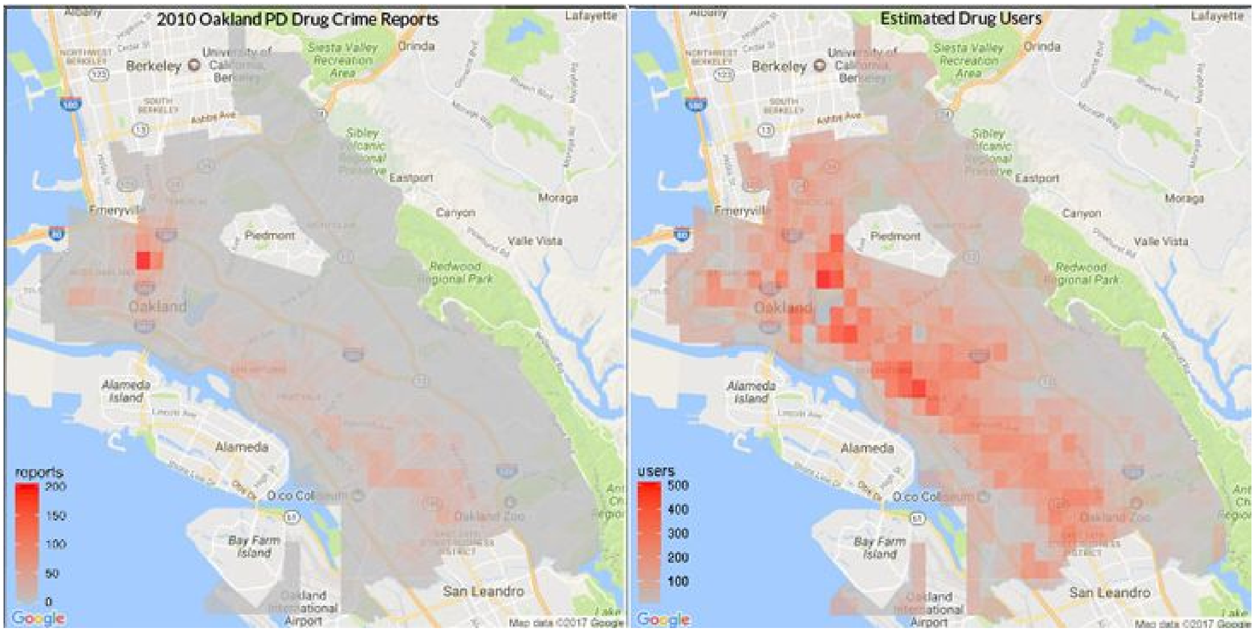
The equation solves for: $\lambda(t, x, y)$, which represents the density of the expected rate of occurrence of crimes in a small neighbourhood around the region (x, y) at time t, conditional upon the history of all occurrences up to that time (Rosser & Cheng, 2019). Essentially, burglars are more likely to return to the same location after an intrusion, following the same aftershock behavior described before. The algorithm also factors the accumulation of prior crimes, making the assumption that historical crime factors into the likeliness of future crime. The algorithm solely relies on three data inputs: crime date and time, crime type, and crime location to fuel its decision making. PredPol believes the few data inputs used in its algorithm removes the potential for bias (PredPol, 2020),

> "No personally identifiable information is ever used. No demographic, ethnic or socio-economic information is ever used. This eliminates the possibility for privacy or civil rights violations seen with other intelligence-led policing models."

The issue however, is not necessarily the algorithm, but instead the data. Earthquakes are not reported the same way that crime is. If an earthquake happens in Los Angeles, there are seismologists all over who can record it unless the earthquake is too weak to even be measured. Earthquakes are discrete events that are objectively quantifiable in the way they are measured in magnitude, giving reasonable confidence that measured seismic activity is an accurate subset of Earth's actual seismic activity. Additionally, there is a universal definition of what an earthquake is. Even though they are unpredictable, their behavior is constricted within the rigid laws set by nature. Crime on the other hand, is a social construct, the definition of crime itself along with its root causes is a conglomeration of a myriad of socio-economic and political influences. There is no universal definition or method of measurement of crime. What could be punishable by death in one place in the world could be perfectly legal in another. While potentially devastating events, earthquakes are tame in the scope of how they are quantified allowing its behavior to be accurately modeled with a single mathematical formula. Crime on the other hand, is a wicked problem that cannot be reliably modeled by a tame mathematical solution. Unlike the recorded data of earthquakes, the LAPD's historical crime data is not an accurate representation of actual crime.

