Artificial Intelligence and Predictive Policing:

A Roadmap for Research

Ethics + Emerging Sciences Group | Cal Poly

Artificial Intelligence+ Predictive Policing
AN ETHICAL ANALYSIS

Primary Authors:

Ryan Jenkins
ryjenkin@calpoly.edu
Ethics + Emerging Sciences Group
Philosophy Department
College of Liberal Arts
California Polytechnic State University, San Luis Obispo

Duncan Purves
dpurves@ufl.edu
Department of Philosophy
College of Liberal Arts and Sciences
University of Florida

Thanks:	Kiana Alikhademi; Ami Asar; Olivia Chan; Emma Drobina; Eric Grossman; Humberto Itriago; Sterling Lankford; David McCullough; Diandra Prioleau; Irina Raicu; Zachary Rentz; Briana Richardson; Jessie Smith; Jonathan Suarez; Shelby Trudeau; Cal Poly College of Liberal Arts
Cover photo:	Joel Filipe, Unsplash
Suggested citation:	Jenkins, Ryan and Duncan Purves. "Artificial Intelligence and Predictive Policing: A Roadmap for Research." Report funded by US National Science Foundation, awards #1917707 and #1917712. September, 2020.
Stable link:	http://www.aipolicing.org/year-1-report.pdf
Last major revision:	September 30, 2020
Version:	1.0.2

Contents

Introduction

Against a backdrop of historic unrest and criticism, the institution of policing is at an inflection point. Policing practices, and the police use of technology, are under heightened scrutiny. One of the most prominent and controversial of these practices centrally involves technology and is often called "predictive policing." Predictive policing is the use of computer algorithms to forecast when and where crimes will take place — and sometimes even to predict the identities of perpetrators or victims. Criticisms of predictive policing combine worries about artificial intelligence and bias, about power structures and democratic accountability, about the responsibilities of private tech companies selling the software, and about the fundamental relationship between state and citizen. In this report, we present the initial findings from a three-year project to investigate the ethical implications of predictive policing and develop ethically sensitive and empirically informed best practices for both those developing these technologies and the police departments using them.

> *"We are at a crossroads… the world of policing is reeling."*

* * *

One of the most promising applications of artificial intelligence (AI) is to identify patterns, which are used for prediction: algorithms trained on massive data sets are used to predict health outcomes, stock market activity, driving behavior, and the likelihood of recidivism for convicted criminals (Dery 2016). Algorithmic prediction is also changing policing: law enforcement resource allocation is increasingly guided by crime forecasts generated by predictive algorithms. This is so-called "predictive policing": the use of predictive algorithms trained on massive troves of historical crime data to anticipate and respond to crime. Predictive policing is attractive to police departments because it seems to promise cost-effective crime analysis that is more accurate, and less susceptible to bias, than human-generated crime analysis.

Recently, predictive policing has come under withering criticism from civil rights groups, academics, media, and the communities that have been subject to the practice. These criticisms include charges that it replicates or amplifies racially biased patterns of policing; that it unfairly burdens marginalized communities; or that it infringes the liberty of targeted communities. We are currently undertaking a project to evaluate the technologies that enable predictive policing in light of these criticisms, and to recommend best practices for both computer and data scientists developing these technologies, police departments that are considering using data-driven technologies to inform their operations, and policy-makers. This report presents many of the insights from the first year of this project.

As the principal investigators of this project, we invite the academic community and the public to see this report as a preliminary roadmap for research into the ethical, legal, and social implications of predictive policing technologies. As this project unfolds into 2021 and beyond, we propose the collection of issues here be seen as a guide to future investigations and conversations, demonstrating the relationships between various ethical challenges as well as the collaborative, interdisciplinary effort required to meet them.

Many of our experts expressed skepticism about the effectiveness of defunding the police—at least where defunding was taken to be a sweeping move to significantly reduce funding for law enforcement. However, a commonly shared view was that society must reconsider the appropriate distribution of burdens for crime prevention across various institutions. We explore this theme more below [SEE Q#21 ON PAGE 21]. Further, and ironically, many observed a close connection between cutting funding for the police and an increased interest in adopting new policing technologies. Predictive policing, after all, was originally touted as a saving grace of police departments facing twin challenges: funding cuts in the wake of the Great Recession, and charges of systemic bias in the face of the death of Michael Brown in Ferguson, Missouri and others. This time around, even if police are driven to adopt new technologies, we expect them to meet with intensified skepticism from the public about the role and accountability of those technologies.

About this Report

This report synthesizes the insights from a group of virtual expert interviews. We have organized the issues below by stakeholder group: (1) designers and developers of predictive policing technologies; (2) police departments and law enforcement agencies who are considering or are already using these technologies; and (3) policymakers who are considering how to regulate the use of these technologies, or how they are best implemented alongside other law enforcement initiatives. Within each stakeholder group, topics correspond to three rough categories: theoretical issues, concrete ethical concerns, and empirical concerns, though each topic tends to overlap with more than one category. Members of the specific audiences can skip to the section that applies to them, but all are invited to read the entire report.

Because the insights of our report are organized to directly address developers, police departments, and policymakers, it may seem that we've overlooked the most important stakeholder group: community members. This appearance is misleading, however. The insights contained herein are intended to be used to *inform* community members about issues of central ethical concern. By laying out some of the key ethical and empirical questions facing predictive policing, this report hopes to overcome one informational hurdle to public engagement.

The report aims to surface the most important ethical issues with these technologies, even if some important issues are left out. Nonetheless, we hope that key stakeholders will find it useful in their deliberations about how to ethically develop and deploy predictive policing, and about whether and how to receive it in their communities.

When organizing and authoring this report, it became clear that this constellation of issues defies neat categorization and division. Whether topics are organized by theme—for example: bias, transparency, the use and abuse of data, etc.—or whether topics are organized by audience, as they are here, central issues straddle the divisions. As a result, we have included hyperlinked cross-references throughout the document.

Interviews with the experts were carried out under the Chatham House Rule (Chatham House n.d.), which allows quoting discussions verbatim, but forbids attribution of any quotation to any particular person. Phrases that you see throughout this report that are in quotation marks—and attributed to "one expert" or "many experts," for example—are taken directly from these discussions. In other places, we have paraphrased experts who expressed similar sentiments.

