**AI and Law Enforcement: The Effects of Predictive Policing**

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 Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

**David Dimmett**

Spring 2022

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 _____

David Dimmett

Signature _____ Date _____

Hannah Rogers, Department of Engineering and Society

# AI and Law Enforcement: The Effects of Predictive Policing


## Abstract

Predictive policing technology uses AI to assist police in being proactive about criminal activity. Langdon Winner's framework of political artifacts shows us politics permeates the creation and use of technology (Winner, 1980). Based on this framework, the politics of racial bias, intelligence-led policing, and big data companies have shaped the use of predictive policing. AI's history of bias problems can explain how predictive policing can mirror the same bias found in traditional policing. A change in police strategy following 9/11 has led to a promotion of intelligence-led policing, but predictive policing may not meet all the criteria of such a system. The companies, PredPol and Palantir, explain the business interest in predictive policing. Finally, questions on privacy ethics show how public perception of predictive policing is not entirely positive.


## Introduction

Advances in computer hardware and data processing have enabled traditionally manual processes to become data-driven and automated. This transformation has occurred in a variety of industries, from healthcare to manufacturing, with artificial intelligence, or AI, leading the way. AI itself is a broad term that covers several applications, including machine learning. No matter the application, AI strives to replicate human thinking and processes with computation (IBM Cloud Education, 2020).

Law enforcement is one group that has found a use for AI in the form of predictive policing. This technology involves using computation to make predictions either about where crimes will occur, who will commit them, or both. Massive amounts of data feed these prediction algorithms, with common sources of data being surveillance cameras, online activity, and police

records. The goal of these programs is to make better use of police resources and stop crimes before they happen (Lau, 2020).

One example of predictive policing comes from Pasco County, Florida. This case in particular effectively illustrates the police-community dynamic that can result from predictive policing. In their work, "Targeted," McGrory and Bedi document the Pasco County Sheriff's Office predictive policing program and the response from citizens in the community. The program, which was established in 2011, uses an algorithm to generate an initial list of suspects. Each person on the list is given a score that indicates their likelihood of criminal activity. Many data sources feed this algorithm, including social media activity, police reports, and school records. Analysts then manually select suspects for further investigation. This further investigation usually involves visits from officers at home, work, and other places in the community at arbitrary times. From the suspect's perspective, there can be much confusion and fear because they actually have not committed a crime, so their interactions with the police have an increased potential to turn confrontational. With officers writing citations for arbitrary offenses like tall grass, the sheriff's office also encourages officers to be aggressive with suspects. The resulting effect is an increase in tensions between police and members of the community (McGrory & Bedi, 2020).

Clearly, there is more to predictive policing than straightforward data analysis based on the Pasco, County example. In fact, there is a complex web of interests in the adoption and use of predictive policing. In this thesis, we will examine how politics have shaped predictive policing. One, intelligence-led policing has created a role for AI to fill. Second, techniques like machine learning make predictive policing easy, but bias problems strain police-community relationships.

There are many law enforcement entities throughout the world. Given this scale, this paper will limit its focus to issues and police forces inside the United States. In discussing the different aspects of predictive policing, we may compare and contrast with other countries' practices. It should be assumed otherwise that "predictive policing" refers to US predictive policing. This narrower scope still creates a sufficient discussion, since there are many cities and counties that have predictive policing programs. One could argue that this scope is too small to determine all the politics involved in predictive policing, but other countries have entirely different political climates that would need to be explained by a separate paper. For example, China and the United States both have predictive policing systems. Compared to the US, China's system is much more top-down and unified. Additionally, China has used its system to explicitly target minority populations, like the Uyghurs (Davidson & Ni, 2021). These differences show that it is sufficient to discuss US predictive policing alone. Furthermore, this scope will show how the politics and history of a country matter in their relationship to predictive policing.

**Bias in AI and Predictive Policing**

Understanding the history and bias issues in AI can help explain the inequality in predictive policing. In 1950, Alan Turing first conceptualized AI as computers using information to automatically make decisions (IBM Cloud Education, 2020). For much of its history, the real-world use of AI has been limited by computer hardware. Most computers did not have enough memory or computational power for AI. With cheaper hardware, more industries have access to AI capabilities. Furthermore, massive amounts of data can be collected and shared over the internet to train these AI models. Often, these models take the form of neural networks. Researchers were writing about neural networks as early as the 1940s. Frank Rosenblatt's 1967 perceptron was an early example, and many modern techniques, like deep learning and

reinforcement learning, retain some of the original perceptron's ideas. All of these factors have contributed to AI's increased scale and presence, far exceeding what was originally predicted of the field (*What Are Neural Networks?*, 2020).