Even if the LAPD could perfectly model criminal activity, their glaring misconduct of racial bias and mismanagement of sensitive data will still produce unintended consequences and allocative harm. To demonstrate this, Lum and Isaac (2016) conducted a study to compare PredPol's performance when trained by data from the Oakland Police Department and the 2011 National Survey on Drug Use and Health (NSDUH). While some drug users do conceal their drug use in public health surveys, Lum and Isaac made the assumption that the data from the NSDUH would still provide a better estimate of actual drug use than police records. Additionally, to back their assumption, the US Bureau of Justice Statistics has favored data from the NSDUH over police reports for measurements of drug use (Langan, 1995). From the two data sources, they constructed two heat maps of drug crime from the Oakland Police Department and estimated drug crime from the NSDUH.

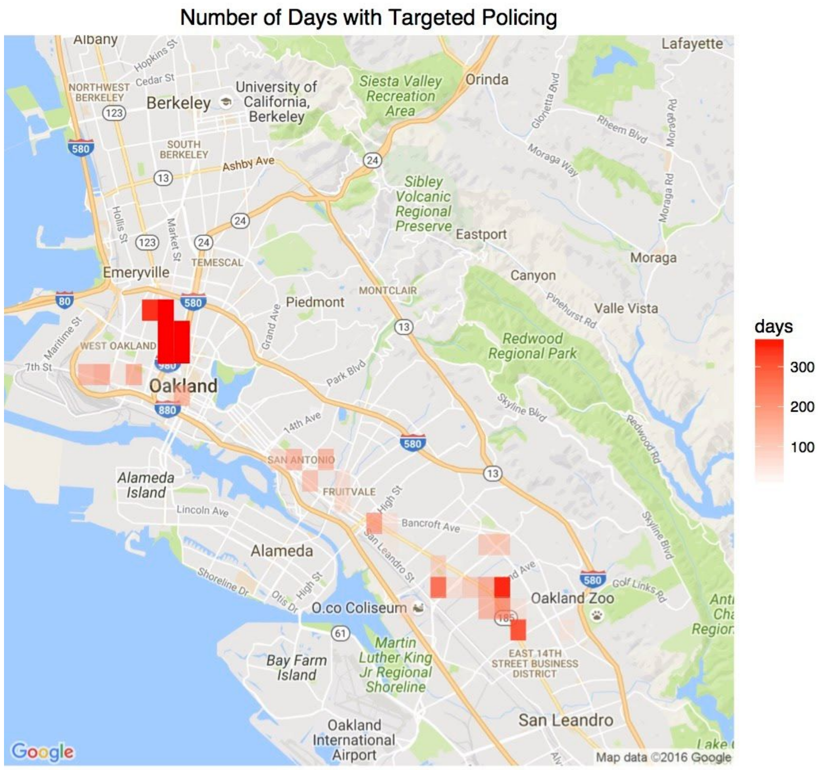
Figure 2: Map on the left shows the distribution of drug arrests and on the right is the estimated drug use in Oakland



Note Reprinted from "To Predict and Serve?", by Lum, K and Isaac W., 2016 *The Royal Statistical Society,* p. 4

