**Five Ethical Challenges Facing Data-Driven Policing**

**Abstract**: This paper synthesizes scholarship from several academic disciplines to identify and analyze five major ethical challenges facing data-driven policing. Because the term “data-driven policing” emcompasses a broad swath of technologies, we first outline several data-driven policing initiatives currently in use in the United States. We then lay out the five ethical challenges. Certain of these challenges have received considerable attention already, while others have been largely overlooked. In many cases, the challenges have been articulated in the context of related discussions, but their distinctively ethical dimensions have not been explored in much detail. Our goal here is to articulate and clarify these ethical challenges, while also highlighting areas where these issues intersect and overlap. Ultimately, responsible data-driven policing requires collaboration between communities, academics, technology developers, police departments, and policy-makers to confront and address these challenges. And as we will see, it may also require critically reexamining the role and value of police in society.

**Keywords**: big data, policing, algorithms, bias, transparency, accountability

**Introduction:**

The use of data-driven approaches by law enforcement is at an inflection point. Heralded in the 1990s as the next great revolution in police work, almost a decade of criticism has led police departments to rein in their ambitions of fighting crime using big data and computer algorithms [1]. For example, after years of local resistance and national criticism about racial bias and other civil liberties concerns, the Los Angeles Police Department, in the spring of 2020, ended its decade-long experiment with PredPol, an algorithmic crime forecasting system designed to predict the location of property crimes [2,3]. Others have followed suit: in June of 2020, Santa Cruz, California became the first U.S. city to ban predictive policing technology [4]. In January of 2021, the city council in Oakland, California voted unanimously to ban both predictive policing and biometric surveillance technology [5].

At the same time that some municipalities are erecting barriers to the use of data-driven methods in policing, more departments use data-driven methods than ever before. Given the constantly evolving landscape of data-driven policing methods and the inconsistent publicity about the use of such technologies, it is difficult to know precisely how many departments employ data-driven methods, but it is safe to say that most major cities across the U.S., as well as several smaller cities and towns, use one or more forms of data-driven policing technology. On its website, Geolitica (formerly PredPol) claims its software currently “help[s] protect roughly one out of every 30 people in the United States.”[6] Furthermore, even some departments that have discontinued using certain methods still rely on others. For example, despite abandoning PredPol, the LAPD Strategic Plan for 2019-2021 lists “enhancing data-driven policing” as the #1 activity in its initiative to reduce crime and victimization [7].

**Methods**

This paper synthesizes scholarship from several academic disciplines to identify and analyze five major ethical challenges facing data-driven policing. Because the term “data-driven policing” emcompasses a broad swath of technologies, we first outline several data-driven policing initiatives currently in use in the United States. We then lay out the five ethical challenges. Certain of these challenges have received considerable attention already, while others have been largely overlooked; in many cases, the challenges have been raised in the context of related discussions. Our goal here is to articulate and clarify these ethical challenges, while also highlighting areas where these issues intersect and overlap. Ultimately, responsible data-driven policing requires collaboration between communities, academics, technology developers, police departments, and policy-makers to confront and address these challenges. And as we will see, it may also require critically reexamining the role and value of police in society.

Before beginning, however, two caveats are in order: First, our focus in this paper is the deployment of data-driven policing tools in the US context. We draw on international scholarship where applicable, but the distinctive and complex relationship between policing and race in the US makes it challenging to import lessons from case studies outside of the US. With this in mind, we have tried to proceed cautiously. Second, given the limited space available in a journal-length article, it would be impossible to offer an in-depth survey of every piece of scholarship that bears on the challenges we raise in this paper. Therefore, we have not attempted it. Instead we have supported our analysis of the emerging major challenges with what we take to be representative work in the field. What we provide in this article is neither intended to be comprehensive nor to be the final word on data-driven policing. For example, some scholars have raised worries about the prospect of mass public surveillance enabled by automated license plate readers and facial recognition-enabled CCTV cameras of the sort found in the United Kingdom and China. Others have hailed mass surveillance as a potential solution to existential threats to humanity [8–11]. While we touch on these technologies, we do not devote significant space to speculating about large scale societal ramifications of these technologies such as how they might transform the relationship between the citizen and the state. And while we attempt to give adequate attention to so-called “structuralist” critiques of data-driven systems, much more could be said [12]. Nonetheless, we hope what follows gives the reader a sense of some of the obstacles and open questions facing any attempt at a holistic assessment of data-driven policing.

**What is data-driven policing?**

Data-driven policing is part of a larger “big data” revolution across sectors, including medicine, finance, transportation, and criminal justice. What characterizes big data systems, of which data-driven policing is one instance, is that they make use of vast quantities of data from a variety of data sources, and, using rapid computer processing and algorithms, identify correlations or patterns in the data--often patterns that humans would otherwise overlook--in order to classify new instances in the domain of interest [8,13,14]. This process of pattern recognition and classification occurs through the use of computer algorithms, which are sometimes developed through the process of machine learning [15]. A computer algorithm is a set of logical operations expressed in computer code designed to solve a problem or accomplish a task (e.g., rank-ordering websites on the basis of an input keyword). Machine learning is a method of computer science by which a computer system is “trained” to reason or make inferences from inputs. In contrast with a method whereby inference rules are hand coded for the system to follow, the aim of machine learning is to enable the system to develop its own rule or set of rules from a given set of examples in order to classify new instances in the domain of interest. This requires providing the computer system with a large set of examples and a set of tools to enable it to infer rules for correctly classifying new instances [16].

With this background in mind, data-driven policing can be understood as the use of big data systems to make classifications of interest to law enforcement. As we discuss below, machine learning can produce data-driven policing systems whose operations are very difficult for humans to understand, when the example sets from which the system learns are very large and the classification rules the system learns are very complicated. At its core, then, data-driven policing involves police agencies making decisions by harnessing vast quantities of data and identifying patterns in that data with assistance from computer systems. This characterization of data-driven policing includes a variety of big data applications in law enforcement. Let us consider some examples.

The LAPD pioneered perhaps the most well known and widely critiqued data-driven policing application. Starting in 2011 the LAPD began an experiment with PredPol. PredPol is a computer program developed by anthropologist Jeffrey Brantingham and computer scientist George Mohler. Its purpose is to predict crime using a proprietary computer algorithm trained on data about the type, location, and time of crime. Premised on a criminological phenomenon known as the “near repeat effect,” wherein one crime event in an area tends to give rise to a spike in similar crimes in that area, PredPol was used by the LAPD to forecast future property crime—vehicular theft and theft from a vehicle—at highly specific locations and times during a police officer’s patrol shift [17]. These forecasts are displayed on a computer screen as “heat maps” of 500x500 square foot boxes indicating high-risk areas in the officer’s beat. Additional information about time, location, and type of recent crime incidents can be incorporated each day to update the algorithm’s forecasts. While the LAPD discontinued its use of PredPol in the spring of 2020, PredPol has become one of the most widely used pieces of predictive policing software in the United States [18]. Internationally, since 2015, the city of New Delhi's CMAPS program (Crime Mapping Analytics and Predictive System) has provided police with algorithmically-determined maps of predicted crime hotspots, based on historical data and continuously updated data from emergency calls and “first information reports” [19].