Neural network applications are essentially black boxes in that it is not possible to view the reasoning behind a network's output. This fact becomes more problematic when bias is involved. An interesting case study of machine learning bias comes from the ImageNet database. ImageNet is a popular dataset of images used in computer vision research. The dataset contains 14 million categorized pictures, including people. Importantly, workers manually label the images. The art project ImageNet Roulette by Trevor Paglen and Kate Crawford showed how certain disparities arise with this method. Paglen and Crawford essentially trained a facial recognition program using the ImageNet database and hosted it online. While the program identified white-skinned users by occupation, it identified darker-skinned people mostly by race and more negatively. The program's implementation was not at fault for this issue; rather, it was the labels themselves in the dataset. The workers who had labeled the faces leaked their bias in the process, so the classifications were biased too (Solly, 2019).

While predictive policing may employ facial recognition technology or other machine learning techniques, simpler forms of data analysis can still be used and have bias. What creates bias is the over or under-representation of a certain characteristic or group in a population. As a result, there may be many sources of bias in predictive policing. As an example, police departments will often use past arrest records to feed their prediction algorithms, and keep those records in the system despite any dropped charges or not guilty verdicts. Studies have also shown that police arrest patterns are racially biased. This combination means that predictions made using biased arrest records will suffer the same deficiencies (Brayne et al., 2015).

Data selection can also be more sinister. In a survey, some higher-ranking NYPD officers admitted to manipulating crime statistics. Others admitted to pressuring victims to not file police reports and to planting evidence. For example, some officers planted drugs on innocent people during stops. While this activity was successfully legally challenged, nothing stops the data generated in these activities from entering the predictive policing system. Once there, it can influence predictions just like other biased data (Richardson et al., n.d.).

Data used in predictive policing is a part of the technology, just like the lines of software that control the algorithm or hardware used for computation. The selection of data is an intentional choice and determines the success or failure of the application. However, police departments gloss over this fact. One will often see more idealized language surrounding predictive policing. In this context, predictive policing and AI is "objective" as it relies on computation whereas traditional policing is inaccurate and "hunch" based (Chan, 2021).

Some may argue that such language is simply police departments endeavoring to sell their programs to their city councils. However, it is more likely that predictive policing is part of a larger buy-in trend to AI. Even though self-driving cars have been in development for some time, there are still not any in full use on the road. Despite this, companies like Tesla and Uber continue to pump money into self-driving car research. Likewise, police departments have stuck by their predictive programs, probably because the payoff of success is greater than the continued cost to develop the program. One can imagine the time, resources, and most importantly, lives that would be saved by a truly successful predictive policing program.

**Intelligence-led Policing**

Often the origins of technology are murky, as different inventors or groups may converge on the same idea and approaches at different times. This is the case with intelligence-led

policing, which is an approach to law enforcement that predictive policing falls under. Some claim that intelligence-led policing was born out of the 9/11 terrorist attacks and resulting war on terror (*What Is Intelligence-Led Policing?*, 2020). Meanwhile, others claim that intelligence-led policing came from Kent and Northumbria in England. Regardless of the source, the goal of intelligence-led policing is to utilize data to inform police decisions. In contrast, traditional policing models are more reactive. Furthermore, intelligence-led policing is frequently combined with other policing strategies like community-oriented and problem-oriented policing. The Boston Ceasefire program is an example of this combination. The police worked with Harvard analysts and community leaders to successfully reduce youth gun violence. The analysts first identified where the gun violence came from. Following some more prolific arrests, community leaders and the police worked together to deter further criminal activity among the at-risk population (McGarrell et al., 2007).

On the surface, intelligence-led policing seems to have a positive reputation. The framing of this style of policing is often business-like. Just as a company does market research and focus groups, an intelligently led police department will use data to strategize on ways to fix crime problems (McGarrell et al., 2007). Based on this, it seems reasonable that law enforcement would adopt predictive policing, but upon closer inspection, predictive policing may not entirely fit the intelligence-led model.

A Bureau of Justice Assistance report outlines some key components of intelligence-led policing. These points are illustrative of law enforcement's predictive policing goals. I summarize a few of these points below, which I will discuss later:

- Problem clarity: Having a clear understanding of the problem the intelligence is supposed to help solve

- Effective intelligence: Gathering the right information based on, “reasonable suspicion of relationship to a crime”
- Information Sharing: Intelligence should be shared across teams and departments (Bureau of Justice Assistance, n.d.)