A note on the term "predictive policing" is in order. Paraphrasing definitions provided in the literature, we define predictive policing as using sophisticated computational methods to collect and analyze data about previous crimes (and possibly non-crime data) in order to predict which individuals or geospatial areas are at increased probability of criminal activity in order to more efficiently deploy policing intervention and prevention strategies and tactics (Meijer and Wessels 2019). "Predictive policing," as we have defined it, does not perfectly capture all of the technologies we examine here. For example, proponents of Risk Terrain Modeling, which seeks to *diagnose* underlying features of an area that make it vulnerable to crime, argue that it is not a form of "predictive policing." Moreover, the term itself has fallen out of favor, plummeting from popularity in recent years as the practice has come under critical scrutiny. As one of our experts said, "Nobody is *for* 'predictive policing' any more." A superior term might be "data-driven policing," but this is overly broad and unhelpfully vague. A more descriptive term might be "artificial intelligence-driven crime-prediction, crime-forecasting, or crime-diagnosing technologies." But this catchall is unacceptably cumbersome. For better or worse, then, we stick to the term "predictive policing" throughout to refer to such technologies, bowing to the term's popularity and comparative simplicity.

This material is based upon work supported by the National Science Foundation under Grant Nos. 1917707 and 1917712. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF). This work has also been supported by a summer research stipend awarded by the Cal Poly College of Liberal Arts.

Overarching Ethical Themes

How do we conceptualize the problem of crime?

A complete assessment of predictive policing requires asking the question, "What is the goal of the system we are developing and deploying?" In the case of predictive policing systems, the answer is very broadly "to prevent crime." But this is too vague to be of much help, because a predictive policing system cannot, on its own, prevent crime. A more helpful answer is "to predict crime," for that is a reasonable goal for a machine learning (ML) system to achieve. But even this is too vaguely specified. We must specify the reasons we have for predicting crime in the first place. Is the reason for predicting crime that we want to know when and where to place officers on patrol, or is the reason that we want to diagnose and correct the underlying features of a place that make it prone to crime? Even the way this question is posed presupposes a law enforcement emphasis on street crime, which is more amenable to spatial analysis. Perhaps our reason for predicting crime should be to locate digital networks of individuals involved in human trafficking or child pornography. An emphasis on crimes perpetrated via digital networks rather than street crime will suggest very different predictive policing systems, and these systems will distribute the benefits and burdens of law enforcement very differently across society. In practice, most predictive policing systems focus on street-level crimes, particularly property crimes, precisely because they are amenable to spatial analysis. But this focus shifts law enforcement priorities in ways that have significant social effects, if, for instance, minorities commit more property crime and whites commit more crime facilitated by digital platforms.

What's wrong with bias?

Perhaps the single greatest criticism of predictive policing offered by its detractors is that it is a thinly disguised form of racial profiling, i.e. technologically-sanctioned race-based discrimination. Yet for all of this discussion, the two crucial claims that compose this objection remain controversial: first, the evidence that predictive policing results in racist or biased outputs is inconclusive[1]; second, the best explanation for why (and when) discrimination is wrongful is disputed by academics. Multiple experts, even those broadly skeptical of the use of police force, and those with a strong sympathy for complaints about unfair bias and discrimination, pointed out that more discussion and research are needed on both points.

Discrimination is often divided into two kinds. (1) First is intentional discrimination, which is probably the kind that comes to mind for most people. Intentional discrimination takes place when a person purposefully treats someone differently on the basis of an irrelevant characteristic, such as by denying them a bank loan because they are black. (2) The second species of discrimination is "disparate impact" discrimination. This takes place when a policy is formally blind to a person's protected characteristics (such as race or ethnicity), but still ends up affecting them differently because of their membership in one of these sensitive groups. Suppose a police force decided to confiscate all of the weapons in a city and, while the city was racially balanced, suppose that all of the guns were owned by white citizens. This policy would have a disparate impact on whites, even though it was formally race-blind. That is, even though the people who set out to craft the policy did not intend to treat whites differently than black citizens, that was the actual effect.

Determining whether a predictive policing system is discriminatory requires identifying the kind of discrimination at issue and whether it is problematic in this case. It is much less controversial that it is problematic to intentionally discriminate against historically disadvantaged groups. Plausible explanations include that it is disrespectful or unfair to decide how to treat someone on the basis of their racial characteristics. This kind of discrimination fails to treat them as a person entitled to equal consideration, and it fails to take their interests into account, discounting those interests for the wrong reason.

However, it's not clear that computer programs can commit intentional discrimination. For one thing, computers cannot intentionally do anything, at least in the way we normally think about intentions. Moreover, we are not aware of any predictive policing algorithms that use race as a factor when generating crime forecasts. A more realistic prospect is that the algorithms use factors *correlating closely* with race, such as ZIP code. In cases where bias

1 A note on terminology is important here. In a sense, all of the outputs of machine learning algorithms are *biased* because the purpose of the technology is to draw distinctions between different groups (Barocas and Selbst 2016). While the terms "bias" and "discrimination" have negative connotations, we point out below that's not clear when and why bias is wrongful [SEE PAGE 5]. We will use "wrongful" or "unethical" bias to clarify when appropriate.

manifests as disparate impacts, can those disparate impacts be sufficiently problematic to render a policy morally wrong (Boonin 2011)? Many examples of disparate impacts seem unremarkable and unproblematic, like the example of confiscating weapons discussed immediately above. On the other hand, if this kind of bias *is* problematic, how can its wrongness be outweighed?

What is the standard of success for predictive policing?

As indicated above, a common refrain among critics of predictive policing is that predictive policing systems will discriminate against people of color. However, whether explicit or subconscious, human decision-makers are far from perfect when it comes to being influenced by racial bias. This raises a key question for a complete ethical assessment of predictive policing: when assessing a predictive policing system's accuracy, transparency or fairness, what is the relevant performance standard? Must current predictive policing systems, to be ethically acceptable, be more accurate, transparent, and fair than even an ideal human agent, or is being more accurate, fair, or transparent than the average actual human decision-maker enough to justify adopting a predictive policing system?

> *"There is no placebo in policing."*

A recurring problem that haunts this discussion is how the effectiveness of a policing intervention can be measured effectively. In one sense, all policing interventions are experiments, since there is no placebo in policing: it would clearly be ethically unacceptable for the police to merely withdraw from a community entirely to establish a "baseline" for crime. Instead, each intervention represents a departure from the status quo, but the status quo already represents a particular policing approach which is itself not neutral. (For example, some scholars would urge the discussion towards even more fundamental questions—for example, to focus not on the issue of how police decide to distribute street patrols, but on how police distribute resources between policing *different kinds of crimes* entirely [SEE PAGE 5; SEE Q#9 ON PAGE 13].) Moreover, even more fundamental is the question of how a baseline for crime can be reliably established. Thus, establishing the effectiveness of predictive policing as one element of policing strategy requires settling methodological disputes in criminology and obtaining data that is notoriously difficult to gather.

Questions for Designers and Developers

1 **What norms of fairness govern technology development?** To what extent are designers responsible for the effects of the technologies they create? What if we have good reason to think these effects are unjust? One perspective is that designers should be concerned about equality in the distribution of *outcomes* of their products—a concern for which predictive policing technologies have been harshly criticized [SEE Q#6 ON PAGE 11]. For example, consider: it might be that being detained before trial is significantly more harmful to black defendants than white defendants, perhaps because black defendants are less likely on average than white defendants to be able to take time off of work. Are the designers of predictive policing algorithms obligated to take this into account? Or is this downstream effect of the algorithm outside the scope of their concern?