Whereas PredPol was primarily used by the LAPD to assist in-house decision-making about where police officers on patrol spent their uncommitted time, other data-driven systems are advertised as being more holistic in their application. Risk Terrain Modeling (RTM) is a crime forecasting system developed by Rutgers criminologists Joel Caplan and Leslie Kennedy. RTM is theoretically grounded in so-called “opportunity theory,” which explains criminal activity in terms of the social and physical features of places that give rise to opportunities to commit crime [20]. To understand where opportunities for crime are most abundant, one must be able to identify the features of places that give rise to those opportunities. As such, RTM forecasts crime at a location by incorporating in its model the features of a location that are the underlying causes of crime. Once the features of locations that give rise to crime are identified, geographic units are assigned a value depending on the presence or absence of those features. The higher the value at a location the greater the risk of crime at that location. One key difference between RTM and PredPol is that RTM forecasts take into account both crime data and information about the features of a place that make it vulnerable to crime, whereas PredPol merely makes use of past crime data. A further difference is that RTM is not designed to predict *specifically* when crime will occur. It simply identifies areas facing the greatest risk. It therefore does not lend itself as easily to allocating officer patrols. Instead it suggests solutions to crime that target the features of places that give rise to crime. In one project, RTM was used to identify locations in need of measures to prevent shootings in New Jersey. Spatial features indicating areas of greatest need included the addresses of known gang members, retail businesses and drug arrests for sale or possession [21].

Still other data-driven policing methods are designed not to predict high-risk places but high-risk people. Around the same time that the LAPD began its experiment with PredPol, it adopted another predictive policing method called “Operation LASER,” short for Los Angeles Strategic Extraction and Restoration Program [22]. Part of Operation LASER involved identifying a list of “chronic offenders.” The list of chronic offenders was compiled using information gathered from patrol units, field interview cards, traffic citations, crime and arrest reports and criminal histories. Chronic offenders are ranked by a points system. Points-accruing features included having a violent crime conviction, a known gang affiliation, having been arrested for possession of a handgun, and police contacts. From this list, the LAPD generated “Chronic Offender Bulletins,” which included information about the individual’s arrest history, physical markers, gang affiliation, parole status, warrants, and recent police contacts, among other information. These bulletins were circulated within departments so as to inform officers on patrol about which individuals required the most attention [13]. After an internal audit in 2019 revealed gross inconsistencies in the application of Operation LASER by officers in the field, LAPD discontinued the program.

Person-based policing programs remain in use by police departments across the United States and abroad. For example, the Tampa Bay Times recently revealed some of the inner workings of Pasco County Sheriff’s Office’s Intelligence Led Policing program [23]. Part of the Sheriff’s Office’s intelligence-led policing approach is to look for “prolific offenders” in Pasco County, described in the intelligence-led policing manual as individuals who are “not likely to reform.” The prolific offender designation is determined in part by using an algorithm to assign points to members of the community if they have been arrested for a violent crime, narcotics violation, theft, or if they are *suspected* of any such offenses. Point “enhancements” are assigned for involvement in five or more crime reports either as a reporting person, victim, or witness. The more points a person has, the higher risk they are deemed to be of committing a future crime. District analysts review the points assigned to individuals by the algorithm and may designate them as prolific offenders. These individuals are then singled out for greater police scrutiny. We say more about Pasco County’s Intelligence Led Policing program below.

The London Gangs Matrix has been used by the London Metropolitan Police since 2011 to identify individuals at high risk of involvement in the city’s gang activity and includes both individuals who are associated with committing gang-related crime as well as victims. The London Gangs Matrix is a database jointly maintained by the different borough divisions of the Metropolitan Police of the city of London. Its purpose is to track individuals – especially youths – who are likely to be involved with gangs. Each borough uses its own standards to determine who should be included (or “Matrixed”), but the official operating procedures require that a police recommendation for inclusion be corroborated by at least one other source or partner organization, such as local social services authorities. Once an individual has been made a “gang nominal”, they are assigned an algorithmically-determined, color-coded status meant to show their likelihood to be involved in gang violence. Those coded as “Red” are supposed to be the most likely to be involved in violent crimes, followed by the moderate-risk “Amber” and the low-risk “Green” designations. Risk scores are based on history of violence according to police records, assessments by police intelligence analysts, and assessments or recommendations from partner organizations. At any given time, a significant portion – roughly 38% prior to 2019 – of the Matrix will be made up of nominals who are not only Green but have a risk score of zero, meaning that the police and partner organizations have determined that they are involved with gangs but have no chance of being involved in violence [24,25]. Individuals who have been Matrixed do not suffer any explicit legal consequences. They do, however, seem to face increased police surveillance. Although no causal relationship between inclusion on the Matrix and police activity can be assumed, nominals are more than twice as likely as non-nominals to be stopped and searched [24].

In addition to these systems, police departments have begun to employ methods relying upon and furthering the use of existing technologies. For example, facial recognition technology uses algorithms that allow police to identify wanted criminals or suspicious individuals without the use of human analysts. Such systems rely on existing surveillance technology infrastructure, such as CCTV and social media, and are expected to be integrated into police-worn body cameras in the near future [26]. Large-scale trials of live facial recognition software have recently been performed by the London Metropolitan Police, and similar technologies are already widely used by the Chinese government [27–29].

Facial recognition methods raise ethical concerns that are distinct from other algorithmic systems we have discussed. For one thing, facial recognition systems collect (and store) data on countless civilians via widespread and persistent surveillance in various public spaces. This introduces heightened risks of privacy violations, misuse or abuse, and accidental harm to innocents. While these harms are not entirely unique to facial recognition, they do arise in a distinctive form as a result of these systems. By contrast, the other algorithmic programs we have canvassed focus in general on existing crime data and other relevant demographic data. Much of the ongoing scholarly debate surrounding facial recognition underscores these and other issues [30]. However, it is important to understand facial recognition not simply in isolation, but as part of a broader pattern of algorithmic policing, especially as these different systems are often used in tandem.

Having laid out some common data-driven policing applications, we now turn to several of the most significant challenges for the use of data-driven policing that have emerged from the scholarly literature. Several of these challenges concern the specific conditions of implementation of the applications; other challenges call into question some of the goals of policing that using the applications presupposes.

**Challenge #1 Racial bias and civil liberties**

The most prominent objection to data-driven policing found in academic and media publications is that it leads to discriminatory treatment against people of color, or against economically disadvantaged classes. This discriminatory treatment can result from a combination of two flaws in some predictive policing systems. First, the data used to predict high-risk places and people can have dubious origins. According to differential selection theory, racial disparities in arrests are a function of racial discrimination by police in selecting whom to investigate and arrest [31]. If, as differential selection theory suggests, police target people differently on the basis of race, then arrest data, particularly for drug and nuisance crimes, is significantly influenced by racial bias in police officers’ choices about whom to investigate. If arrests for drug and nuisance crimes are a reflection of racial bias on the part of police, and arrests are used to generate the forecasts of predictive policing systems, then these systems will tend to forecast greater crime than actually exists in communities where racial minorities are concentrated. Thus, residents of those communities will often face unnecessary, and sometimes unwanted, additional police attention. Second, once police inundate an area, this can lead to even more police contacts, incidents, and arrests in minority communities. This data is fed back into the system and used to make future predictions, leading to a “ratchet effect” of escalating police attention [32]. Danielle Ensign et al have called this cycle of ever-escalating police attention caused by predictive policing a “runaway feedback loop” [33]. Insofar as the predictive success of predictive policing systems hinges on ratcheting up the unjust and discriminatory policing patterns of the past, these systems are not ethically justifiable. Nor are they particularly useful.