The predictive policing we have discussed so far is more general purpose than what the report seems to recommend based on these points. Another key difference is the wording “reasonable suspicion of relationship to a crime.” Predictive policing casts a wide net, and, as we have seen with the Pasco County example, picks up individuals that do not have much relation to criminal activity.

**Sources of Predictive Policing Software**

Since predictive policing is algorithmic in nature, complex software must be written to craft the prediction systems. Some police departments construct their own systems, while others contract with software vendors. Two vendors which have worked with police departments include PredPol and Palantir. These companies have different backgrounds and motivations in the predictive policing space. Understanding the profit motives and operations of these companies will give further insight into predictive policing.

PredPol is a software as a service (SaaS) company whose main product is their location-based predictive policing software. According to their website, the data points fed into the software include: crime type, location, and timestamp of incident. These data points are supposed to model three aspects of criminal behavior: repeat victimization, near-repeat victimization, and local search. Another important claim is that PredPol information is anonymized and protected. The company has worked with police departments including the LAPD and NYPD (*Predict Prevent Crime | Predictive Policing Software*, n.d.). Although being

an early predictive policing company, PredPol has distanced itself from predictive policing. In March 2021, the company rebranded to Geolitica, citing that predictive policing has come to include, "activities – such as facial recognition or 'predicting' that certain individuals will commit crimes (PredPol, 2021)." It appears then that person-based predictive policing has a certain amount of risk or consequence that PredPol was not willing to accept. Even so, PredPol was found to replicate police racial bias when used by the LAPD (Ahmed, 2018).

In contrast with PredPol, Palantir deals in many more areas than just predictive policing. Palantir is mostly in the market of big data, selling their services to the US government, military, and local police departments. The LAPD and New Orleans Police Department (NOPD) have had contracts with Palantir. In New Orleans, Palantir worked with the NOPD mostly in secret to create a person-based predictive model. Data sources for the project included social media and police data. The NOPD essentially gave Palantir all the data it had, even on people who had never been arrested. A study of a drop in crime during the NOPD and Palantir's partnership could not attribute the decrease solely to Palantir (Winston, 2018b). In LA, police used Palantir's technology to construct a chronic offenders list. A person on this list would receive letters and visits from the police. Instructions even encouraged patrol officers to stop and interview them on the street (Ahmed, 2018).

As discussed above, the actual effectiveness of these two products is not proven. Their products seem to be good at identifying potential criminals, albeit still racially biased. Even with the knowledge of who might be a criminal, it is still up to the police department to prevent crime from happening. It is good for the software company if predictions continue, however, as it means the department will continue to use their software. Without Palantir or PredPol's software being open sourced or some comprehensive study, the true effectiveness of their software will

remain unproven. The data collection these companies carry out also present a challenge to ongoing efforts for data privacy. Usually, discussion of data privacy occurs when a user initiates interaction with some application. For instance, a user logs into Facebook, and Facebook gathers some amount of information about them. For predictive policing, the data collection is harder to track, since an individual has no way of knowing which information they generate will end up in the predictive system. The mere existence of this collected information is also an opportunity for hacks and other abuses of privacy. Discussions concerning user control of data should also include predictive policing, especially since the use of such data is so consequential.

**Privacy Ethics and Predictive Policing**

Recent discourse has surrounded how large technology companies use our data and the rights we have over the data we generate. Companies, like Google and Facebook, rely on selling advertisements based on user data. These ads are often innocuous, but can be more sinister, as was the case with political advertisements in the 2016 Presidential election. These privacy concerns have moved some governments to pass strict regulations, such as the GDPR in the EU. Similar legislation has only been proposed at the state level in the US. A current example is the California Consume Privacy Act. Notably missing from the bill are rules on law enforcement, which can amass considerable amounts of data on individuals (Hospelhorn, 2018).

Police can collect publicly available information on people, along with any data generated from police interactions. The police have also been able to buy data on citizens from social media companies (Cooke, 2016). Coupled with the fact that police departments are often secretive about their predictive policing processes, a person would find it very difficult to find out which information the police had on them and what predictions they have made (Winston, 2018a).

Another consideration is whether predictive policing violates the Fourth Amendment in the Constitution. The Supreme Court has upheld that probabilistic thinking can be reasonable suspicion for a stop, but whether a determination made by an entirely computerized would be upheld in court is unknown. According to Ferguson, it is likely that predictions could not form the entire basis of a stop, as police observation would still be required (Ferguson, 2012).

Concerns about privacy and racial profiling have led to the creation of groups entirely against the practice of predictive policing. One such group is Stop LAPD Spying Coalition. Its stated goal is to protest against data-driven policing. They claim data-driven policing, "is how police automate and justify their racism, violence, and banishment," in their communities.

**Conclusion**

Law enforcement is in a continuing race with criminals to keep up with the pace of crime. With shrinking budgets and pressures to be intelligence-led, law enforcement has turned to predictive policing. As a form of artificial intelligence, predictive policing is billed with AI's magic promise of being able to do what a human can do better. This idealization overshadows problems with racial bias, which seems almost unavoidable with the current use of predictive policing. Meanwhile, the benefits of these systems seem no better than traditional policing. In all, predictive policing is representative of how a technology can interact with different groups and the politics of inequality.

Two actions may affect the perception of predictive policing. Open sourcing the methods and algorithms used in predictive systems may increase trust between the community and the police. Disclosing the data used could also increase trust. Second, enabling citizens to view and challenge data collected against them would further increase trust. Such a provision goes in hand with empowering people to control their digital footprint.

## Works Cited


Ahmed, M. (2018, May 11). Aided by Palantir, the LAPD Uses Predictive Policing to Monitor Specific People and Neighborhoods. *The Intercept*. https://theintercept.com/2018/05/11/predictive-policing-surveillance-los-angeles/

Brayne, S., Rosenblat, A., & Boyd, D. (2015). Predictive Policing. *Data & Civil Rights*, 11.

Bureau of Justice Assistance. (n.d.). *Reducing Crime Through Intelligence-Led Policing*. Retrieved March 27, 2022, from https://bja.ojp.gov/sites/g/files/xyckuh186/files/Publications/ReducingCrimeThroughILP.pdf

Chan, J. (2021). The future of AI in policing: Exploring the sociotechnical imaginaries. In *Predictive Policing and Artificial Intelligence*. Routledge.

Cooke, K. (2016, October 11). U.S. police used Facebook, Twitter data to track protesters: ACLU. *Reuters*. https://www.reuters.com/article/us-social-media-data-idUSKCN12B2L7

Davidson, H., & Ni, V. (2021, October 19). Chinese effort to gather 'micro clues' on Uyghurs laid bare in report. *The Guardian*. https://www.theguardian.com/world/2021/oct/19/china-predictive-policing-surveillance-uyghurs-report

Ferguson, A. G. (2012). PREDICTIVE POLICING AND REASONABLE SUSPICION. *EMORY LAW JOURNAL*, *62*, 67.

Hospelhorn, S. (2018, November 5). *California Consumer Privacy Act (CCPA) vs. GDPR*. https://www.varonis.com/blog/ccpa-vs-gdpr

IBM Cloud Education. (2020, June 3). *What is Artificial Intelligence (AI)?*

https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

Lau, T. (2020, April 1). *Predictive Policing Explained | Brennan Center for Justice*. https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained

McGarrell, E. F., Freilich, J. D., & Chermak, S. (2007). Intelligence-Led Policing As a Framework for Responding to Terrorism. *Journal of Contemporary Criminal Justice*, *23*(2), 142–158. https://doi.org/10.1177/1043986207301363

McGrory, K., & Bedi, N. (2020, September 3). *Targeted*. https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/intelligence-led-policing

*Predict Prevent Crime | Predictive Policing Software*. (n.d.). PredPol. Retrieved March 28, 2022, from https://www.predpol.com/

PredPol. (2021, March 2). *Geolitica: A New Name, A New Focus*. https://blog.predpol.com/geolitica-a-new-name-a-new-focus

Richardson, R., Schultz, J. M., & Crawford, K. (n.d.). DIRTY DATA, BAD PREDICTIONS: HOW CIVIL RIGHTS VIOLATIONS IMPACT POLICE DATA, PREDICTIVE POLICING SYSTEMS, AND JUSTICE. *NEW YORK UNIVERSITY LAW REVIEW*, *94*, 42.

Solly, M. (2019, September 24). *Art Project Shows Racial Biases in Artificial Intelligence System*. Smithsonian Magazine. https://www.smithsonianmag.com/smart-news/art-project-exposed-racial-biases-artificial-intelligence-system-180973207/

*What are Neural Networks?* (2020, August 17). https://www.ibm.com/cloud/learn/neural-networks

*What is Intelligence-Led Policing?* (2020, October 2). Kent. https://onlinedegrees.kent.edu/sociology/criminal-justice/community/intelligence-led-policing

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.

Winston, A. (2018a, January 27). Transparency Advocates Win Release of NYPD "Predictive Policing" Documents. *The Intercept*. https://theintercept.com/2018/01/27/nypd-predictive-policing-documents-lawsuit-crime-forecasting-brennan/

Winston, A. (2018b, February 27). *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