We sometimes hold technologists responsible for anticipating the way their creations will interact with the socio-technical world into which they are deployed—whereby seemingly benign design choices can become ethically fraught as a result of their interactions with pre-existing social conditions. When these negative effects are especially pronounced, is it fair to criticize technologists for this oversight?

On the other hand, designers could reasonably complain that they cannot be concerned with *all* of the downstream effects of the things they create. If all they need to be concerned with is *procedural fairness*, then perhaps they can discharge their obligation by avoiding the use of certain variables in the design of their products. Relatedly, for that matter, what kind of features are appropriate to use? For example, is it morally objectionable to use features of a person in determining a risk score that are outside of their control?

2 **How should developers balance accuracy and fairness when designing their systems?** There is a well-known tradeoff in machine learning: while it is relatively easy to optimize an algorithm for either accuracy or for fairness—where "fairness" is defined mathematically so as to conduce

to measurement[2]—it is challenging to optimize for both simultaneously (Kamiran and Calders 2012).

Suppose, for instance, that we find that a predictive policing algorithm assigns risk scores to areas of a city in such a way that areas of predominantly black communities are labeled as high risk 30% of the time, and areas within predominantly white communities are labeled high risk 20% of the time. Suppose we want to eliminate this racial disparity because of a fairness-based concern about historical discrimination against black citizens by the criminal justice system. There are numerous methods for doing this (Berk et al 2018), but, in general, eliminating the racial disparity will require reducing the algorithm's reliance on race itself, but also any factors that *correlate* strongly with race, when making risk classifications. When those correlating factors are highly predictively useful, eliminating them *can* (but may not always) reduce the accuracy of the predictive system (Kamiran and Calders 2012). This loss of accuracy can have ethically significant costs, such as a citizen being assaulted in a high-risk area due to a mistaken crime forecast. So, designers and users of the algorithm face a challenge: either (a) find a way to correct the racial disparity that does not compromise accuracy or violate one of our other commitments to fairness or (b) accept some compromise between accuracy and fairness.

3 **What data should(n't) we collect?** What data is AI appropriate for or equipped for? One of the key decision points for developers of any predictive machine learning system is, "What kind of data should we use?" Data that are most amenable to use in algorithmic systems like those used in predictive policing are data that are quantifiable. But many factors that are both aggravators or mitigators for crime are not easily quantified. For example, consider a system meant to predict a person's likelihood of recidivating, such as the COMPAS risk score algorithm (Angwin et al. 2016). That algorithm uses data such as the number of juvenile felony arrests the person has or how many of their friends or acquaintances have been arrested (Angwin n.d.). It does not easily capture information such as whether the defendant has entered rehab, found religion, been reunited with a guardian or parent, and so on. As one of our experts suggested, "Counting one thing often means not counting

2 Even then, there is no mathematical measure of fairness that enjoys consensus support among data scientists. Witness, for example, the dispute between Northpointe, the company behind the COMPAS risk score model, and ProPublica, the investigative website that criticized COMPAS harshly in their landmark piece, "Machine Bias." Much of the disagreement between the two organizations rests on which tests of statistical bias were appropriate to justify the claims being made (Angwin et al. 2016; Larson et al. 2016; Northpointe 2018; Larson and Angwin 2016).

something else." In this case, the things that are easiest to quantify are the aggravators for crime rather than the mitigators.

There are two problems here. One is that, while these data can be captured as binary "Yes" or "No" check boxes, just like the questions on the COMPAS form, these mitigators constitute an "open set," an infinite number of potential factors that reduce a person's likelihood of committing a crime. Second, for whatever reason, these questions are often overlooked when constructing algorithmic prediction systems. This is likely because these data are much more difficult to come by, whereas many agencies regularly collect data about the aggravators of crime. But this stance also reveals a philosophical position that inclines designers and departments towards primarily or exclusively considering aggravators of crime. Ultimately, society's vision of what crime is, and what causes it, becomes distorted by these choices [SEE PAGE 5].

There are many other more mundane, but still important, questions involved. For example: Does the system lump together violent crimes as inputs? Are the sources of data incomplete, and who is under- or overrepresented? And so on (Selbst 2017). Constructing algorithms demands special attention to these questions during the development phases.

4 **What information do the designers of algorithms owe to communities and the police who use them?** One of the most striking features of predictive policing systems is how little information the general public has about how they work. Most are protected as private intellectual property, and some information is withheld by police departments to maintain strategic advantage over potential offenders. But lack of information can make external oversight of the system nearly impossible and secrecy can lead to frustration by community organizations and their leaders, who might want to participate more actively in crime prevention. Sharing the burden of crime prevention with other community organizations or city agencies might require greater informational transparency [SEE Q#20 ON PAGE 20]. This poses a difficult question about how to balance information sharing against strategic advantage and legal protections for proprietary trade secrets.

5 **Is *any* degree of bias too much?** Suppose (a) that certain types of crime occur more often in minority communities and (b) that a particular predictive policing system moderately overestimates the amount of that type of crime in minority communities because the system is partly contaminated by historical data that were created by racist practices in the criminal justice system, such as over-policing of minority populations. It is plausible that both (a) and (b) are true in the real world. In this kind of circumstance, is it permissible to use the system's predictions in deciding where to allocate police resources? It's an open

question how big the difference between the reality and the prediction has to be to make this ethically unacceptable. How much crime reduction would be required before disparate impacts could be justified? [SEE PAGE 6]

6 **Is predictive policing a technologically-veiled form of racial discrimination?** The most prominent ethical objections to predictive policing center on the claim that it is biased, or a form of technologically-veiled racial discrimination. We have known for some time that the data can reflect the biases of the actors who construct it (Silberg and Manyika 2019; Manyika, Silberg, and Presten 2019). There is good reason to think that police focus their activities such as patrols disproportionately in minority communities (Scheindlin 2013; US Department of Justice 2011). Even when intensified police patrols lower crime rates, the greater police presence simultaneously increases the risk of wrongful search, arrest and physical altercation between police and innocent members of the targeted community. Predictive policing thus imposes a risk of harm on innocent members of designated 'high crime' communities. There is strong evidence, for example, that police disproportionately arrest minorities for crimes that minorities and whites commit at roughly equal rates (Ross 2015; United States Department of Justice Civil Rights Division 2015; Police Accountability Task Force 2016; Ferguson 2017; O'Neil 2016; Kochel 2011). Moreover, an analysis of a three-year experiment of predictive policing algorithms in Los Angeles found that arrest rates were higher in areas designated for special attention by a predictive policing algorithm than they were in areas designated for special attention by human crime analysts—in particular, arrests of blacks and Latinos more than doubled, while arrests of whites remained the same (Brantingham, Valasik, and Mohler 2018).