 Facial recognition systems face similar concerns. Whereas much of the public concern surrounding such systems concerns privacy, the disparate rates of false positives between genders and races is also a cause for concern. If members of minority groups have higher rates of false positive identification as persons of interest, they will be more likely to receive unwanted police attention, which could yield greater risks of harms during these interactions [34].

Whether the concern about discriminatory treatment applies to particular data-driven systems will depend on a number of factors, including the kind of data the system uses to generate predictions. For example, the LAPD adopted PredPol in part because the system does *not* rely on arrest data to generate crime forecasts. In this way it seems to avoid influence from police officer selection bias. Sean Malinowski of the LAPD describes this rationale in Andrew Guthrie Ferguson’s seminal book on constitutional issues related to data-driven policing.

*...arrests are not part of the equation [in generating PredPols forecasts]. We felt this was important because we heard from some community members that they were concerned about the program creating a kind of self-fulfilling prophecy from under which a community could not recover...In our model, we would hope to deploy the officer based on crime only and then deny the criminal the opportunity to commit the crime in the first place* [8].

Even without arrest data, a ratchet effect can occur, depending on what non-arrest data is used in a predictive model. For example, police reports can come from community members reporting crimes and from officers looking for crimes, so police reports are susceptible to influence from selection bias. Emergency calls for service, on the other hand, are entirely community-driven, so they offer an attractive alternative to arrest data and crime reports. But even data about emergency calls is not immune from influence by human bias. Marda and Narayan have identified several potential sources of bias within the New Delhi place-based predictive policing system CMAPS, which makes predictions solely on the basis of data from past emergency calls for service and police reports [19]. First, historical police records in Delhi may reflect the biases of the reporting officers, who might have been more likely to record crimes perpetrated by individuals from minority communities. Second, calls to Dial 100 may come disproportionately from wealthier and more socially empowered citizens, and wealthier, better educated callers might be better equipped to provide the detailed information that is operationalizable by CMAPS.

Other police departments using data-driven policing systems have seemingly ignored the threat of a ratchet effect altogether. Pasco County’s Intelligence Led Policing identifies “prolific offenders” using a variety of data sources, each of which is susceptible to influence by police bias. By assigning points to individuals based on arrests and mere *suspicion* of criminality, the Pasco County system faces a clear threat from differential selection. By “enhancing” point assignments for mere involvement in crime reports, either as a reporting person, victim, or witness, the system risks placing undue burden on innocent members of the community.

Perhaps the most remarkable fact about the public and academic conversation around data-driven is how little evidence exists either of bias and discriminatory impacts, or of data-driven policing’s efficacy in preventing crime. In the U.S. context only a small handful of studies have been published assessing the efficacy of data-driven programs [35], and only *one* published study has examined the racially discriminatory impact of PredPol on arrest rates [36], finding no statistically significant effect on racial disparities in arrests. Of course, absence of evidence is not evidence of absence, particularly when access to data-driven policing systems is denied to the public or to academic investigators, as it often is [35].

Something else that has been lost in the discussions of bias and discriminatory impacts is how to better include members of minority communities in decision-making about data-driven systems and additional police attention being directed toward their community. No studies have been published about public attitudes toward the use of algorithms to predict where crime will occur or who will commit it. National surveys paint a complicated picture of minority attitudes toward police presence in their communities. Black Americans express much less favorable attitudes toward police than do White Americans [37], and black Americans have a long-documented distrust of the legal system and of police in particular [38–40]. And yet in a recent Gallup poll, 61% of Black Americans wanted police presence in their area to remain the same, and 19% wanted police to spend *more* time in their area [41]. The details of when, how, and what sort of police attention in minority communities is an unwanted *burden* is a complicated matter.

Person-based predictive policing programs like Pasco County’s prolific offender program raise a host of distinct concerns relating to civil liberties and surveillance. What academics and journalists have found most troubling about these programs is that they often target individuals who are not suspected of any particular crime, but, because of their social networks, have become *associated* with criminal activity. For example, an Enforcement Notice by the Information Commissioner, responsible for enforcing data regulation for the Metropolitan Police, found that the London Gangs Matrix used by the Metropolitan Police included both victims and perpetrators of gang violence, and that it failed to distinguish between them [42,43]**.**

Once police have fixed their gaze on a so-called prolific offender, and information about their prolific offender status circulates to other institutions, it can become virtually impossible to get back onto the right side of the law [13]. Data about criminal justice contacts is increasingly integrated and shared across institutions. One consequence of labeling some individuals as chronic or prolific offenders is therefore that people with records of contact with the criminal justice system may shy away from other institutions that keep formal records. Sociologist Sarah Brayne has argued that individuals with police contacts, arrests, or convictions are less likely to participate in medical, financial, or educational institutions, or the labor market than individuals who do not have those recorded contacts with the criminal justice system [44]. Thus, what begins as an inequality in the criminal justice system can entrench inequality in access to other vital social services.

Being marked by an algorithm as a prolific offender has implications for constitutional protections as well. The fourth amendment protects citizens from unreasonable stops, searches, and seizures by law enforcement. Whether a stop counts as reasonable depends in part on whether a police officer had reasonable suspicion that the subject of the search was involved in a criminal activity. As Andrew Ferguson has noted in several places, what constitutes reasonable suspicion for a search might be affected by algorithmic classifications [45]. Behavior that would otherwise be insufficiently suspicious to justify a search may warrant a search if the officer observing the person knows that he or she has been flagged as a “prolific offender” by a predictive policing program. But this additional suspicion has nothing to do with the *current* *activities* of the suspect. The suspicion is grounded in a statistical inference made by a computer algorithm.

Addressing discrimination and violations of civil liberties in data-driven policing requires a multi-pronged approach. As we’ve suggested, care must be taken in data source selection and processing to avoid a ratchet effect. Beyond that, continuous monitoring of the impacts of data-driven systems on communities is required to guard against disparate impacts [46–48]. This has not been done effectively in the past. One of the key deficiencies of the PredPol program identified in the Office of the Inspector General’s audit of LAPD’s data-driven policing programs was inconsistent recording by officers of time spent in the designated high-risk zones [49]. The program had been in effect for approaching ten years by the time the OIG made this discovery.

**Challenge #2 What data for which purposes?**

Our second challenge concerns two assumptions made by proponents of data-driven policing that its critics have called into question.

**Assumption 1** advocates of data-driven policing often tacitly assume that once a department adopts certain data-driven methods, these methods can be used in a value-neutral way. That is, it is taken as obvious how a given technology is to be incorporated into existing practices, what roles or tasks it will assume, and what the parameters of its function will be. But a growing body of academic literature calls this assumption into question [13,46,50].

We just saw in the previous section that data-driven policing methods have the potential to further entrench discriminatory elements of the criminal justice system. But there are a number of choice points along the way that lead to this problem. One key choice point is where and how in a given department’s strategic decision-making process these technologies are introduced.

To begin, take the case of place-based predictive policing. There are several junctures in the decision-making process where this technology can be employed. First, it could be consulted at an early stage, to determine where to distribute the department’s resources at a given time—i.e., which macro-level places (e.g., beats, city wards, districts, or neighborhoods) ought to be prioritized. Alternatively, the system could be consulted after having made judgments about how to allocate police resources among macro-level places. At this point, the predictive software could help focus the officers’ attention on a given block or address within the broader area they have selected.