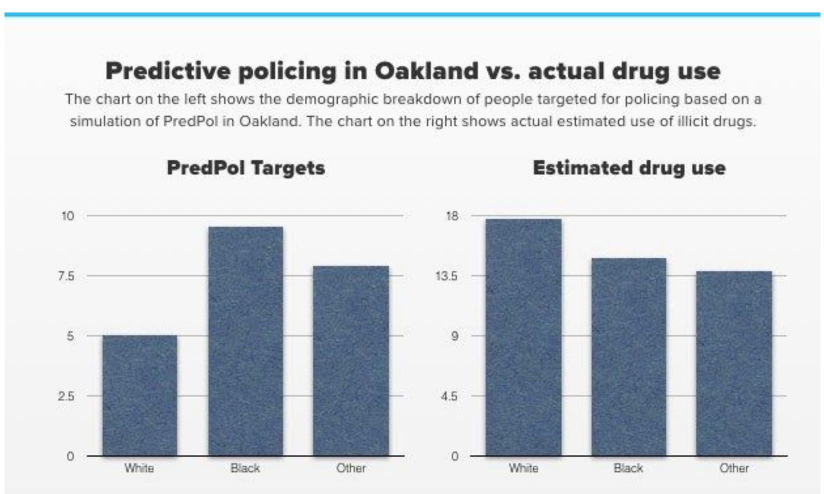
From the data visualization in Figure 2 the discrepancies in drug arrests vs the estimated actual drug use become very apparent. The drug arrests were most heavily concentrated in West Oakland and International Boulevard which are areas with largely non-white and low-income populations (Lum & Isaac, 2016, p. 4). However, the estimated crime data shows a much different pattern with actual drug use being fairly evenly distributed in the city as the discrepancies in drug users shown on the map is more of a result of variations in population density. Interestingly enough, the area with the highest drug arrests was an area with a relatively moderate to low number of total estimated drug users. Lum and Isaac then trained PredPol's algorithm with data from the Oakland Police Department's drug arrest records to see which areas the algorithm would direct officers to target.

Figure 3: Map produced from PredPol's analysis of Oakland Police Department's drug arrest data showing the number of day police should target certain locations.



Note Reprinted from “To Predict and Serve?”, by Lum, K and Isaac W., 2016 *The Royal Statistical Society,* p. 5

Figure 4: Graph showing the racial breakdown of drug arrests vs estimated drug use in Oakland



Note Reprinted from “To Predict and Serve?”, by Lum, K and Isaac W., 2016 *The Royal Statistical Society,* p. 5

Despite PredPol's claim of its algorithm not having bias, the results from this simulation portray a much different reality. Even though white people in Oakland have a considerably higher drug usage than black people or other races, they were targeted at a much lower rate by PredPol's algorithm even with the software including no data inputs regarding personally identifiable information such as race, ethnicity, or socioeconomic status. In response to the study, PredPol's CEO claims that they do not use their software for drug crimes "The reason we do not predict for drug crimes is that these can be selectively enforced in different neighborhoods or by different officers… our practice is to use the most objective data available, and 'drug crime' data does meet our criteria of inclusion" (as cited in Smith IV, 2016). The researchers only used drug crime data because, with drug use, they could use the public health survey to estimate actual drug usage unlike with other forms of crime. Regardless, despite the beliefs of PredPol's CEO, all crime is in fact selectively enforced. The results showed such disparity in drug arrests because certain neighborhoods in Oakland are under-policed and others are over-policed. The police weren't just selectively looking for drug crimes, but instead for all crimes, and drug crime is not some sort of statistical anomaly in crime data.

Even though the PredPol example was just a simulation of drug crime in Oakland, the LAPD helped create the software tool and is heavily relied on by the department today. And certainly, with the LAPD's poor policing practices and low quality data, the police department should not be able to freely implement a tool that is so easily capable of perpetuating bias. An internal audit conducted by the LAPD revealed that predictive policing tools such as PredPol "lacked oversight in their implementation and often strayed from their stated goals" (Macias Jr, 2019). In response, the LAPD completely ended one program named LASER and modified several others. Sarah Brayne, a sociologist at the University of Texas at Austin who shadowed LAPD officers using PredPol stated in an interview “The fact that we don’t know if this thing works is indefensible… This is $900,000 worth of federal grant money for literally, specifically, evidence-based policing, and our evidence is so bad at this point in the game” (Moravec, 2019). Police departments across the country have adopted the use of AI agents in policing as a way to cover up or maximize the distribution of a deficient number of resources, costing not only precious time and money but also the livelihoods of individuals. It is truly telling that the LAPD, one of the most data-driven police departments in the entire country carries such negligence in both their policing strategies and data management.