> *"Do we need to fix the humans before we use human-generated data?"*

The claim that predictive policing is biased or discriminatory is a linchpin that connects almost all of the criticisms of the practice. The academic community and the public are in desperate need of empirical data that would confirm or disconfirm the allegation.

7 **Does predictive policing lead to a self-fulfilling prophecy?** One of the most prominent concerns in the public debate about predictive policing's discriminatory potential is that it leads to a kind of feedback loop of escalating police attention visited on people of color. If algorithmic predictions put more police in minority communities, this will lead to more documented incidents of crime, including police contacts or arrests. Data about these arrests are then fed back into the model, which will forecast greater crime in the minority communities, leading to even more police, contacts or incidents, and arrests in

minority communities, and so on (Lum and Isaac 2016). Despite this objection's popularity, it is not clear how many predictive policing systems it applies to. For example, LAPD officials have been adamant that in their experiment with the PredPol predictive policing software, they ignored arrest data and focused on emergency calls for service and crime reports. Crime reports are often the result of members of the community calling police for assistance, but crime reports do still carry a risk of causing a feedback loop. This is because police who *initiate* stops will often end up writing an incident report leading to criminal charges. If police are looking for crimes disproportionately in black communities, this incident report data can be fed back into the predictive system, leading to a feedback loop. However, emergency calls for service do not in any obvious way lead to a feedback loop of escalating police attention [SEE Q#9 ON PAGE 13]. Because of intellectual property protections, it is unclear exactly what type of data is being used by a given predictive policing system.

There are similar worries that the LAPD's PredPol system could have been a harbinger of a more expansive and intrusive form of surveillance—the kind that we see arising in authoritarian regimes like China. While a serious worry, this should be kept distinct from the empirical claim that the predictions of these systems are, in fact, biased.

8 **How can we eliminate bias in the choice of data or the construction and outputs of algorithms?** Nonetheless, there are serious concerns, and some evidence to think, that the crime data used to train predictive algorithms can reflect the biases of its human creators (Selbst 2017). What are the best tests for bias in the training data, in the resulting model, or in the model's outputs, and how can these biases be eliminated? What measures of fairness are available for computer scientists, and which are the most appropriate for this specific application [SEE Q#2 ON PAGE 8]?

A more pessimistic perspective holds that our data are the artifacts of an unequal society—where data are logged and curated by imperfect human beings—thus the data will always recapitulate the biases of its creators. As Kant asked almost three hundred years ago: From the crooked timber of humanity, can any straight thing be fashioned? If the answer is, "No," then we should consider why and under what conditions bias is ethically problematic [SEE PAGE 5; SEE Q#5 ON PAGE 10].

Questions for Police Departments and Police Officers

9 **What is the aim of our predictive systems?** Different predictive policing systems suggest different solutions to crime. Some predictive policing systems generate crime forecasts based on a limited subset of crime data (e.g., the place, time, and location of crime).

Forecasts based on these data can predict when and where crime will occur, but they cannot diagnose the underlying causes of crime. For this reason, such a system lends itself to a patrol- or enforcement-oriented response to crime. If all a police department knows is when and where the crime is likely to occur, the natural response is to send patrol officers to the location in order to deter or apprehend the offender.

> *"Change doesn't necessarily mean reform."*

Moreover, if these systems are trained on data that represent the behavior or activities of the police department itself, such as arrest records or police contacts, then it raises the possibility that these recommendations become a self-fulfilling prophecy, which we address above [SEE Q#7 ON PAGE 11]. While predictive policing technologies are often billed as an unbiased, technologically informed method of reforming police behavior, in this case they may merely serve to legitimize the status quo and stymie innovation in policing tactics. Algorithms that essentially tell departments to continue what they have been doing and veil those recommendations in a layer of opaque technology threaten to "reify police power," as one of our experts worried.

Compare this system with one that incorporates data from non-law-enforcement agencies about features of high crime places. Such a system might, for example, find correlations between poor street lighting or multi-family housing and auto vehicle theft. But here the system has moved away from crime prediction to diagnosis of the underlying causes or strong correlates of crime, and it therefore suggests non-enforcement-oriented solutions. Addressing the underlying features of places that make them vulnerable to crime requires engaging non-law-enforcement agencies like public works, sanitation, or urban planning [SEE Q#21 ON PAGE 21].

A further conceptual question arises with the use of person-based crime prediction. Some of these systems identify citizens who are likely to be perpetrators of crime, e.g. because they are likely to be gang affiliated. Other systems identify people who are likely to be *victims* of a crime. Once again, these different systems naturally suggest different kinds of intervention—and even intervention by different agencies. However, note that it's possible that these lists would have overlap, i.e. some citizens might be likely to *both* be gang-affiliated and, say, for that reason, also likely to be a victim of crime. Police agencies must reflect on the aim of a predictive system before incorporating it into crime fighting operations.

10 **What is the social role of police?** What portion of crime prevention should police be responsible for? A seemingly simple decision about which predictive policing system to implement in fact requires reflecting on the scope of the role that police play in crime prevention. A diagnostic (rather than merely predictive) policing system suggests a model on which law enforcement shares the burden of crime prevention with other agencies. Much of the agitation for "defunding" the police has centered around the suggestion to distribute the responsibility for addressing crime more widely among government agencies, rather than concentrating it within the police. Many of our experts expressed a frustration that much of what police do now seems to be "social control for poverty," i.e. responding to the crimes such as homelessness and addiction for which the root cause is poverty at the last stage in the "life cycle" of crime. Of course, including other city agencies is a drastic change to the status quo of patrol-based crime prevention. It requires that police departments give up control over crime prevention, something they may resist.

11 **What's the appropriate normative analogy for understanding police ethics?** Police ethics suffers from "normative dissonance." It is highly under-theorized, especially compared to military ethics, for example. This is surprising. Both domains concern the use of force by governments, and their apparent isomorphism has not been overlooked by critics (Miller 2016). Still, the military ethics community enjoys a broad agreement about the general framework that governs the behavior of armies during wartime (i.e. just war theory). The same cannot be said for police ethics and, instead, individual jurisdictions differ widely in their practices and norms concerning the use of force—to say nothing of the differences between nations. Police ethics remains a field in desperate need of a systematized, overarching, consensus normative framework.

There are two popular but starkly opposed analogies for police ethics: waging war and fighting disease. Aligning with one of these—or rejecting both—becomes crucial as the field matures, because analogies can suggest normative frameworks, possible justifications, and potential guardrails or criticisms of practices. First, consider the military ethics analogy. Police often think of themselves as warriors under threat, waging a kind of battle against the darker elements of society. Borrowing from the military ethics literature would provide rich, mature frameworks for determining when the use of force is justified (*jus ad bellum*), how force might be permissibly used (*jus in bello*), and against whom it may be directed. For example, how can police distinguish between those who are liable to be harmed and those who are not? Similarly, police might appeal to some doctrines in military ethics—namely, the Doctrine of Double Effect—to argue that it is sometimes permissible to impose grave risks on innocent parties under the right conditions.

Compare this to another popular analogy for law enforcement: that crime is a kind of contagion that society must prevent and treat. This is the "public health" analogy for crime. This analogy is imperfect, but understanding the task of police as a kind of public health mission does provide us with a different normative lens through which to view their operations. It suggests that law enforcement agencies should focus more on identifying and preventing the risk factors for crime, according to the philosophy that an ounce of prevention is worth a pound of cure. (Where the "cure" often means intervening with force, at the end of the "life cycle" of crime.) It suggests that communities should seek to use data and algorithms to identify the causes of crime rather than just the symptoms, as it were [SEE PAGE 3; SEE Q#9 ON PAGE 13]. And the public health literature furnishes us with frameworks for evaluating how society can distribute interventions that might burden the liberty of whole communities in ways that are fair and effective. While this analogy may seem to suggest a kinder, gentler form of policing, note also that quarantine and surveillance are common interventions in public health, and that these bring along their own ethical problems.

The military analogy and the public health analogy both offer advantages and disadvantages. The attraction of each analogy might be due to the fact that police perform a patchwork of services to the community, some of which share features with war, others of which are more akin to public health. Some of the work to take down serious organized crime is more like war, with specific battles, strategies and counter-strategies, technological arms races and so on. Street-level policing of drug dealing, prostitution, gangs and so on, is much more like public health interventions. Often, the individuals involved in street-level crimes are as much the victims of the illegal activity as they are perpetrators. Given the multifarious nature of police work, it may not be that we need to choose between the two analogies, but rather that we need to know when each applies. After a normative framework is settled upon, the urgent question becomes how to operationalize that framework into education and onboarding, acculturation, regulations, training programs, and so on.

Some of the commonalities between policing and both war and public health ethics can be seen in the earliest principles governing policing. Sir Robert Peel, known as the "father of modern policing," established the London Metropolitan

Police Force in 1829. His commissioners established nine principles to guide the conduct of police, which are as important today as ever. A core theme of these principles is that good policing takes a *preventative* approach, taking measures to stop crime before a forcible arrest is required. This is akin to public health approaches which stress the importance of preventative care to avoid the need for an emergency response. Another core theme of Peel's principles is that police should only use physical force only to the extent necessary to restore order. All other means must have been exhausted. This is closely analogous to the Principles of Necessity and Proportionality from Just War Theory, which permit a resort to war only if it is a necessary and proportional response to an aggressor state.

12 **What are the cultural prerequisites for successfully integrating AI-driven technologies?** The relationship between technology and institutional culture is a reciprocal one. On the one hand, certain cultural conditions must be in place before an institution is willing to adopt a technology. On the other hand, adopting that technology will act reflexively on that culture. We should expect that some police departments will be more open to adopting data-driven technologies, and that in turn the practice of policing in those departments will change after those technologies have been adopted. There is already work that suggests that new technologies can pervert the *telos* or goals of an institution (Miller 2010). This question merits more investigation.

> *"There is no purely technological solution."*

The importance of these cultural prerequisites is significant, as imposing a technology on an unwelcoming department can guarantee its failure. Multiple experts with firsthand experience of working with police departments who have adopted data-driven technologies suggested that police view the technologies skeptically—just as any workers would who fear being "de-skilled," replaced, or usurped by a fancy new toy. For example, a police officer with thirty years of experience on the beat is likely to trust their instincts over the outputs of a system that is mysterious or inscrutable. Officers are also sometimes left without clear instructions about what to do once inside of a designated high-risk area (Boba Santos 2020). As a result, police often go where the computer tells them to simply in order to check a box, and then move on. This underscores the need, explored further below [SEE PAGE 23], to create the next generation of data-driven policing technologies as a collaborative effort between police as primary users, ethicists, and other community representatives, and to pilot them mindfully in the communities where they will be deployed to continually evaluate their effects.

An important question for future study will be to examine and catalogue, from both failed and successful adoptions of new technologies within police departments, the cultural prerequisites for technology adoption. What is required of the departments themselves before they can be expected to adopt new technologies critically and carefully and, ultimately, successfully?

13 **What role should predictive policing systems play in strategic decision-making?** Once a police department has a predictive policing system at their disposal, it must determine how to incorporate that system into their strategic and operational decision-making. At this stage a key question arises: at what point in the strategic decision-making process should the predictive system be consulted? Taking the case of place-based predictive policing, two options immediately present themselves: (1) The system could be consulted early on when deciding which macro-level places (e.g., beats, city wards, councilman districts, or neighborhoods) to prioritize when allocating patrols or other resources; or (2) the system could be consulted only when determining where *within* a place to focus patrol attention. Option (2) helps to address the concern that predictive policing systems will lead to racial bias in allocation decisions, because the system could not be used to justify additional scrutiny for entire minority communities. That decision is left to be made holistically, perhaps in consultation with a variety of stakeholders. A third option is to ignore macro places altogether in decisions about resource allocation, focusing solely on the micro places (e.g., blocks or addresses). However, if the predictive system identifies high-risk micro places in a way that clusters them in communities of color, concerns about racial bias resurface.

14 **What do police owe citizens as an explanation of decision-making?** Predictive policing systems, along with many AI systems based on machine learning, have been criticized as *opaque*. A system is opaque just in case the contribution of any single feature of the world to the final prediction—whether the subject of prediction is an individual or a place—cannot be easily understood, either by the human decision-maker or the person directly affected by the prediction. When and why opacity is problematic is one of the central questions currently confronting the development of artificial intelligence. Police departments face a particular version of this criticism: That the police might be making decisions about how to engage with communities, where to patrol, and where to distribute resources (i.e. how to spend taxpayer dollars) based on artificially intelligent systems that they do not understand and cannot explain.

Is this problematic? Might the opacity of these systems undermine critical reflection within police departments about best practices by obfuscating responsibility for decisions that do not actually improve their tactics [SEE Q#9 ON PAGE 13]? Do police owe an explanation of decision-making to communities, city councilmen and women, and judges in the context of civil action that they can understand and scrutinize? How granular or detailed should such an explanation be? Would providing community members with the opportunity to be involved in strategic decision-making help to discharge this obligation?

A possible response is to point out that *human* decision making is often opaque as well: humans are subject to well-documented biases, obfuscation, rationalization, and so on. If AI systems are opaque, they might still be *no worse* than the humans who preceded them [SEE PAGE 7]. What is new about algorithms that would generate a new obligation? The opacity of predictive policing algorithms poses a challenge for policymakers trying to create oversight mechanisms for the use of predictive policing technologies by police agencies, especially because many of these technologies are privately developed, and their source code is often protected by intellectual property laws.

In sum: Does implementing predictive policing technologies change the explanatory burden that police departments have towards their communities, or is that burden *symmetrical* with respect to human or machine decision making?

15 **What is the appropriate role for human oversight and expertise throughout the life-cycle of data collection, analysis, and choice of intervention?** The automation of decision-making in several contexts has caused concern. The fundamental question that presents itself is: Should we trust computers to make decisions that humans used to make? Further, criminal justice seems an especially sensitive context in which to outsource human expertise to machines, since the rights and liberties of constituents are at issue. Consider another context: the military context. There, opponents of automated weapons systems—which could theoretically select and engage human targets without human intervention—have demanded "meaningful human control" of these systems.[3] Likewise, police departments will need to consider the role of the "human element" throughout the process of collecting and analyzing data, and then designing interventions based on those analyses. Current best practices in data science recommend that the use of data be guided by human expertise and hypothesis testing, rather than simply feeding data into a "black box" and trusting the outcome. This also calls for a healthy dose of skepticism and scrutiny.

16 **What evidence is there of the efficacy of predictive policing over other methods of resource allocation?** The evidence base for

3 See, for example, (Santoni de Sio and van den Hoven 2018), and the work of the group Article 36, for example, (Roff and Moyes 2016).

predictive policing is limited. Only a handful of peer-reviewed studies have been published comparing the efficacy of predictive policing forecasts with the efficacy of forecasts provided by human analysts, and only one study has been published investigating racial disparities in predictive policing outcomes (Boba Santos 2020; Meijer and Wessels 2019). Building this evidence base is crucial to answering many of the ethical questions raised in this report [SEE Q#17 ON PAGE 19].

17 **What sort of police response is most effective at preventing crime when paired with predictive policing (patrols versus problem-oriented)?** Predictive policing algorithms merely provide forecasts of criminal activity. The forecasts do not dictate the police response to crime. As discussed above, the police have an array of responses available, some of which are more enforcement-oriented than others [SEE Q#9 ON PAGE 13]. A key question for the future of predictive policing is what police action is most effective in response to a predictive policing forecast. Here again, the evidence base is shallow. At present, there are few studies comparing the efficacy of different police responses to predictive policing forecasts (Boba Santos 2020). The evidence base for hotspots policing, by contrast, is quite robust (Committee on Proactive Policing: Effects on Crime, Communities, and Civil Liberties et al. 2018). Building this evidence base is crucial to answering many of the ethical questions raised in this report.

18 **How does predictive policing change police behavior in designated high-risk boxes?** As one of our experts put it, "Everything we do is shaped by how we see the world." Predictive policing technologies present the world to police officers in a certain way, specifically, by telling them that certain areas are likely to be more dangerous than others. When police are "primed" to see the world this way, it presumably changes the way they behave towards people they might encounter in one of these "high-risk" zones.

> *"As a citizen, is my living in a red box good or bad for me?"*

This raises several concerns. For one, it could prime police officers to be more suspicious, i.e. to treat a person's presence in a high-risk zone as probable cause for a search, or to heighten their level of suspicion even before entering a red box. This raises 4th Amendment concerns. Second, it is obviously important to the citizens who happen to live or be passing through these boxes at the time: does their presence there, in the wrong place at the wrong time, make them *de facto* suspects (Ferguson 2017)?

Questions for Policymakers

19 **How should policymakers weigh disparate impacts against efficiency when evaluating predictive policing programs?** Suppose for simplicity's sake that there are two neighborhoods of equal size in a town. One is entirely white, and one is entirely black. Suppose in this town that the police department's predictive policing system consistently predicts that certain blocks in the black neighborhood are 40% likely to see a burglary during an officer's shift, compared with 10% likelihood in the white neighborhood. Suppose also that, in fact, more burglaries are committed in the black neighborhood, but that the predictive system overestimates their probability because it is contaminated by racially biased data. Were the system not so contaminated, the algorithm would predict that the high-risk blocks in the black neighborhood were 30% likely to see a burglary. So, while the system over-predicts crime in the high-risk blocks in the black neighborhood, because of racial bias, it is still efficient, from the standpoint of crime reduction, for officers to spend more time there. Of course, the algorithm will lead to some disparate impact on the black neighborhood, in part because of racial bias. (In turn, this increased police presence will increase the likelihood of wrongful arrest, search, seizure, etc. for innocent residents [SEE Q#6 ON PAGE 11].) In this case, is it acceptable to use the biased algorithm, so long as its predictions are *close enough* to the ground truth to distribute police patrols efficiently [SEE Q#5 ON PAGE 10]? Is avoiding disparate impact a rigid *constraint* on police activity?

20 **How can communities ensure accountability and oversight of predictive technologies?** Communities must balance competing concerns, namely, the public interest in transparency with the reasonable demand for operational secrecy in domestic security. The police have a reasonable claim to keep secret at least some of their operations and decision making—to do otherwise unacceptably compromises the mission of policing and the safety of police officers [SEE Q#4 ON PAGE 10]. However, secrecy threatens to undermine the ability of communities and citizens to hold law enforcement accountable. Allowing too little oversight exacerbates criticisms that police technologies are anti-democratic, since they are often created by those in society that enjoy structural privileges such as access to education and venture capital, and then employed by those who have a monopoly on the

state-sanctioned use of force. Resolving these democratic deficits requires substantive oversight and participation by multiple parties.

Policy makers must strike an appropriate balance, preserving law enforcement's strategic advantage while permitting operational oversight. This might be a challenging task, because providing adequate oversight of a predictive policing system requires the overseer to be (1) impartial in their assessment, (2) technically competent, and (3) worthy of trusting with sensitive information about crime prevention operations and strategy. Achieving (1)–(3) might require restructuring existing oversight mechanisms or introducing new ones. For example, many police forces in the United Kingdom now have ethics committees, and some even have data ethics committees (West Midlands Police and Crime Commissioner n.d.). Any oversight effort will be complicated by the opaque nature of the complicated computer models most commonly in use in these domains [SEE Q#4 ON PAGE 10].

Relatedly, to provide some measure of accountability to the citizens who are subjected to these systems, communities must decide what these "data subjects" are owed. Consider, as an analogy, the way that credit reports are tabulated and the way they are made accountable. Citizens are entitled to the information from credit bureaus that combines to calculate their credit score. They are free to audit that information and to appeal the information if it is mistaken. A similar approach for predictive policing technologies would guarantee citizens the right to audit and appeal the decisions of those systems. Now, the problem of providing too great a shield of secrecy for the police becomes clear: if public complaints are the primary form of redress and appeal, this becomes effectively useless if the public does not understand how these systems work.

Another approach communities could take is to require these systems to undergo "algorithmic impact assessments" (Selbst 2017)—which New York City recently required something similar of the NYPD (Heaven 2020). This process requires police departments to go through an ethical checklist of sorts, prompting them to consider and justify their choice of data, fairness metrics, interventions, and so on. There are no "correct answers" in this exercise, but the hope instead is that the mere act of asking these questions prompts departments to make explicit the implicit choices that go into implementing these systems which can have serious consequences for citizens.

21 **What allocation of crime prevention resources (city/state/national) is optimal across law enforcement and non-law enforcement agencies?** As recent protests have brought to the fore in the wake of the death of George Floyd in police custody, cities across America are reconsidering the enforcement-oriented approach to crime in their communities. It remains an

open question whether, and to what extent, non-law enforcement agencies (e.g., sanitation and urban planning), can aid in crime prevention. Answering this question is crucial for determining whether and how to make use of predictive policing technology; for, as we saw earlier [SEE Q#9 ON PAGE 13], different predictive policing tools are conducive to different approaches to crime prevention.

Conclusion: Predictive Policing's Second Act

The questions outlined here are meant to serve as a scaffold for future research, to assist in what we hope will be a burgeoning discussion combining insights from the intersection of moral and political philosophy, artificial intelligence, criminology, and police ethics. They are by no means meant to be exhaustive, nor the last word. We have tried to avoid giving recommendations, but rather to point the way towards possible positions and responses, identify intersections between issues, and highlight where more work is needed.

> *"The data is always going to exist, so someone's going to be analyzing it."*

The field of predictive policing is rapidly changing and, in fact, the term has already fallen out of favor. But these questions will not go away. As the cost of collecting, storing, and analyzing data falls to nearly zero, we should expect a proliferation of data analysis tools and algorithmic mediation between the citizen and state. This new incarnation of algorithmically informed law enforcement is now often called "data-driven policing." This is predictive policing's second act. It remains to be seen whether this next generation of tools can improve upon predictive policing in meaningful ways.

We expect the conclusions that arise from these investigations to inform analyses of adjacent technologies, which raise similar issues. These include the state's use of facial recognition, drones, and other surveillance technologies. It includes recidivism prediction and risk scores. More broadly, it touches on questions of the police use of force and the interplay between technology, culture, and behavior. For example, are police emboldened when they possess more militaristic tools of law enforcement?[4]

At multiple levels and across multiple disciplines and professions, the next step is to move from theory to practice. This likely includes ethics education in the engineering education pipeline, cultural change in police departments, enhanced awareness and sensitivity among computer scientists, and empirically informed regulations that balance the competing concerns

4 For a relevant analogous discussion from the military ethics literature, see (Kahn 2017).

sketched here. Moving from theoretical positions to operationalizable best practices continues to bedevil technology ethics, but here the challenge is particularly pronounced and urgent. It will require the input from multiple stakeholders throughout the process of creating and deploying these technologies—but most importantly, it will require this input early on. One of our experts opined, "I don't think there was ever a moment where the people creating the current generation of predictive technologies sat down to think about ethics." As a new generation of data-driven technologies replaces the current one, communities, technologists, and police cannot afford to overlook ethics again.

One of the most urgent questions challenges us to reconsider the provenance and possession of these data in the first place. Is there a way to put data about crime to use for other purposes besides "tip of the spear" law enforcement? How can the tools built for police be repurposed to identify the basic drivers of poverty and social ills in communities, to interdict crime before it materializes? How can this be done in a way that guarantees accountability, democratic control, and oversight by the people being surveilled and analyzed? For example, can these same algorithmic prediction tools be repurposed to analyze policing practices themselves, such as identifying patterns of police use of force, decisions to arrest, etc., turning the tools of surveillance into those of *sousveillance*, i.e. observation from below, in an attempt to equalize the current power differential (Ferguson 2017) [SEE Q#9 ON PAGE 13]?

As these conversations mature, scholars and stakeholders should consider whether a more promising future is one in which the process of technology creation is democratic, transparent, and accountable, where the responsibility for combating crime is shared throughout multiple institutions in society, and where the technological optimism of algorithmic governance is tempered by a serious scrutiny of the ethical and social effects of these tools, even when they are created with the best intentions. We welcome these ongoing discussions and we hope that this report, this continuing project, and the work of other researchers can provide clarity, structure, and nourishment for their development.

About the Contributors

Principal Investigators

Ryan Jenkins is an associate professor of philosophy and a senior fellow at the Ethics + Emerging Sciences Group at California Polytechnic State University in San Luis Obispo. He studies the ethics of emerging technologies, especially automation, cyber war, autonomous weapons, and driverless cars. His work has appeared in journals such as *Ethical Theory and Moral Practice* and the *Journal of Military Ethics*, as well as public fora including the *Washington Post*, *Slate* and *Forbes*.

Duncan Purves is an assistant professor in the Department of Philosophy at the University of Florida. His research focuses on the ethical dimensions of emerging technologies, especially decision-making aided by artificial intelligence. He has published his work in leading philosophy and ethics journals including Phi*losophical Studies*, *Ethics and Information Technology*, and *Ethical Theory and Moral Practice*.

Senior Personnel

Juan E. Gilbert is the Andrew Banks Family Preeminence Endowed Professor and Chair of the Computer & Information Science & Engineering Department at the University of Florida where he leads the Human Experience Research Lab. Dr. Gilbert has research interest in Human-Centered AI, machine learning, advanced learning technologies, usability and accessibility, and Ethnocomputing (Culturally Relevant Computing).

Keith Abney is a senior lecturer in the Philosophy Department and a senior fellow at the Ethics + Emerging Sciences Group at Cal Poly, San Luis Obispo. His work focuses on the ethics of emerging technologies, especially ethics and moral reasoning in robots and other autonomous systems in various domains, from space to autonomous vehicles.

Patrick Lin is the director of the Ethics + Emerging Sciences Group at Cal Poly, where he is a full philosophy professor. Prof. Lin is well published in technology ethics—especially related to security and defense—and is regularly invited to provide briefings on the subject to industry, media, and government.

Experts Interviewed

David Boonin, PhD, is Professor of Philosophy and Chair of the Philosophy Department at the University of Colorado Boulder. Dr. Boonin's research focuses on the areas of applied ethics, ethical theory, and the history of ethics.

Joel Caplan, PhD, is a professor at the Rutgers University School of Criminal Justice. Dr. Caplan's research focuses on geographic information systems (GIS), risk assessment, crime prevention, policing, and police-community relations.

Clinton Castro, PhD, is an assistant professor in the Department of Philosophy at Florida International University. His published work focuses on issues in ethics and epistemology, mostly within the context of emerging technologies.

Nick Evans, PhD, is a Professor of Philosophy at the University of Massachusetts Lowell. Dr. Evans' research focuses on national security, emerging technologies, and the ethics of infectious disease, with a focus on clinical and public health decision making during disease pandemics.

Andrew Guthrie Ferguson, JD, LLM, is Professor of Law at the University of the District of Columbia. He is a national expert on juries, predictive policing, and the Fourth Amendment.

Katerina Hadjimatheou, PhD, is a criminologist and applied philosopher working on the ethical aspects of technological developments and the use of data in security, policing, and criminal justice. She chairs a police force ethics committee in the UK and is a member of ethics committees of the National Crime Agency and Metropolitan Police.

Lyria Bennett Moses, JSD, LLM, LLB, BSc (Hons), is Professor of Law at UNSW Sydney and Director of the Allens Hub for Technology, Law and Innovation. She is an expert on law and technological change, legal and policy issues for artificial intelligence, and legal and policy issues for the use of data analytics for law enforcement and national security.

Andrew Selbst, JD, is an assistant professor at UCLA School of Law. His research examines the relationship between law, technology, and society.

Works Cited

Angwin, Julia. n.d. "Sample-COMPAS-Risk-Assessment-COMPAS-"CORE"." DocumentCloud. Accessed September 24, 2020. https://www.propublica.org/documents/item/2702103-Sample-Risk-Assessment-COMPAS-CORE.

Angwin, Julia, Jeff Larson, Surya Mattu, and Laura Kirchner. 2016. "Machine Bias." ProPublica. May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Boba Santos, Rachel. 2020. "Predictive Policing: Where's the Evidence?" In *Police Innovation: Contemporary Perspectives*, 2nd ed., 366–396. Cambridge: Cambridge University Press.

Boonin, David. 2011. *Should Race Matter? Unusual Answers to the Usual Questions*. New York: Cambridge University Press.

Brantingham, P. Jeffrey, Matthew Valasik, and George O. Mohler. 2018. "Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial." *Statistics and Public Policy* 5 (1): 1–6. https://doi.org/10.1080/2330443X.2018.1438940.

"Chatham House Rule | Chatham House." n.d. Accessed September 24, 2020. https://www.chathamhouse.org/chatham-house-rule.

Committee on Proactive Policing: Effects on Crime, Communities, and Civil Liberties, Committee on Law and Justice, Division of Behavioral and Social Sciences and Education, and National Academies of Sciences, Engineering, and Medicine. 2018. *Proactive Policing: Effects on Crime and Communities*. Edited by David Weisburd and Malay K. Majimundar. Washington, D.C.: National Academies Press. https://doi.org/10.17226/24928.

Dery, Sebastian. 2016. "On the Wrong Side of Algorithms." November 28, 2016. https://medium.com/@sderymail/on-the-wrong-side-of-algorithms-part-1-f266a4c3342b.

Ferguson, Andrew G. 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: New York University Press.

Heaven, Douglas. 2020. "Predictive Policing Algorithms Are Racist. They Need to Be Dismantled." MIT Technology Review. July 17, 2020. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/.

Kahn, Leonard. 2017. "Military Robots and the Likelihood of Armed Conflict." In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, 212. Oxford University Press.

Kamiran, Faisal, and Toon Calders. 2012. "Data Preprocessing Techniques for Classification without Discrimination." *Knowledge and Information Systems* 33 (1): 1–33. https://doi.org/10.1007/s10115-011-0463-8.

Kochel, Tammy Rinehart. 2011. "Constructing Hot Spots Policing: Unexamined Consequences for Disadvantaged Populations and for Police Legitimacy." *Criminal Justice Policy Review* 22 (3): 350–74. https://doi.org/10.1177/0887403410376233.

Larson, Jeff, and Julia Angwin. 2016. "Technical Response to Northpointe." ProPublica. July 29, 2016. https://www.propublica.org/article/technical-response-to-northpointe?token=yIzUwxhFBlPUbcgl8aMwDk7gQ6AdqbgA.

Larson, Jeff, Surya Mattu, Lauren Kirchner, and Julia Angwin. 2016. "How We Analyzed the COMPAS Recidivism Algorithm." ProPublica. May 23, 2016. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Lum, Kristian, and William Isaac. 2016. "To Predict and Serve?" *Significance* 13 (5): 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x.

Manyika, James, Jake Silberg, and Brittany Presten. 2019. "What Do We Do About the Biases in AI?" *Harvard Business Review*, October 25, 2019. https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai.

Meijer, Albert, and Martijn Wessels. 2019. "Predictive Policing: Review of Benefits and Drawbacks." *International Journal of Public Administration* 42 (12): 1031–39. https://doi.org/10.1080/01900692.2019.1575664.

Miller, Seumas. 2010. *The Moral Foundations of Social Institutions: A Philosophical Study*. Cambridge ; New York: Cambridge University Press.

———. 2016. *Shooting to Kill: The Ethics of Police and Military Use of Lethal Force*. Oxford ; New York: Oxford University Press.

Northpointe. 2018. "Response to ProPublica: Demonstrating Accuracy Equity and Predictive Parity." *Equivant* (blog). December 1, 2018. https://www.equivant.com/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity/.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. First edition. New York: Crown.

Police Accountability Task Force. 2016. "Recommendations for Reform: Restoring Trust between the Chicago Police and the Communities They Serve." https://chicagopatf.org/wp-content/uploads/2016/04/PATF_Final_Report_Executive_Summary_4_13_16-1.pdf.

Roff, Heather, and Richard Moyes. 2016. "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons." In . Geneva, Switzerland.

Ross, Cody T. 2015. "A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014." Edited by Peter James Hills. *PLOS ONE* 10 (11): e0141854. https://doi.org/10.1371/journal.pone.0141854.

Santoni de Sio, Filippo, and Jeroen van den Hoven. 2018. "Meaningful Human Control over Autonomous Systems: A Philosophical Account." *Frontiers in Robotics and AI* 5 (February): 15. https://doi.org/10.3389/frobt.2018.00015.

Scheindlin, Shira. 2013. "Floyd v. City of New York." New York Times. August 12, 2013. https://www.nytimes.com/interactive/2013/08/12/nyregion/stop-and-frisk-decision.html.

Selbst, Andrew D. 2017. “Disparate Impact in Big Data Policing.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2819182.

Silberg, Jake, and James Manyika. 2019. “Tackling Bias in Artificial Intelligence (and in Humans).” Article. McKinsey & Company. June 6, 2019. https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans#.

United States Department of Justice Civil Rights Division. 2015. “Investigation of the Ferguson Police Department.” https://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_department_report.pdf.

US Department of Justice. 2011. “Justice Department Findings in Its Investigation of the Maricopa County Sheriff’s Office.” DocumentCloud. December 15, 2011. http://www.documentcloud.org/documents/274910-justice-department-findings-in-its-investigation.html.

West Midlands Police and Crime Commissioner. n.d. “Ethics Committee.” West Midlands Police & Crime Commissioner. Accessed September 25, 2020. https://www.westmidlands-pcc.gov.uk/ethics-committee/.