It may appear that the choice between these two options is governed entirely by strategic considerations; however, the evaluative assumptions embedded in this decision raise important ethical issues. In particular, while consulting the predictive policing system at the outset may seem more efficient, it is also more likely to result in targeting minority communities that have been historically and systemically burdened by excessive police scrutiny. This, after all, is the clearest example of how data-driven policing reproduces the history of racist policing. The second option—only consulting the data at later stages—precludes this: entire communities would not be subject to police scrutiny on the basis of a predictive policing forecast.

A second choice point is the choice of data-driven technology in the first place. Choices about which technological system to adopt suggest different solutions to crime. For example, PredPol uses data only on timing, location, and type 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. Compare this system with one like RTM 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, and it therefore suggests different non-enforcement-oriented solutions. Addressing the underlying features of places that make them vulnerable to crime might require engaging non-law-enforcement agencies like public works, sanitation, or urban planning. Therefore, choices about which data to use in data-driven policing implicitly or explicitly involve choices about the proper role of police in crime prevention. While choices about technology adoption can shape policing practice, they will not satisfy all of the demands of reformers. RTM, for example, is designed to facilitate collaboration between police and city agencies on non-enforcement-oriented solutions to crime, but RTM does not tackle the underlying socio-economic conditions of peoples’ lives that make them vulnerable to crime or criminality. And so RTM does not directly address calls to redirect law enforcement funding to other social services such as education and mental health so as to target the underlying social drivers of crime.

Further, much of the focus of big data policing systems centers on ‘street crime’—e.g., property damage, theft, assault, and so forth. This focus is a choice on the part of police departments and those responsible for designing and programming these systems. What is often referred to as ‘white-collar crime’, as well as other types of harmful and socially disruptive criminal activity, is not often actively pursued by local police departments. There are several possible explanations for this focus, such as the significant data available on street crime, which better enables the discovery of patterns within the data that facilitates future predictions. Or it could be attributed to political pressures, departmental expertise, and so forth. In any case, it is important to recognize that, as a result of this narrower focus, the measurements of success or failure of data-driven systems will be sensitive only to their ability to thwart this particular subset of crimes. This leads to several underappreciated issues, such as the creation of incentives for police departments to increase focus on crimes disproportionately committed by the poor; the increased criminalization of certain minor offenses; and the heightened potential for harassment of those with prior criminal activity, and unwanted attention to those in certain geographic areas [46].

These examples illustrate several aspects of the broader ethical concern relating to implementation of these data-driven methods. More generally, police departments implementing such technologies face the following challenge: What data should we use, and for what purposes should we use it? Much of the discussion surrounding the implementation of data-driven policing systems has thus far ignored this question. As a first step, departments planning to adopt such methods ought to scrutinize some of the values and priorities that adoption assumes. And this scrutiny must not occur in a vacuum. Insofar as data-driven policing systems encourage more aggressive enforcement-oriented policing tactics than other systems, community members have a significant stake in participating in the deliberative process about which systems are adopted. For example, citizen-led police advisory councils must be empowered to communicate to police the needs and concerns of community members. As we discuss in Challenge #4 below, this will require greater transparency about the role that data-driven systems play in policing.

**Assumption 2** The second major assumption taken for granted by advocates of data-driven policing, and which we have been granting so far, is that the primary function of data-driven tools in policing is to promote social well-being. A growing chorus of critics has called this assumption into question by calling into question the social value of police [51,52]. Calls to abolish or defund the police have grown in volume and frequency over the past decade as killings of unarmed Black men and women by police continue to make national headlines. If policing is not a legitimate institution, then the tools it uses are not legitimate either. Police abolitionists have argued that the institution of policing is, and always has been, an unjust mechanism of social control of minority bodies, and that one need only glimpse the history of policing to see that this is true [53]. On this view, data-driven policing tools only serve to maintain that control [54]. Indeed, some of the staunchest opponents of data-driven policing are explicitly abolitionist. Hamid Khan is the director of the Stop LAPD Spying Coalition, the leading anti-data-driven activist group in Los Angeles. In interviews he has described the group as “fiercely abolitionist” [55]. Khan’s opposition to data-driven technologies like PredPol and Operation Laser is founded in a commitment to the idea that the system of which these tools are a part is irredeemably racist.

 Opposition to data-driven policing that is rooted in the police abolitionist movement leaves little room for reform. From this perspective, any tool of the police will, necessarily, be used for oppressive purposes and so must be dismantled. Data-driven technologies are a tool of the police. Therefore they must be dismantled. Whether or not one accepts the premise of the abolitionist movement, the critique raises a vital complication for the question we started with: “which data for which purposes?” If existing data-driven policing technologies are tools of oppression and racial control, then we shouldn’t be using data *at all* in the service of these technologies [56]. Instead what data we use, and how we use it, requires reimagining the fundamental goals guiding technology development. Ruha Benjamin is a leading academic voice of this perspective, arguing that the social aim of technology should be to promote equity and social justice, and not be driven by concerns for maximal efficiency and profit motive [12].

**Challenge #3 The standard of success**

As we noted above, a common refrain among critics of predictive policing is that predictive policing systems will discriminate against people of color. However, human decision-makers are far from perfect when it comes to being influenced by racial bias. This raises a key question: when evaluating a data-driven system’s accuracy, transparency, or fairness, what is the relevant standard of success? More specifically, what constitutes adequate performance by the data-driven system such that a law enforcement agency is justified in implementing it in their decision-making process?

This question faces applications of algorithms and machine learning across all sectors, but it arises in high relief in criminal justice contexts, where the stakes are very high. Is it good enough for data-driven policing systems to make more accurate classifications, offer greater transparency with respect to the factors that play a role in institutional decision-making, and produce criminal justice outcomes that better align with societal commitments to fairness than the status quo in policing? Alternatively, should algorithmic systems be held to a significantly higher standard than the status quo in each of these regards? While the algorithms will never be perfect, neither is the status quo. And waiting too long to adopt the technology might allow inferior policing operations to prevail in the meantime.

 To illustrate the issue at stake, let us consider Chicago’s experiment with a person-based predictive policing system called “Strategic Subject’s List,” also known as “The Heat List.” The Heat List, now defunct, used an algorithm to identify individuals at high risk of perpetrating or being victimized by a crime. The algorithm assigned rankings to individuals on the basis of certain features, including arrests for drugs, illegal possession of a firearm, assaults, and gang membership. Investigations of this program revealed significant racial disparities in risk assignments. Black men were vastly overrepresented among subjects on the Heat List [57]. A similar racial disparity was found during the course of an internal audit of the London Gangs Matrix [24]. Do these disparities alone mean that Chicago Police should have stopped using the Heat List to assess risk of criminality and victimization?

Answering this requires first establishing a standard of success for data-driven tools in criminal justice. One simple standard is that implementing an algorithmic tool is justified if the tool surpasses human decision-making in the same domain of application [58]. If one aim of these tools is to produce less discriminatory outcomes than human decision-makers, this aim might still be achieved even if The Heat List over-represents members of some racial groups.First, scholars specializing in the evaluation of risk assessment tools in criminal courts have argued that tradeoffs between accuracy and fairness apply in this context regardless of whether or not one uses algorithms in the decision-making task [59]. Second, adopting data-driven tools might sometimes be preferable to human decision-makers in the same domain even if the tools do not produce perfectly fair outcomes. For, even without perfect knowledge of a data-driven system’s inner workings, algorithmic systems provide unique opportunities to identify the *sources* of discriminatory impacts and suggest remedies for those sources. A growing field sometimes called “explainable AI” is concerned with making the determinations of algorithmic systems interpretable or understandable to the human subjects affected by them [60,61]. One subfield of explainable AI is concerned with explaining the discriminatory behavior of data-driven systems. As Jon Kleinberg et al put it, when it comes to addressing discriminatory outcomes, “The opacity of the algorithm does not prevent us from scrutinizing its construction or experimenting with its behavior – two activities that are impossible with humans” [62]. Furthermore, they argue, a properly calibrated algorithm can counteract discriminatory intent by human decision-makers in criminal justice by providing a “counterweight” to their judgment [62]. That is, the determinations of a non-discriminatory risk assessment system can be used as a check against discriminatory decision-making by human officials. We note in Challenge #4 below various obstacles to inspecting the inner workings of data-driven systems for discriminatory potential, but these obstacles are not obviously *greater* in the case of data-driven systems than they are for human decision-making. The inner workings of the human brain that give rise to our choices remain largely an impenetrable black box, the sources of human bias are often unknown to the decision-maker or to observers, and bias in humans is often deeply intractable [63]. So, if the standard for success is that the system performs better than a human with respect to fairness, transparency, and accuracy, success might be within reach.

One concern with setting human performance as the standard of success is that it might set the bar too low. For example, in the one academic study measuring PredPol’s predictive accuracy in the field in comparison with human crime analysts, PredPol’s predictions were compared with human crime analysts in Kent, UK and Los Angeles, California. In Los Angeles, PredPol predicted 4.7% of crimes compared with 2.2% predicted by human crime analysts. On one reading of the results, PredPol was a vast improvement over human crime analysts, more than doubling their accuracy. But this feat is only impressive if we think that human crime analysts are already pretty good at predicting where crime is going to occur. Does predicting 2.2% of the locations where crime will occur amount to expertise? And is it worth investing in an algorithmic system that predicts 2.5% more crime than human analysts? As Andrew Ferguson says, describing the study, “To say X is better than Y is only really meaningful if you have a baseline understanding of the value of Y. Maybe both the algorithm and the analyst are terrible, so being better than terrible is not necessarily worth the investment” [8]. And as Ferguson notes, there are no academic studies measuring the accuracy of human crime analysts, so it is at present impossible to assess their adequacy as a baseline of comparison.

Brantingham and Valasik’s study examining discriminatory arrest patterns when using PredPol found no significant differences in the proportion of arrests by racial-ethnic groups when comparing police patrols allocated by PredPol to patrols allocated by human crime analysts [36]. But once again, achieving parity with human decision-makers may set the bar for success too low if arrest patterns are themselves racially biased. If too many Blacks and Latinos are arrested when patrol officers are distributed in accordance with the predictions of human analysts, *maintaining* that same proportion of arrests of Blacks and Latinos is not something to be celebrated. We should instead demand that the algorithmic system *improve* on the status quo. And indeed, the differential selection of arrestees by race and ethnicity is a well-established phenomenon for certain types of crime [13,64,65]. An advocate of police reform will not be satisfied by the fact that data-driven policing systems maintain, but fail to improve on, discriminatory patterns of policing. Rather, they will claim that this maintenance of the status quo is *precisely* why these systems are unacceptable.

**Challenge #4 Explanation and transparency**

Like many other algorithmic systems, data-driven policing systems have been criticized for their relative lack of transparency—that is, their opacity [8,66–68]. A system is opaque in this way when the contribution of any single feature of the world to the final prediction cannot be easily understood, either by the human decision-maker or the person directly affected by the prediction [69].

Data-driven policing systems are opaque in several ways. First, the algorithms at the core of such systems are extraordinarily technically complex. Certainly, most of those subjected to them are incapable of understanding them. But this is also true of most of the officers who employ them. In her research on predictive policing, Sarah Brayne spent time alongside LAPD officers as they used this technology in their work. Most of them, she says, admitted that they do not understand how or why the algorithm works.

Data-driven systems developed using “deep learning” tools only amplify technical obscurity. These systems are programmed to ‘learn’ inference rules over time—i.e, correct from previous mistakes, and thus to make more accurate classifications. As a result, with increased use, the model at the heart of these systems evolves over time, and the rules that these systems learn may not be easily understood by humans [58]. Even the most highly trained professionals would struggle to understand why these systems generate the predictions and assessments they do. In other words, as these systems improve, their opacity likewise increases.

Second, most data-driven systems are shrouded in secrecy. Some of this is due to the lack of transparency about the extent of their implementation; whether and to what extent a human is ‘in the loop’; and whether, if ever, these systems are audited. A deeper concern, however, is that these systems are generally created and maintained by private companies, who retain extensive intellectual property rights over their proprietary source code. As a result, many of those subject to these systems are incapable of scrutinizing them. In *State v. Loomis*, the defendant, Eric Loomis, argued that his due process rights were violated, since he and his legal team were not granted access to the COMPAS algorithm that deemed him high risk, since this information was proprietary [70]. While COMPAS is a tool used in courthouses, the same issues arise with the predictive technologies in use by police departments nationwide.

Whether it originates in the technical features or proprietary features of data-driven technology, opacity in the use of data-driven policing technologies is problematic on several fronts. First, government institutions ought to be capable of publicly justifying their treatment of citizens *to* the citizens affected. But technical and proprietary opacity hinders their ability to do so.

Moreover, both sorts of opacity have the potential to sow distrust and skepticism toward police, which may worsen many of the problems these technologies are intended to resolve. For example, many police departments have adopted data-driven systems in response to concerns about racial bias in policing that occurs when police departments rely entirely on officers’ individual judgment and discretion. In response to a critical exposé of their Intelligence Led Policing Program practices in the Tampa Bay Times, the Sheriff’s Department in Pasco County, Florida wrote that their system:

removes any opportunity for bias by removing descriptive features and focuses strictly on the criminal history of the individual, regardless of their race, gender, creed or any other identifying factors. We are surprised to see a system that is blind to anything but the criminal history of an individual is under attack, instead of being celebrated as an important step forward in our country. [71]

Eliminating discriminatory impacts from policing is of course an admirable goal. However, it is questionable whether making a data-driven system blind to individuals’ protected characteristics is the most effective way to eliminate discriminatory impacts ([59,62]. As we discuss in Challenge #1, race-blind data-driven systems can reflect and reinforce discriminatory practices, if the data used by those systems reflects discriminatory practices. Furthermore,when the public is unable to hold these technologies accountable because of opacity, the public has good reason to suspect that the technologies will perpetuate bias and injustice. This, in turn, can lead to even greater distrust of police, decreased perceptions of legitimacy, and more injustice.

**Challenge #5 Accountability and community oversight**

The democratic legitimacy of law enforcement institutions requires that there are appropriate mechanisms for holding them accountable. Accountability mechanisms foster public trust, enable the achievement of valuable institutional goals, and identify and address any problematic elements within the institution itself [58]. But the value of accountability is also intrinsic: part of what constitutes a legitimate institution—particularly those, like law enforcement, that wield significant authority—is that they are democratically accountable to those they are intended to serve.

 The rapid introduction of, and increasing reliance upon, data-driven methods throughout law enforcement introduces a number of challenges for citizens’ ability to hold accountable the institution of law enforcement and those who play a role in it. One such challenge concerns an issue mentioned in the previous section—namely, opacity. While law enforcement requires some measure of opacity or secrecy for their operations to be at all effective, secrecy threatens citizens’ ability to hold law enforcement accountable.

Most of the data-driven technologies employed by police departments across the country suffer from various sorts of opacity. For one thing, many of the most high-profile technologies have been developed and maintained by private companies, which are legally afforded (and typically retain) the right not to disclose certain proprietary elements of these systems to the departments that use them, or to the general public. Police departments are often given full authority to procure these systems from vendors directly and to implement them without oversight from the public or from city commissions [72,73]. Privatization thus poses a significant obstacle to the efforts of concerned citizens, advocacy groups, and those directly impacted by these systems to hold law enforcement accountable. Without full access to these systems, any attempts at holding law enforcement accountable will be incomplete.

 One potential step toward addressing this issue is to require technology companies to release proprietary information to the relevant parties in certain cases. Companies have thus far largely resisted this idea, and police departments have shown little interest in insisting on it as a condition of contracting with technology firms, but courts have recently indicated that this may change. In a recent case, *New Jersey v. Pickett,* the court ruled that defendants have a right to inspect and understand the software that is used to provide evidence against them, after a defendant had previously been denied access to the proprietary DNA software that served as the basis for his arrest [73]. Perhaps other courts will follow suit for algorithmic technologies used in other facets of the criminal justice system, including policing.

While this case suggests a form of legal recourse for defendants subjected to algorithmic decision-making, it faces legally and morally unsettled questions about how to weigh individuals' rights to explanation for algorithmic decision-making against the intellectual property rights of private firms. If the source code of algorithmic systems are protected as trade secrets, even in criminal proceedings, then there are legal limits to the inspection of those systems by defendants. The extent to which algorithmic systems should be protected as trade secrets remains in dispute [75]. Settling that dispute is beyond the scope of this paper. Even once these questions are settled, the right to inspect and understand algorithmic systems is highly limited. The right to inspection was extended only to those accused of crimes but not other interested parties, such as advocacy groups and other oversight organizations, who might play a role in helping to prevent abuses of these systems before they begin. Furthermore, courts are in general a weak mechanism for oversight of law enforcement, since the courts are constitutionally limited in the ways they can provide a check on law enforcement practices. As Erik Bakke argues, “The Fourth Amendment’s protections apply only after there is an invasion of a reasonable expectation of privacy. Because predictive policing is used for surveillance, which does not require reasonable suspicion, police can observe citizens without any court interference” [68]. Thus, full accountability is unlikely to be achieved through the courts alone.

 It is widely thought that accountability can be improved by subjecting data-driven systems to ‘algorithmic audits’ and ‘algorithmic impact assessments’, in which an independent organization evaluates the system for bias, unfair treatment, improper data, corrupt source codes, harmful impacts and other problems [76,77]. Legislative momentum is building to require that private firms regularly evaluate and remedy any flaws or biases in their algorithmic systems. In April 2019, U.S. Senators Cory Booker and Ron Wyden, along with Rep. Yvette D. Clarke introduced The Algorithmic Accountability Act, which would authorize the Federal Trade Commission to require companies to conduct regular impact assessments of certain “automated decision systems” [78]. Such a requirement should extend to those private algorithms being leased by U.S. law enforcement agencies and courts. If algorithmic auditing only looks at the system’s inputs and outputs, without inspecting the system’s source code, it might circumvent some of the intellectual property concerns raised above [79].

Joy Buolamwini’s work auditing commercial facial analysis models used by IBM, Microsoft, and Megvii (Face++) provides an encouraging example of how audits can address biases embedded in AI systems. Within seven months of an initial audit demonstrating disparities in accuracy of facial analysis between men and women, and between light and dark skinned people, IBM, Microsoft, and Megvii released new versions of their software that significantly reduced those disparities [80].

Another encouraging example of algorithmic auditing in practice comes from a collaboration between ShotSpotter Technologies and NYU’s Policing Project. ShotSpotter Inc. is a private technology firm whose product uses computer sensors to identify and analyze gunshots and then notify law enforcement. The Policing Project, run out of NYU’s School of Law, “partners with communities and police to promote public safety through transparency, equity, and democratic engagement” [81]. ShotSpotter voluntarily approached the Policing Project to conduct an audit of potential privacy threats to bystanders whose voices might be picked up by the ShotSpotter sensors. While the audit revealed little threat of unintended voice surveillance, the Policing Project’s final report on the audit happily describes that ShotSpotter “adopted nearly all of our [privacy protection] recommendations verbatim” [81].

These stories are promising, and algorithm audits are, to be sure, an essential step in ensuring law enforcement accountability; however, several key issues remain. First, the auditing process itself requires significant transparency—that is, measures must be put in place to ‘audit the auditors’, or to otherwise ensure the independence of the auditing process. Without this, there will remain incentives for departments to employ only those auditors that are most favorable to them, and for auditors to provide only the most favorable results.

Moreover, even if such audits are fully independent, measures must be put in place to ensure the full unredacted reports are released without spin. Recently, a company called HireVue, which uses AI during job interviews to evaluate job candidates, hired an outside consulting firm, headed by AI expert Cathy O’Neil, to perform an audit of their software [82]. As requested, O’Neil’s team focused on a specific piece of HireVue’s technology; but the company used the largely positive audit to suggest that their entire system had been vindicated—a conclusion that extended far beyond the evidence from the audit. While HireVue is a public company, similar concerns about transparency in the audit process itself affect public institutions like law enforcement. To be sure, algorithm audits are an important step, but they can also be used as a form of “accountability theater” that obscures more than it illuminates.

A further way to promote accountability is through legislative action. City commissions that oversee police departments can create citizen-led councils to act as a conduit through which the community can offer recommendations and advice to police about the impact of police work on their communities. These councils should include as part of their charge the evaluation of new technological policing systems, particularly their impact on minorities. These bodies, sometimes called “citizen advisory councils,” are useful tools of communication and trust building between police and communities, but they provide limited accountability insofar as the council serves a merely advisory role; city officials or police departments can opt to ignore the recommendations. A more demanding approach would require *prior* legislative authorization by city officials of any new data-driven technologies by police [83,84].

**Conclusion**

Our discussion in this essay is far from exhaustive. Each of the ethical challenges we discussed raises many more complex and nuanced issues than we have addressed here. Furthermore, there are many other urgent technical, legal, and ethical challenges facing data-driven policing.

For certain of these challenges, there are clear pathways forward. For instance, as we have noted, several of the problems, such as those concerning opacity and oversight, can be partially addressed via legislation or departmental policies that require that any companies contracted to provide algorithmic technologies to police departments waive the relevant intellectual property rights in certain contexts and to submit to regular external independent audits. None of these steps is a panacea, and they each raise further issues that must be addressed as well.

Other challenges, such as those relating to bias, data selection, standards of success, and deeper questions about the social value of police require more collective deliberation. This deliberation must be informed by academics, technologists, law enforcement, and, most importantly, the lived experience of the public. Some scholars have recently considered whether addressing algorithmic bias directly in the design and programming might mitigate or eliminate these negative effects, and the initial results appear promising [59]. There are, however, concerns with this approach; and some will find the possibility of any amount of encoded bias unacceptable.

While police departments nationwide vie for better data-driven technologies to support their mission to protect and serve their communities, advocacy groups and the broader public are calling for the discontinuation of data-driven policing programs. As we have attempted to show here, what ought to be done about data-driven policing is a complex and multi-faceted issue. It will be virtually impossible to satisfy all sides of the debate. Those who wish to continue on the path to better data-driven policing must work to address the challenges we have raised above, and many other issues that are raised by this technological shift. And those who oppose data-driven policing practices ought to be aware of and sensitive to the full range of challenges, their complexity, and the various possible pathways forward. Socially responsible data-driven policing requires collaboration between academics, technology developers, police departments, and policy-makers to confront and address these challenges. Most importantly, it requires engagement with and oversight by affected communities, especially those who have become deeply skeptical about the enterprise of policing.

**Conflict of interest statement:** On behalf of all authors, the corresponding author states that there is no conflict of interest.

1 Heaven WD. Predictive policing algorithms are racist. They need to be dismantled. MIT Technol. Rev. 2020.https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/ (accessed 15 Feb 2021).

2 Haskins C. Academics Confirm Major Predictive Policing Algorithm is Fundamentally Flawed. *Vice.com* 2019.https://www.vice.com/en/article/xwbag4/academics-confirm-major-predictive-policing-algorithm-is-fundamentally-flawed

3 Moravec ER. Do Algorithms Have a Place in Policing? *The Atlantic* 2019.https://www.theatlantic.com/politics/archive/2019/09/do-algorithms-have-place-policing/596851/ (accessed 15 Feb 2021).

4 Sturgill K. Santa Cruz becomes the first U.S. city to ban predictive policing. Los Angel. Times. 2020.https://www.latimes.com/california/story/2020-06-26/santa-cruz-becomes-first-u-s-city-to-ban-predictive-policing (accessed 15 Feb 2021).

5 Burbank K. City First In Nation To Ban Predictive Policing, Biometric Surveillance Tech. Sfgate.com. 2021.https://www.sfgate.com/news/bayarea/article/City-First-In-Nation-To-Ban-Predictive-Policing-15872642.php

6 Geolitica.com. www.geolitica.com (accessed 25 May 2021).

7 Los Angeles Police Department. The Los Angeles Police Department Strategic Plan 2019-2021. 2019.

8 Ferguson AG. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: : NYU Press 2017.

9 Dou E. China built the world’s largest facial recognition system. Now, it’s getting camera-shy. Wash. Post. https://www.washingtonpost.com/world/facial-recognition-china-tech-data/2021/07/30/404c2e96-f049-11eb-81b2-9b7061a582d8\_story.html (accessed 9 Sep 2021).

10 Woollacott E. UK Government Accused Of Sneaking Through New Live Facial Recognition Rules. Forbes. https://www.forbes.com/sites/emmawoollacott/2021/08/23/uk-government-accused-of-sneaking-through-new-live-facial-recognition-rules/ (accessed 9 Sep 2021).

11 Bostrom N. The Vulnerable World Hypothesis. *Glob Policy* 2019;**10**:455–76. doi:10.1111/1758-5899.12718

12 Benjamin R. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity 2019.

13 Brayne S. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York: : Oxford University Press 2020.

14 Lazer D, Radford J. Data Ex Machina: Introduction to Big Data. *Annu Rev Sociol* 2017;**43**:19–39.

15 Krafft T, Zweig KA, Konig PD. How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regul Gov* 2020.

16 Barocas S, Rosenblat A, Boyd D, *et al.* Data & Civil Rights: Technology Primer. Data Soc. 2014.https://datasociety.net/library/data-civil-rights-technology-primer/ (accessed 12 May 2021).

17 Short MB, D’Orsogna MR, Brantingham PJ, *et al.* Measuring and Modeling Repeat and Near-Repeat Burglary Effects. *J Quant Criminol* 2009;**25**:325–39.

18 Miller L. LAPD will end controversial program that aimed to predict where crimes would occur. Los Angel. Times. 2020.https://www.latimes.com/california/story/2020-04-21/lapd-ends-predictive-policing-program (accessed 15 Feb 2021).

19 Marda V, Narayan S. Data in New Delhi’s Predictive Policing System. *Proc ACM Conf Fairness Account Transpar* 2020.

20 Cohen LE, Kluegel J, Land K. Social Inequality and Predatory Criminal Victimization: An Exposition and Test of Formal Theory. *Am Sociol Rev* 1981;**46**:505–24.

21 Caplan J, Kennedy L, Miller J. Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting. *Justice Q* 2011;**28**:360–81.

22 Uchida CD, Swatt M, Gamero D, *et al.* The Los Angeles Smart Policing Initiative: Reducing Gun-Related Violence through Operation LASER. U.S. Department of Justice, Bureau of Justice Assistance 2012. http://newweb.jssinc.org/wp-content/uploads/2014/11/Spotlight-on-Operation-LASER.pdf

23 McGrory K, Bedi N. Targeted. Tampa Bay Times. 2020.https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/intelligence-led-policing/ (accessed 15 Feb 2021).

24 Review of the MPS Gangs Matrix. Mayor’s Office of Policing and Crime 2018. https://www.london.gov.uk/mopac-publications-0/review-mps-gangs-matrix (accessed 25 May 2021).

25 Review of the MPS Gangs Violence Matrix - Update. Mayor’s Office of Policing and Crime 2021. https://www.london.gov.uk/mopac-publications-0/review-mps-gangs-violence-matrix-update (accessed 25 May 2021).

26 Ferguson AG. Illuminating Black Data Policing. *Ohio State J Crim Law* 2018;**15**:503–25.

27 Fussey P, Murray D. Independent Report on the London Metropolitan Police Service’s Trial of Live Facial Recognition Technology. University of Essex: : The Human Rights, Big Data, and Technology Project 2019.

28 Kostka G, Steinacker L, Meckel M. Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Underst Sci* 2021;**forthcoming**.

29 Chen S. Is China’s corruption-busting AI system “Zero Trust” being turned off for being too efficient? South China Morning Post. 2019.https://www.scmp.com/news/china/science/article/2184857/chinas-corruption-busting-ai-system-zero-trust-being-turned-being (accessed 25 May 2021).

30 Brey P. Ethical aspects of facial recognition systems in public places. *J Inf Commun Ethics Soc* 2004;**2**:97–109. doi:10.1108/14779960480000246

31 Piquero A, Brame B. Assessing the Race-Crime and Ethnicity-Crime Relationship in a Sample of Serious Adolescent Delinquents. *Crime Delinquency* 2008;**54**:390–422.

32 Harcourt BE. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. Chicago: : University of Chicago Press 2006.

33 Ensign D, Friedler SA, Neville S, *et al.* Runaway Feedback Loops in Predictive Policing. In: *Conference on Fairness, Accountability and Transparency*. PMLR 2018. 160–71.http://proceedings.mlr.press/v81/ensign18a.html (accessed 16 Apr 2021).

34 Buolamwini J, Gebru T. Gender Shades:  Intersectional Accuracy Disparities in Commercial Gender Classification. *Proc Mach Learn Res* 2018;**81**:1–15.

35 Boba Santos R. Predictive policing: Where’s the evidence? In: Weisburd D, Braga AA, eds. *Police Innovation: Contrasting Perspectives*. Cambridge: : Cambridge University Press 2020.

36 Brantingham J, Valasik M. Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial. *Stat Public Policy* 2018;**5**:1–6.

37 Ekins EE. Policing in America: Understanding Public Attitudes Toward the Police. Results from a National Survey. 2016.https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2919449 (accessed 15 Feb 2021).

38 Bobo LD, Thompson V. Unfair By Design: The War on Drugs, Race, and the Legitimacy of the Criminal Justice System. *Soc Res* 2006;**73**:445–72.

39 Brooks RRW. Fear and Fairness in the City: Criminal Enforcement and Perceptions of Fairness in Minority Communities. *Calif Law Rev* 2000;**73**:1228–9.

40 Hagan J, Albonetti C. Race, Class, and the Perception of Criminal Injustice in America. *Am J Sociol* 1982;**88**:329–55.

41 Saad L. Black Americans Want Police to Retain Local Presence. Gallup. 2020.https://news.gallup.com/poll/316571/black-americans-police-retain-local-presence.aspx (accessed 15 Feb 2021).

42 Supervisory Powers of the Information Commissioner, Advisory Notice. https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/about-us/gangs-violence-matrix/ico-enforcement-notice.pdf (accessed 24 May 2021).

43 Densley JA, Pyrooz DC. The Matrix in context: Taking stock of police gang databases in London and beyond. *Youth Justice* 2020;**20**:11–30.

44 Brayne S. Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment. *Am Sociol Rev* 2014;**79**:367–91.

45 Ferguson AG. Big Data and Predictive Reasonable Suspicion. *Univ Pa Law Rev* 2015;**163**:327–410.

46 Barocas S, Selbst AD. Big Data’s Disparate Impact. *Calif Law Rev* 2016;**104**:671–732.

47 Corbett-Davies S, Goel S. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. 2018.https://arxiv.org/abs/1808.00023 (accessed 15 Feb 2021).

48 Krafft PM, Young M, Katell M, *et al.* An Action-Oriented AI Policy Toolkit for Technology Audits by Community Advocates and Activists. *Proc 2021 ACM Conf Fairness Account Transpar* 2021;:772–81.

49 Los Angeles Police Commission. Review of Selected Los Angeles Police Department Data-Driven Policing Strategies. Office of the Inspector General 2019.

50 Selbst A. Disparate Impact in Big Data Policing. *Ga Law Rev* 2017;**52**:109–95.

51 Adams M, Rameau M. Black Community Control Over Police. *Wis Law Rev* 2016;**515**.

52 Vitale AS. *The End of Policing*. Verso Books 2017.

53 Muhammad KG. *The Condemnation of Blackness: Race, crime, and the Making of Modern Urban America*. Harvard University Press 2019.

54 Capers IB. Race, Policing, and Technology. *N C Law Rev* 2017;**95**.

55 Ryan-Mosley T, Strong J. The activist dismantling racist police algorithms. MIT Technol. Rev. 2020.https://www.technologyreview.com/2020/06/05/1002709/the-activist-dismantling-racist-police-algorithms/ (accessed 25 Feb 2021).

56 Noble SU. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press 2018.

57 Dumke M, Main F. A look inside the watch list Chicago police fought to keep secret. Chic. Sun-Times. 2017.https://chicago.suntimes.com/2017/5/18/18386116/a-look-inside-the-watch-list-chicago-police-fought-to-keep-secret (accessed 25 Feb 2021).

58 Vestby A, Vestby J. Machine Learning and the Police: Asking the Right Questions. *Polic J Policy Pract* 2019.

59 Skeem J, Lowenkamp C. Using algorithms to address trade‐offs inherent in predicting recidivism. *Behav Sci Law* 2020;**38**:259–78.

60 Edwards L, Veale M. Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For. *Duke Law Technol Rev* 2017;**16**:18–84.

61 Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. *Sci Robot* 2017;**2**. doi:10.1126/scirobotics.aan6080

62 Kleinberg J, Ludwig J, Mullainathan S, *et al.* Discrimination in the Age of Algorithms. *J Leg Anal* 2018;**10**:113–74.

63 Zerilli J, Knott A, Maclaurin J, *et al.* Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philos Technol* 2019;**32**:661–83. doi:10.1007/s13347-018-0330-6

64 Beck A, Blumstein A. Racial Disproportionality in U.S. State Prisons: Accounting for the Effects of Racial and Ethnic Differences in Criminal Involvement, Arrests, Sentencing, and Time Served. *J Quant Criminol* 2018;**34**:853–83.

65 Blumstein A. Bringing Down the U.S. Prison Population. *Prison J* 2011;**91**.

66 Goodman B, Flaxman S. European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation.” *AI Mag* 2017;**38**:50–7.

67 Ferguson AG. Policing Predictive Policing. *Wash Univ Law Rev* 2017;**94**:1109–89.

68 Bakke E. Predictive Policing: The Argument for Public Transparency. *Annu Surv Am Law* 2018;**74**:131–72.

69 Burrell J. How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data Soc* 2016;**3**.

70 State v. Loomis. 2016.

71 Tampa Bay Times. Read the Pasco Sheriff’s Office response to our investigation. Tampa Bay Times. 2020.https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/sheriffs-response/ (accessed 15 Feb 2021).

72 Crump C. Surveillance Policymaking by Procurement. *Wash Law Rev* 2016;**91**.

73 Joh E. The Undue Influence of Surveillance Technology Companies on Policing. *NU Law Rev* 2017;**92**.

74 Fasciale PJAD. State of New Jersey v. Corey Pickett. 2021. https://www.eff.org/files/2021/02/03/decision\_appdiv\_232021.pdf

75 Wexler R. Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System. *Stanford Law Rev* 2018;**70**:1343–430.

76 Barocas S, Selbst AD. Big Data’s Disparate Impact. *Calif Law Rev* 2016;**104**:671–732.

77 Reisman D, Schultz J, Crawford K, *et al.* Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability. AI Now Institute https://ainowinstitute.org/aiareport2018.pdf (accessed 24 May 2021).

78 Booker C. Booker, Wyden, Clarke Introduce Bill Requiring Companies To Target Bias In Corporate Algorithms. 2019. https://www.booker.senate.gov/news/press/booker-wyden-clarke-introduce-bill-requiring-companies-to-target-bias-in-corporate-algorithms (accessed 15 Feb 2021).

79 Wachter S, Mittelstadt B, Russell C. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harv J Law Technol Harv JOLT* 2017;**31**:841.

80 Raji ID, Buolamwini J. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *AIES 19 Proc 2019 AAAIACM Conf AI Ethics Soc* 2019;:429–35.

81 Policing Project. 2020.www.policingproject.org (accessed 26 Feb 2021).

82 Engler AC. Independent auditors are struggling to hold AI companies accountable. Fast Co. 2021.https://www.fastcompany.com/90597594/ai-algorithm-auditing-hirevue (accessed 15 Feb 2021).

83 ACLU. Community Control Police Surveill. TakeCTRL. 2018.https://communityresourcehub.org/resources/community-control-over-police-surveillance-takectrl/ (accessed 26 Feb 2021).

84 Ferguson AG. It’s time for D.C. to regulate police surveillance technology. Wash. Post. 2020.https://www.washingtonpost.com/opinions/local-opinions/its-time-for-dc-to-regulate-police-surveillance-technology/2020/06/25/9e94feb6-b57a-11ea-aca5-ebb63d27e1ff\_story.html (accessed 26 Feb 2021).