Moving Forward

We now see the consequence of police departments using AI agents as a means to cover up poor policing practices and limited resources. As shown in the PredPol simulation of drug crime in Oakland, AI tools often lead to harmful feedback loops. Essentially, when AI agents are trained with low quality data where certain communities have been overly targeted, that AI agent will then send police back to those same areas causing police to find more crime making the AI more confident in its already biased predictions. We see the paradox of artificial agency play out in law enforcement, AI agents mirroring harmful policing habits. As a wicked problem, crime cannot be solved with tame solutions. The use of AI agents in policing cannot be viewed as an universal solution. Before police departments adapt more AI technology and predictive policing programs, they should instead adapt better oversight and policing practices first. Police departments with targeting policing strategies and low quality data collection and management standards should not be able to use AI tools in policing. Otherwise, the implementation of these AI tools will only lead to unintended consequences and reinforce bad policing.

With the use of AI accelerating all over, law enforcement is not the only entity that should be more mindful of its practices. Racial bias and poor data collection practices are prevalent in education, healthcare, financial services, hiring, only to name a few examples. The issue is that data is often viewed as the unequivocal truth, as the saying goes, numbers never lie. When actually, data might only tell a partial truth or could be entirely wrong. The way in which data is collected and managed is more important than the data itself. While advancements in the power and capabilities of AI grows, the same emphasis has not been placed on data. Without high quality data, our technological advancements become highly constrained. And, if racial bias continues to seep into our data, we will continue to aggravate the pre-existing racial disparities within society. This is an issue of human behavior, not AI.

Works Cited

Anbarci, N., & Lee, J. (2014). Detecting racial bias in speed discounting: Evidence from speeding tickets in Boston. *International Review of Law and Economics*, *38*, 11–24. https://doi.org/10.1016/j.irle.2014.02.002

Beckett, K., Nyrop, K., Pfingst, L., & Bowen, M. (2005). Drug Use, Drug Possession Arrests, and the Question of Race: Lessons from Seattle. *Social Problems*, *52*(3), 419–441. https://doi.org/10.1525/sp.2005.52.3.419

Brantingham, P. J. (2017). *The Logic of Data Bias and Its Impact on Place- Based Predictive Policing*. http://paleo.sscnet.ucla.edu/Brantingham-2018-OSJCL.pdf

California State Auditor. (2015). *The CalGang Criminal Intelligence System: As the Result of Its Weak Oversight Structure, It Contains Questionable Information That May Violate Individuals' Privacy Rights*. https://www.auditor.ca.gov/pdfs/reports/2015-130.pdf

Collins, D. (2018). Police departments sued over predictive policing programs. https://www.policeone.com/legal/articles/police-departments-sued-over-predictive-policing-programs-oEOyziTEEgyMrLNv/

Compton, J. (2019). *Intel AI BrandVoice: Data Quality: The Risks Of Dirty Data And AI*. https://www.forbes.com/sites/intelai/2019/03/27/the-risks-of-dirty-data-and-ai/#665042a92dc7

Crawford, K. (2017). *The Trouble with Bias.* https://www.youtube.com/watch?v=fMym_BKWQzk

Davis, Z. (2020). *LAPD Chief Says Its Gang Database Abuse Scandal Now Has 'Criminal Aspects'* https://reason.com/2020/01/15/lapd-chief-says-its-gang-database-abuse-scandal-now-has-criminal-aspects/

Dickson, B. (2018). *There's a huge difference between AI and human intelligence.* https://bdtechtalks.com/2018/08/21/artificial-intelligence-vs-human-mind-brain/

Gascón, G. *COMPSTAT Plus—Los Angeles Police Department*. (n.d.). http://www.lapdonline.org/inside_the_lapd/content_basic_view/6364

Hanover County Sheriff. *Hanover Facts, VA*. (n.d.). https://www.hanoversheriff.com/249/Hanover-Facts

Hersher, R. (2016). *LAPD Report Finds Racial Disparity In Approval And Trust Of Police: The Two-Way: NPR*. https://www.npr.org/sections/thetwo-way/2016/11/15/502193667/lapd-report-finds-racial-disparity-in-approval-and-trust-of-police

Kim, W., Choi, B.-J., Hong, E.-K., Kim, S.-K., & Lee, D. (2003). A Taxonomy of Dirty Data. *Data Mining and Knowledge Discovery*, *7*(1), 81–99. https://doi.org/10.1023/A:1021564703268

Krasadakis, G. (2017). *Data quality in the era of Artificial Intelligence.* https://medium.com/ideachain/data-quality-in-the-era-of-a-i-d8e398a91bef

LA police search black drivers most – even though white people have more drugs, report finds. (2019). The Guardian.

http://www.theguardian.com/us-news/2019/oct/08/los-angeles-police-stop-search-black-latino

Langan, P. (1995). *The Racial Disparity in U.S. Drug Arrests.pdf*. https://www.bjs.gov/content/pub/pdf/rdusda.pdf

Legewie, J. (2016). Racial Profiling and Use of Force in Police Stops: How Local Events Trigger Periods of Increased Discrimination. *American Journal of Sociology: Vol 122, No 2*. https://www.journals.uchicago.edu/doi/abs/10.1086/687518

Lloyd, K & Hamilton, B. (2018). *Bias Amplification in Artificial Intelligence Syst.pdf*. https://arxiv.org/ftp/arxiv/papers/1809/1809.07842.pdf

Lum, K., & Isaac, W. (2016). *To predict and serve?* Royal Statistical Society, 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Macias Jr, M. (2019). *Audit Finds LAPD Predictive Policing Programs Lack Oversight*. https://www.courthousenews.com/audit-finds-lapd-predictive-policing-programs-lack-oversight/

Mason, Sandra. (2019). *Bias In, Bias Out*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3257004

Mohler, G., Short, M., Brantingham, P., Schoenberg, F., & Tita, G. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, *106*, 100–108. https://doi.org/10.1198/jasa.2011.ap09546

Moravec, S. (2019). *Do Algorithms Have a Place in Policing?* The Atlantic. https://www.theatlantic.com/politics/archive/2019/09/do-algorithms-have-place-policing/596851/

Nunan D, Bankhead C, Aronson JK. *Catalogue of Bias Collaboration*. (2017). Catalog of Bias. https://catalogofbias.org/biases/selection-bias/

Osaba, O. A., & Welser, W. I. (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1744.html

Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR233.html

Rey, E. (2020). Predictive Policing Definition. (2015). PredPol. https://www.predpol.com/predictive-policing-definition/

Richardson, Rashida and Schultz, Jason & Crawford, Kate. (2019). *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*. 94 NYU Law Review. https://ssrn.com/abstract=3333423

Ritchey, T. (2013). Wicked Problems: Modelling Social Messes with Morphological Analysis. *Acta Morphologica Generalis*, *2*.

Rosser, G. & Cheng, T. (2016). *Improving the Robustness and Accuracy of Crime Prediction with the Self-Exciting Point Process Through Isotropic Triggering* Springer Link. https://link.springer.com/article/10.1007/s12061-016-9198-y

Smith IV, J. (2016). *Crime-prediction tool PredPol amplifies racially biased policing, study shows*. https://www.mic.com/articles/156286/crime-prediction-tool-pred-pol-only-amplifies-racially-biased-policing-study-shows

The Predictive Policing Company. (n.d.). PredPol. https://www.predpol.com/

Vaughan, J. (2019). *What is Data Quality and Why is it Important?* https://searchdatamanagement.techtarget.com/definition/data-quality

Warren, P., Tomaskovic‐Devey, D., Smith, W., Zingraff, M., & Mason, M. (2006). Driving While Black: Bias Processes and Racial Disparity in Police Stops*. *Criminology*, *44*(3), 709–738. https://doi.org/10.1111/j.1745-9125.2006.00061.x

Winston, A. (2018). *Palantir has secretly been using New Orleans to test its predictive policing technology.* The Verge. https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd