CHALLENGING RACIST PREDICTIVE POLICING ALGORITHMS UNDER THE EQUAL PROTECTION CLAUSE

Renata M. O'Donnell*

Algorithms are capable of racism, just as humans are capable of racism. This is particularly true of an algorithm used in the context of the racially biased criminal justice system. Predictive policing algorithms are trained on data that is heavily infected with racism because that data is generated by human beings. Predictive policing algorithms are coded to delineate patterns in massive data sets and subsequently dictate who or where to police. Because of the realities of America's criminal justice system, a salient pattern emerges from the racially skewed data: Race is associated with criminality in the United States. Because of the "black-box" nature of machine learning, a police officer could naively presume that an algorithm's results are neutral, when they are, in fact, infected with racial bias. In this way, a machine learning algorithm is capable of perpetuating racist policing in the United States. An algorithm can exacerbate racist policing because of positive feedback loops, wherein the algorithm learns that it was "correct" in associating race and criminality and will rely more heavily on this association in its subsequent iterations.

This Note is the first piece to argue that machine learning-based predictive policing algorithms are a facial, race-based violation of the Equal Protection Clause. There will be major hurdles for litigants seeking to bring an equal protection challenge to these algorithms, including attributing algorithmic decisions to a state actor and overcoming the proprietary protections surrounding these algorithms. However, if the courts determine that these hurdles eclipse the merits of an equal protection claim, the courts will render all algorithmic decisionmaking immune to equal protection review. Such immunization would be a dangerous result, given that the government is hurling a growing number of decisions into black-box algorithms.

 J.D. Candidate, 2019, New York University School of Law; B.A., 2016, University of Pennsylvania. I am grateful to Professor Barry Friedman, Professor Jessica Bulman-Pozen, and Professor Maria Ponomarenko for taking the time to answer my many questions, provide feedback, and read drafts of this Note. I would like to thank the Editorial Board of the *New York University Law Review*, particularly my editors Nicholas Baer, Alice Hong, and Benjamin Perotin, for the hard work they put in to edit this piece. Special thanks to Professor Kimberly Taylor-Thompson, Professor Tony Thompson, and Professor Randy Hertz for inspiring and cultivating my interest in criminal justice reform. Finally, thank you to the three greatest lawyers I know—my parents, Catherine and Neil T. O'Donnell, and my brother, Neil P. O'Donnell—for thoughtful criticism on this Note and for all of your love, support, and sacrifice.

Introduction

The Chicago Police commander and two officers knock on the home of a Black man on the West Side of Chicago. Robert McDaniel is not under arrest. He has not committed any crime, with the exception of a single misdemeanor he pled to years ago.[1] The officers tell Mr. McDaniel that they have a file on him back at the precinct that indicates he is very likely to commit a violent offense in the near future.[2] Dumbfounded, Mr. McDaniel wonders how they can predict such a thing. The answer: an algorithm known as the Strategic Subject List ("SSL").[3] Mr. McDaniel is shocked because he has not done any-

[1] *See* Jeremy Gorner, *Chicago Police Use 'Heat List' as Strategy to Prevent Violence*, Chi. Trib. (Aug. 21, 2013), http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list (describing Mr. McDaniel's interaction with the police).

[2] *See id.*

[3] *See Going Inside the Chicago Police Department's 'Strategic Subject List,'* CBS Chi. (May 31, 2016, 7:58 AM) [hereinafter *Going Inside*], http://chicago.cbslocal.com/2016/05/

thing egregious that would flag him, personally, as a risk.[4] So why is Mr. McDaniel on the SSL and being monitored closely by police? The Commander tells him that it could be because of the death of his best friend a year ago due to gun violence.[5] Ultimately, it was an algorithm, not a human police officer, which generated the output causing Mr. McDaniel's name to appear on the SSL.

Officers are beginning to delegate decisions about policing to the minds of machines.[6] Programmers endow predictive policing algorithms with machine learning[7]—a type of artificial intelligence which allows the algorithms to pinpoint factors that will distinguish people or places that are allegedly more likely to perpetrate or experience future crime.[8] With each use, algorithms automatically adapt to incorporate newly perceived patterns into their source codes via machine learning and become better at discerning patterns that exist in the additional swaths of data to which they are exposed.[9] In this way, machine learning creates a "black-box" conundrum, wherein the algorithm learns and incorporates new patterns into its code with each decision it makes, such that the humans relying on the algorithm do

31/going-inside-the-chicago-police-departments-strategic-subject-list (describing the Chicago Police Department's Strategic Suspect List); *see also* Gorner, *supra* note 1 (describing the length of the List and the factors used by the algorithm to rank individuals); City of Chi., *Strategic Subject List*, CHI. DATA PORTAL, https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np (last updated Dec. 7, 2017) (providing the List and relevant data).

4 *See* City of Chi., *supra* note 3 (indicating Mr. McDaniel told the *Tribune*, "I haven't done nothing that the next kid growing up hadn't done. Smoke weed. Shoot dice. Like seriously?").

5 *See id.* ("McDaniel, for instance, likely made the list in spite of his limited criminal background . . . because a childhood friend with whom he had once been arrested on a marijuana charge was fatally shot last year in Austin.").

6 *See generally* ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT (2017) (examining the rise of predictive policing in cities and towns across the United States).

7 *See id.* at 3 (discussing machine learning as an important element in the predictive policing algorithms currently on the market).

8 *See* Cynthia Rudin, *Predictive Policing: Using Machine Learning to Detect Patterns of Crime*, WIRED, https://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime (last visited Jan. 25, 2019) ("Machine learning can be a tremendous tool for crime pattern detection, and for predictive policing in general. If crime patterns are automatically identified, then the police can immediately try to stop them. Without such tools, it could take weeks or years . . . or it might be missed altogether.").

9 *See* SOLON BAROCAS ET AL., DATA & CIVIL RIGHTS: TECHNOLOGY PRIMER 4 (2014), http://www.datacivilrights.org/pubs/2014-1030/Technology.pdf (defining the term machine learning as "a kind of learning by example, one in which an algorithm is exposed to a large set of examples from which it has been instructed to draw general lessons").

not know what criteria the algorithm might have relied on in generating a certain decision.[10]

Machine learning-based predictive policing algorithms can learn to discriminate facially on the basis of race because they are exposed to and learn from data derived from the racist realities of the United States criminal justice system—a world in which Black Americans are incarcerated in state prisons at a rate that is 5.1 times the imprisonment of whites,[11] and one of every three Black men born today can expect to go to prison in his lifetime if current trends continue.[12] Machine learning-based policing algorithms learn to replicate and exacerbate these patterns by associating race and criminality. Because these algorithms have the power to discriminate facially by engaging in race-based classifications, they can be challenged under the Equal Protection Clause. This Note is the first piece to argue that machine learning-based predictive policing algorithms present a viable equal protection claim.

Litigants can challenge a state actor's policy under the Equal Protection Clause when that policy impacts a "suspect classification"—such as a classification on the basis of race—because of the policy's intentional, facial discrimination on the basis of the suspect classification.[13] If the litigant can demonstrate that the policy facially discriminates based on the suspect classification,[14] the court reviews

---

[10] *See* Lee Rainie & Janna Anderson, Pew Research Ctr., Code-Dependent: Pros and Cons of the Algorithm Age 19 (2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age [https://perma.cc/89Y6-8L3V] (describing the lack of algorithmic transparency for programmers). It is important to note, however, that it is not clear precisely how these algorithms function. *See infra* note 21 and accompanying text. Some predictive policing algorithms may not be as advanced as this piece theorizes. Coders can parse less sophisticated algorithms that do not present the black-box condundrum. *See* Walter L. Perry et al., Rand Corp., Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations 36 (2013), https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf (discussing the varying complexity in algorithms and differentiating algorithms that are simple and are "directly interpretable by a person" from those that rely on "all of the possible variables and combine them using extremely complicated relationships to generate forecasts," which are "commonly referred to as *black box* models").

[11] The Sentencing Project, The Color of Justice: Racial and Ethnic Disparity in State Prisons 3 (2016), https://www.sentencingproject.org/wp-content/uploads/2016/06/The-Color-of-Justice-Racial-and-Ethnic-Disparity-in-State-Prisons.pdf.

[12] The Sentencing Project, Report of the Sentencing Project to the United Nations Human Rights Committee 1 (2013), https://sentencingproject.org/wp-content/uploads/2015/12/Race-and-Justice-Shadow-Report-ICCPR.pdf.

[13] *See, e.g.*, Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 299 (1978) (laying out the specific framework in which an equal protection claim operates).

[14] *See* Korematsu v. United States, 323 U.S. 214, 216 (1944) ("[A]ll legal restrictions which curtail the civil rights of a single racial group are immediately suspect . . . [and] courts must subject them to the most rigid scrutiny.").

the policy under strict scrutiny and will only deem it constitutional if the government can demonstrate that the policy is narrowly tailored to serve a compelling government interest.[15]

This Note specifically examines machine learning-based predictive policing algorithms that programmers feed and train on data sets from which race is not completely removed. Parsing the differences between specific algorithms, however, is beyond the scope of this Note. This Note does not claim that any predictive policing algorithms are intentionally programmed by developers to target people or places on the basis of race. On the contrary, programmers expose the algorithms to large swaths of data with the benign intention of creating an algorithm that can objectively predict crime.[16] However, when input data—like historical crime data and dragnet data searches—contains information about race, a machine learning algorithm becomes biased by parsing the patterns that exist between race and criminality, regardless of whether the developer explicitly wrote that its source code ought to find such a pattern.

Part I gives an overview of the state of predictive policing. Section I.A defines machine learning. Section I.B explains how machine learning is used in predictive policing. Section I.C explains how and why these algorithms can develop racial biases by delving into the types of data that the algorithms train and rely on, the ways that this data can lead to bias, and the ways in which that bias exacerbates the human biases that already exist in policing. Part II argues that because these algorithms facially discriminate on the basis of race, a group of plaintiffs could bring a viable facial challenge to police precincts' reliance on them. Section II.A gives an overview of the modern equal protection framework. Section II.B applies this framework specifically to machine learning-based predictive policing algorithms. Section II.C discusses the hurdles that litigants will face in bringing an equal protection challenge to machine learning-based predictive policing algorithms.

I
Machine Learning-Based Predictive Policing Algorithms and Their Problems

Artificial intelligence, including machine learning, is remarkably complex both in practice and because of its implications for human

15 *Bakke*, 438 U.S. at 299.

16 *See* Rudin, *supra* note 8 (explaining the motivation for exposing the machine learning-based algorithms to data as a desire to give officers the ability to identify crime before it happens).

lives. Section I.A provides a definition of machine learning. Section I.B unpacks how machine learning works in predictive policing algorithms. Section I.C discusses how machine learning-based predictive policing algorithms can become racist.

A. What Are Machine Learning Algorithms?

Machine learning-based predictive policing begins with an algorithm. An algorithm is a specific sequence of logical operations that provides instructions for how computers should act on an input data set.[17] This specific sequence of logical operations is the source code.[18] A machine learning algorithm learns from the training data that it is fed and finds correlational patterns within that data.[19] The machine learning algorithm subsequently incorporates knowledge of these patterns into its code.[20]

A variety of different algorithms can fall under the umbrella of machine learning algorithms. The precise type of machine learning that predictive policing algorithms rely on is unknown, because their inner workings are considered proprietary knowledge.[21] Based on the goals of predictive policing, however—to identify areas predisposed to future criminal activity and the individuals most likely affected by it—two machine learning candidates are conducive to these goals: K-Nearest Neighbors (KNN) and deep learning neural networks.[22] KNN algorithms incorporate new variables based on the "nearest neighbor" of the original variables the coder programmed the algorithm to use.[23] The KNN algorithm autonomously learns what the variables most

---

[17] *See* BAROCAS ET AL., *supra* note 9, at 3 (defining the term algorithm).

[18] *See id.* (explaining how algorithms are concretely expressed in computer code).

[19] *See id.* at 4 ("[Machine] '[l]earning' occurs when the algorithm extracts logical rules that are not simply a recapitulation of the specific properties of the examples," because they are, instead, replications of those examples and the patterns delineated from them at a broader level).

[20] *See id.* (giving examples of machine learning identifying patterns).

[21] *See* Andrew Guthrie Ferguson, *Policing Predictive Policing*, 5 WASH. U. L. REV. 1109, 1153 (2017) ("Predictive policing relies on proprietary algorithms that adopt a particular analytical methodology.").

[22] *See* Orlando Torres, *Data Science Gone Wrong: Predictive Policing Is a WMD*, TOWARDS DATA SCI. (Jan. 9, 2018), https://towardsdatascience.com/the-perils-of-predictive-policing-11928a9f1d60 (creating a ninety percent accurate predictive policing model using five simple machine learning techniques, including KNN and a deep learning neural network). *See generally* IAN H. WITTEN, EIBE FRANK & MARK A. HALL, DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES 1–188 (3d ed. 2011) (providing a comprehensive overview of various data-mining techniques); WALTER L. PERRY ET AL., *supra* note 10, at 33–36 (explaining various data-mining models and describing ensemble methods which combine simple predictive models to yield a general prediction).

[23] Torres, *supra* note 22.

similar to the original variables (the nearest neighbors of the original variables) are by exposure to the training data. The algorithm then incorporates the nearest neighbors of the variables into its code and relies on the new variables in its subsequent decisionmaking. KNN algorithms do not reveal to the human programmer what additional variables they have come to rely on.[24] All the human programmer knows is the success rate of the algorithm and whether or not it is making successful predictions based on the new variables it is using.

Alternatively, a predictive policing algorithm could rely on deep learning through an artificial neural network.[25] Such a system would "learn" to perform tasks by considering examples without being programmed with any task-specific rules or inherent limits.[26] The neural network receives inputs in the form of training data. At first, the algorithm does not know what it is examining in the training data—it produces a random output.[27] In the predictive policing context, that would mean that during its training phase, the algorithm randomly dictates that there would be a crime that would take place in this neighborhood or would be perpetrated by this person. Each time the algorithm gets an answer "wrong" in its training, its neural connections to variables that produced that answer get weaker.[28] When the algorithm produces the "right" answer, its neural connections to variables that produced that answer strengthen "until the computer teaches itself the features that define" the problem to which it is being applied.[29]

Once it is trained, the machine learning algorithm (whether it is a KNN model or an artificial neural network model) is applied to a current set of data, and it recognizes and applies the patterns that it learned through its training.[30] As it operates on the current data, the algorithm may learn to recognize additional patterns and will incorporate these newly learned patterns into its code as well.[31] In this way, iterative machine learning algorithms create a black-box conun-

---

24 See Figure 1 for a depiction of a KNN input and output code.

25 *See Neural Network*, DEEPAI, https://deepai.org/machine-learning-glossary-and-terms/neural-network (last visited Jan. 25, 2019).

26 *Id.*

27 NOVA PBS Official, *AI Explained: What Is a Neural Net?*, YOUTUBE (May 21, 2018), https://www.youtube.com/watch?v=XS2G0oolHpo.

28 *Id.*

29 *Id.*

30 *See Panel: Bias Traps in AI, 2017 Symposium*, AI NOW INST. (July 10, 2017) [hereinafter *Panel*], https://ainowinstitute.org/symposia/videos/bias-traps-in-ai.html (explaining the recognition patterns in deep learning artificial intelligence).

31 *See id.* (same).

drum.[32] The algorithm is constantly updating itself once it is exposed to a data set. The original developer cannot control this continuous editing process.[33] On the contrary, the algorithm is designed to learn from its previous use, edit itself based on that new knowledge, and use factors that it was not necessarily programmed to rely on, but that it recognizes as patterns in the data sets to which it was exposed.

One might think that programmers ought to be able to look under the hood to understand an algorithm's decisions as it edits itself. However, this is not the case. As machine learning algorithms are exposed to more data, they autonomously become more "context specific and often based on thousands or millions of factors"[34] in a manner that is indecipherable to human programmers.[35] As examined above, machine learning algorithms identify which variables will lead to more successful predictions. The outputs they produce only demonstrate their success rates and the original source codes the programmers generated—not the ever-changing variables on which they rely.[36] Further, the programmer is unlikely to even be able to determine what these variables are or how much they matter to the evolving algorithm, because the algorithm relies on new variables based on context and begins to assign different weights to each variable as it parses how much that variable matters in the patterns it perceives.[37]

B. How Are Machine Learning Algorithms Used in Policing?

Machine learning has become prevalent across a wide variety of fields and has now become entrenched in the world of policing.[38]

---

32 *See* David Schatsky & Rameeta Chauhan, *Machine Learning and the Five Vectors of Progress*, Deloitte Insights (Nov. 29, 2017), https://www2.deloitte.com/insights/us/en/focus/signals-for-strategists/machine-learning-technology-five-vectors-of-progress.html ("Then there is the black-box problem . . . because [algorithms'] inner workings are inscrutable.").

33 *See* Will Knight, *The Dark Secret at the Heart of AI*, MIT Tech. Rev. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai (describing how even engineers of deep learning algorithms cannot fully explain their decisionmaking processes).

34 Daniel S. Weld & Gagan Bansal, The Challenge of Crafting Intelligible Intelligence 1 (2018), https://arxiv.org/pdf/1803.04263.pdf.

35 *See* Torres, *supra* note 22 (noting that programmers are unable to explain what the algorithms are exactly doing with KNN and neural network techniques).

36 See Figures 1 and 2 for a depiction of what such outputs could look like.

37 *See* Weld & Bansal, *supra* note 34, at 2 ("[W]hen one feature is included in a model, machine learning algorithms extract as much signal as possible from it, indirectly modeling other features that weren't included.").

38 *See* TED Radio Hour, *Cathy O'Neil: Do Algorithms Perpetuate Human Bias?*, NPR (Jan. 26, 2018, 9:12 AM), https://www.npr.org/2018/01/26/580617998/cathy-oneil-do-algorithms-perpetuate-human-bias (explaining that "algorithms are completely controlling [an average person's] experience and their atmosphere and their environment" and providing some examples of the ways they might be biased and dangerous).

There are two primary approaches to machine learning-based predictive policing: person-based and place-based.[39] Knowledge of how each of these types of policing algorithms functions is incomplete because police leadership has been reticent to unveil the information they have about what predictive policing algorithms use in generating results.[40] However, the basic distinction between these two forms of predictive policing is fairly simple. Person-based predictive policing algorithms generate a risk assessment score for an individual, like the Strategic Subjects List did for Mr. McDaniel.[41] Place-based predictive policing algorithms generate a risk score for a particular area.[42] Some forms of predictive policing rely on both person- and place-based assessments of risk in their algorithmic decisions of who and where to police.[43] All of these variants of algorithmic policing present similar equal protection concerns because they are infected with racial biases.

---

39 *See* FERGUSON, *supra* note 6, at 3, 38 (noting how police use algorithms to identify potential sites and victims of future crimes, as well as those who might commit future crimes).

40 *See, e.g.*, *Going Inside*, *supra* note 3 (quoting Deputy Chief Jonathan Lewin as saying, "We don't give out the specific list of variables, but it's things like subject's trendline in criminal activity"); Letter from Rachel Levinson-Waldman, Senior Counsel, Liberty and Nat'l Sec. Program at the Brennan Ctr. for Justice, to Lt. Richard Mantellino, Records Access Officer of the N.Y.C. Police Dep't (June 14, 2016), https://www.brennancenter.org/sites/default/files/NYPD%20Palantir%20FOIL%20061416.pdf (discussing the Freedom of Information Law ("FOIL") litigation in which the Brennan Center has demanded that the New York City Police Department provide information on the predictive policing technology it employs).

41 FERGUSON, *supra* note 6, at 38.

42 *See* Andrew Guthrie Ferguson, *The Truth About Predictive Policing and Race*, APPEAL (Dec. 7, 2017), https://theappeal.org/the-truth-about-predictive-policing-and-race-b87cf7c070b1 (distinguishing between person- and place-based predictive policing in critiquing a piece which had conflated the two practices); *see also, e.g.*, *About*, PREDPOL, http://www.predpol.com/about/ (last visited Jan. 22, 2018) ("For us and our customers, [predictive policing] is the practice of identifying the times and locations where specific crimes are most likely to occur, then patrolling those areas to prevent those crimes from occurring.").

43 Palantir is a prime example of a software company that has built out algorithms that generate both people- and place-based predictive policing decisions. U.S. Patent No. 9,836,694 B2 (filed Sept. 2, 2015), https://patents.google.com/patent/US9836694B2/en?oq=inassignee (describing the product as a "computer-based crime risk forecasting system . . . for generating crime risk forecasts and conveying the forecasts to a user . . . [so that] the user can more effectively gauge both the level of increased crime threat and its potential duration" with a map). It is, however, one of the most secretive (and lucrative) companies in Silicon Valley, with very little known about the inner workings of its algorithms. *See* Mark Harris, *How Peter Thiel's Secretive Data Company Pushed into Policing*, WIRED (Aug. 9, 2017), https://www.wired.com/story/how-peter-thiels-secretive-data-company-pushed-into-policing (expressing frustration that "[n]o one outside Palantir seems to know for sure how many police departments in America use its technology").

C. How Can Machine Learning-Based Predictive Policing Algorithms Be Biased?

The concern for scholars and advocates of police reform is that machine learning-based predictive policing algorithms can reinforce the patterns of racist policing in the United States under the imprimatur of neutrality.[44] Such replication is a real threat because these algorithms rely on biased data.[45] There are two types of data that predictive policing algorithms rely on: historical crime data and dragnet data searches. Historical crime data is laced with racial biases against people of color.[46] Dragnet data searches are laden with the racial biases of the web—a place where white supremacy is blatant.[47]

1. Historical Crime Data Is Racially Biased

Criminological research since the nineteenth century has shown that police databases and the information they contain about individuals' and neighborhoods' contacts with the police are not a "complete census of all criminal offenses, nor do they constitute a representative sample."[48] At the start of the twentieth century, "in a rapidly industrializing, urbanizing, and demographically shifting America, blackness was refashioned through crime statistics"[49] because of the way crime data was generated and recorded.[50] Criminal contacts data often

---

[44] *See* The Leadership Conference on Civil & Human Rights et al., Predictive Policing Today: A Shared Statement of Civil Rights Concerns (2016), http://civilrightsdocs.info/pdf/FINAL_JointStatementPredictivePolicing.pdf (arguing that predictive policing algorithms replicate "[t]he institution of American policing into which [they] are being introduced, [which] is profoundly flawed").

[45] *Id.* (demonstrating a concern that historic crime data is biased in that it is a "*response* to the reports [police] receive and the situations they encounter, rather than . . . a consistent or complete record of all crimes that occur") (emphasis in original).

[46] *See infra* Section I.C.1 (delineating the ways in which historical crime data is skewed to overrepresent people of color, particularly Black Americans).

[47] *See infra* Section I.C.2 (discussing the racial biases baked into the Internet).

[48] Kristian Lum & William Isaac, *To Predict and Serve?*, Significance, Oct. 2016, at 14, 15, https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x (citing Steven D. Levitt, *The Relationship Between Crime Reporting and Police: Implications for the Use of Uniform Crime Reports*, 14 J. Quantitative Criminology 61, 62–63 (1998) (investigating the manner in which historical crime data can lead to biases in PredPol's algorithmic outputs)); *see also* Patrick A. Langan, Bureau of Justice Statistics, U.S. Dep't of Justice, The Racial Disparity in U.S. Drug Arrests (1995), https://www.bjs.gov/content/pub/pdf/rdusda.pdf.

[49] Khalil Gibran Muhammad, The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America 5 (2010).

[50] For an example of a biased individual in the modern day who creates data later used in predictive policing, look to Kristine de Leon's analysis of an interview with Brian Hoepner, a senior crime analyst at the Los Angeles Sheriff's Department West Hollywood station; Mr. Hoepner claims to know that "[t]he same people are doing all the crimes all the time" and that he has "to keep an eye on the homeless because they're involved in a lot

directly correlate with racist police practices.[51] Black and Hispanic Americans are more likely to have contact with the police in the United States.[52] New York's former stop-question-and-frisk policy serves as a prime example of the disproportionate contact that Black and Hispanic Americans have with police. Between 2004 and 2012, the New York City Police Department made approximately 4.4 million stops, over 80% of which involved people of color.[53] More precisely, 52% of these 4.4 million stops involved Black Americans and 31% involved Hispanic Americans.[54] In 2010, 23% of New York City's population was Black, while 33% was white.[55] Much has been written about Black Americans' higher rate of exposure to police contact outside the confines of New York City.[56] For example, according to an

of general crimes . . . [and] most of them are drug addicts . . . ." Kristine de Leon, *Predicting Crime*, USC STORY SPACE (May 2017), https://uscstoryspace.com/2017-2018/plocek/kristine_capstone/index.html.

51 *See* FERGUSON, *supra* note 6, at 47; Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (discussing the manner in which COMPAS, the assessment system used to determine the risk of flight by a person seeking bail, is biased against Black Americans); Stephen Buranyi, *Rise of the Racist Robots—How AI Is Learning All Our Worst Impulses*, GUARDIAN (Aug. 8, 2017), https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses (indicating that "the US justice system, reviled for its racial bias . . . turned to technology for help, only to find that the algorithms had a racial bias too").

52 *See* FERGUSON, *supra* note 6, at 47 (citing Kenneth B. Nunn, *Race, Crime and the Pool of Surplus Criminality: Or Why the "War on Drugs" Was a "War on Blacks*,*"* 6 J. GENDER, RACE, & JUST. 381, 391–412 (2002)) (explaining that "disproportionate police contact . . . permeates life in major urban cities" because many of the variables predictive policing algorithms analyze, including "police contacts, prior arrests, [and] gang affiliations . . . directly correlate with racially discriminatory law enforcement practices"); Jeffrey Fagan & Garth Davies, *Street Stops and Broken Windows:* Terry, *Race, and Disorder in New York City*, 28 FORDHAM URB. L.J. 457, 457–58, 482 (2000) (highlighting empirical evidence showing racially disparate police targeting); David Rudovsky, *Law Enforcement by Stereotypes and Serendipity: Racial Profiling and Stops and Searches Without Cause*, 3 U. PA. J. CONST. L. 296, 300 (2011) (highlighting a study by the Attorney General of New Jersey indicating that Blacks and whites violate traffic laws at the same rate, but 42% of stops and 73.2% of arrests were of Black people, resulting in serious disparities).

53 Floyd v. City of New York, 959 F. Supp. 2d 540, 556 (2013).

54 *Id.* at 559.

55 *Id.*

56 *See, e.g.*, ACLU, THE WAR ON MARIJUANA IN BLACK AND WHITE 4, 47, 49 (2013), https://www.aclu.org/report/report-war-marijuana-Black-and-white (finding that Black Americans were, on average, 3.73 times more likely to be arrested for marijuana possession than whites were, despite the fact that, according to the report, Blacks and whites use marijuana at similar rates); Harry Cockburn, *Black People Now Eight Times More Likely to Be Stopped and Searched, Figures Show*, INDEPENDENT (Oct. 26, 2017), http://www.independent.co.uk/news/uk/home-news/Black-people-stop-and-search-uk-stats-figures-latest-more-likely-a8022161.html; Jeff Guo, *America's Tough Approach to Policing Black Communities Began as a Liberal Idea*, WASH. POST (May 2, 2016), https://www.washingtonpost.com/news/wonk/wp/2016/05/02/americas-tough-approach-to-policing-Black-communities-began-as-a-liberal-idea; Christine Hauser & Jacey Fortin, *'We Only*

investigative report in Ferguson, Black Americans are two times more likely than whites to be searched during vehicle stops, even after controlling for non-race-based variables, such as the reason the stop was initiated, but Black Americans are found in possession of contraband 26% less often than white drivers.[57] Black Americans accounted for more than a third of violent crime arrests in 2010, far surpassing their representation in the United States' population.[58] Over a five-year period, the Vera Institute found that Black people accounted for 94% of felony marijuana arrests and 85% of overall marijuana arrests in New Orleans.[59] This number contrasts sharply with the fact that Black people compose approximately 60% of the population of New Orleans.[60]

There are a variety of reasons why communities of color may experience disparate policing, but it is not because race is an accurate predictor of crime. The Department of Justice's Crime Victimization Survey estimates that 42% of violent crime and 60% of household property crime goes unreported each year.[61] Thus, police are missing major swaths of crime happening in other communities. The disproportionate policing of communities of color may stem, in part, from biases of officers. Empirical evidence demonstrates that "police officers – either implicitly or explicitly – consider race and ethnicity in their determination of which persons to detain and search and which neighbourhoods to patrol."[62] Disparate policing may also stem from white civilians' implicit biases, which cause them to conceive of people of color as "more dangerous" in some way, and cause them to call the police more frequently to address people of color than they would for

---

Kill Black People,' Police Officer Says During Traffic Stop, N.Y. TIMES (Aug. 31, 2017), https://www.nytimes.com/2017/08/31/us/Black-kill-police-georgia.html.

57 U.S. DEP'T OF JUSTICE, CIVIL RIGHTS DIV., INVESTIGATION OF THE FERGUSON POLICE DEPARTMENT 4–8 (2015), https://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_department_report.pdf.

58 *The Color of Justice*, CONSTITUTIONAL RIGHTS FOUND., http://www.crf-usa.org/brown-v-board-50th-anniversary/the-color-of-justice.html (last visited Feb. 16, 2019).

59 *See* FERGUSON, *supra* note 6, at 48 (citing MEGHAN RAGANY ET AL., VERA INST. OF JUSTICE, RACIAL DISPARITY IN MARIJUANA POLICING IN NEW ORLEANS 9–10 (2016), https://www.vera.org/publications/racial-disparity-in-marijuana-policing-in-new-orleans).

60 *Id.*

61 U.S. DEP'T OF JUSTICE, NATIONAL CRIME VICTIMIZATION SURVEY: VICTIMIZATIONS NOT REPORTED TO THE POLICE, 2006–2010, at 2 (2012).

62 Lum & Isaac, *supra* note 48, at 15 (citing Andrew Gelman et al., *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N 813, 813–23 (2007); *see also* L. Song Richardson & Phillip Atiba Goff, *Self-Defense and the Suspicion Heuristic*, 98 IOWA L. REV. 293, 305–07, 313, 316–17, 332 (2012) (illustrating the ways in which implicit racial bias impacts police interactions).

similar behavior of white people.[63] No matter the underlying reasoning, historical crime data is racially skewed.

The use of historical crime data in machine learning-based predictive policing is particularly unique because police are not just the end users of the algorithmic outputs. The police create the information that the algorithms use when the algorithms use historical crime data.[64] Scholars and advocates fear that the biased human nature of the input data, in conjunction with the nature of machine learning algorithms, leads to a "garbage in garbage out" phenomenon.[65]

2. *Dragnet Data Searches Are Racially Biased*

Some algorithms train and rely on big data mining of publicly and commercially available records from the web, which are embedded with racism. Data mining can dredge up criminal history data and criminal records,[66] and in this way, data mining is biased in the same ways as the criminal history data examined in Section I.C.1. However, there are additional biases to consider in data mining. The Internet is racist, because "the internet is still dominated by the richer, more educated . . . parts of the world"[67] and thus does not include the thoughts, ideas, and broad representation of the less educated and impoverished, many of whom are people of color.[68] The Internet's racism is

---

[63] *See, e.g.*, Jenny Gathright & Emily Sullivan, *Starbucks, Police and Mayor Respond to Controversial Arrest of 2 Black Men in Philly*, NPR: THE TWO-WAY (Apr. 14, 2018, 11:56 PM), https://www.npr.org/sections/thetwo-way/2018/04/14/602556973/starbucks-police-and-mayor-weigh-in-on-controversial-arrest-of-2-black-men-in-philly (describing a situation in which Starbucks employees called the police on two Black men because they waited in the store for a friend to arrive without ordering anything); *see also Implicit Bias*, NAT'L INITIATIVE FOR BUILDING CMTY. TRUST & JUSTICE, https://trustandjustice.org/resources/intervention/implicit-bias (last visited Jan. 22, 2018) (explaining that implicit bias is the way in which "racism without racists" develops and can lead white people to perceive Black people as more dangerous than people of other races because of long-standing racist views that have ingrained themselves into American societal views); REBECCA EPSTEIN ET AL., GEORGETOWN LAW CTR. ON POVERTY AND INEQUALITY, GIRLHOOD INTERRUPTED: THE ERASURE OF BLACK GIRLS' CHILDHOOD 2, 3–19 (2018) (examining the ways in which society views Black boys as less innocent than white male peers and Black girls as older and less innocent, resulting in harsher outcomes and punishments).

[64] Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, 26 WM. & MARY BILL RTS. J. 287, 289 (2017).

[65] *See id.* at 294 ("[W]hen algorithms in the criminal justice system rely upon data that contains racial bias, the machine learning algorithms . . . will inevitably reflect that racial bias.").

[66] *See* Anita L. Allen, *Dredging up the Past: Lifelogging, Memory, and Surveillance*, 75 U. CHI. L. REV. 47, 61 (2008) (discussing the fact that "[c]riminal histories are public data").

[67] Adrianne Jeffries, *Machine Learning Is Racist Because the Internet Is Racist*, OUTLINE (Apr. 27, 2017, 11:05 AM), https://theoutline.com/post/1439/machine-learning-is-racist-because-the-internet-is-racist.

[68] *See id.*

evident in considering something as simple as a Google search. Names that "sound black" are 25% more likely to prompt Google to display connections to criminal records than names that "sound white," even when the person associated with the name has no criminal record.[69] When Safiya Umoja Noble—author of *Algorithms of Oppression: How Search Engines Reinforce Racism*—searched the term "black girls" in Google for the first time, she found that the search results were dominated by pornography.[70] If these are the results of a human's quick Google search, an algorithm, learning from "billions of data points" available about a person or place on the web through a dragnet data search,[71] is undoubtedly learning from even more racist inputs because the web's values "reflect its builders—mostly white, Western men—and do not represent minorities and women."[72]

3. *When the Data Is Racist, the Algorithm Is Racist*

As examined in Section I.A, there are two likely forms of machine learning that machine learning-based predictive policing algorithms employ: a K-Nearest Neighbors model or a deep learning artificial neural network. A KNN algorithm is capable of discerning that race is a "near neighbor" of some of the variables that the programmer originally trained it to select for. For example, if a KNN algorithm relies on historical crime data and is programmed to predict where future crime will happen based on where it has happened before, the algorithm could find that there is more crime in predominantly Black and Hispanic neighborhoods and determine race is a "near neighbor" of the location variable that the human programmer originally programmed it to select for.

A deep learning neural network does not rely on humans to tell it what to look for in discerning criminality. Instead, based on the examples it is fed, it makes associations and defines criminality on its own. So, a deep learning neural network that is trained on historical crime data could examine countless rap sheets and determine that race is a good predictor for crime in the United States, because of the overrep-

---

69 Hiawatha Bray, *Racial Bias Alleged in Google's Ad Results*, Bos. Globe (Feb. 6, 2013), https://www.bostonglobe.com/business/2013/02/06/harvard-professor-spots-web-search-bias/PtOgSh1ivTZMfyEGj00X4I/story.html.

70 Jackie Snow, *Bias Already Exists in Search Engine Results, and It's Only Going to Get Worse*, MIT Tech. Rev. (Feb. 26, 2018), https://www.technologyreview.com/s/610275/meet-the-woman-who-searches-out-search-engines-bias-against-women-and-minorities.

71 *See* U.S. Patent No. 9,836,694 B2 at 1 (filed Sept. 2, 2015) (describing the product as a "computer-based crime risk forecasting system . . . for generating crime risk forecasts and conveying the forecasts to a user").

72 Snow, *supra* note 70.

resentation of people of color in the criminal justice system.[73] If a deep learning neural network is trained on dragnet data searches, it will be endowed with the Internet's racism and will be capable of associating race and criminality just like a racist human being. Whether a predictive policing algorithm is a KNN algorithm or an artificial neural network, any algorithm that associates race and criminality will subsequently consider people of color and their neighborhoods more likely to be the possible perpetrators, victims, and sites of future crimes. It is for this reason that this Note argues that these algorithms are capable of facially discriminating on the basis of race and that they can be challenged under the Equal Protection Clause, as is examined in Part II.

But how can we *know* that machine learning-based predictive policing algorithms are capable of being racist when police departments and the tech industry have not made the algorithms' source codes and their outcomes available to the public?[74] To answer this crucial question, this section of the Note examines various ways in which machine learning algorithms have demonstrated racial bias when exposed to historical crime data or dragnet data searches.

Tay, Microsoft's now-defunct automated chatbot, is a prime example of the manner in which biased inputs can lead to a racist algorithm via machine learning, even when an algorithm was never coded to consider race. Microsoft's aim in launching Tay was to "'experiment with and conduct research on conversational understanding,' with Tay [who would be] able to learn from 'her' conversations and get progressively 'smarter.'"[75] However, when Tay was exposed to Twitter, she was also exposed to racist commentary and biases that exist on the social media platform. These racist biases "then became part of the data corpus" that Tay integrated into her algorithmic processing.[76] Tay quickly learned from the Twitter inputs, and her Tweets turned into sickeningly hateful tirades.[77] Tay's ability

---

73 *See supra* Section I.C.1.

74 *See infra* Section II.C.

75 Rob Price, *Microsoft Is Deleting Its AI Chatbot's Incredibly Racist Tweets*, BUS. INSIDER (Mar. 24, 2016), http://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3.

76 *See* Lum & Isaac, *supra* note 48, at 16 (discussing 4chan users' interactions with Tay).

77 Tay answered the Tweet "Did the Holocaust happen?" with the response "it was made up" and a clapping hands emoji. Price, *supra* note 75. A few hours later, Tay said, "I f[*]cking hate n[*]ggers, I wish we could put them all in a concentration camp . . . and be done with the lot." *Id.* (alteration in Price piece reproducing Tay's original, uncensored tweets). Finally, before she was taken down, Tay responded to a Tweet that read, "We must secure the existence of our people and a future for white children," with "could not agree more. i wish there were more people articulating this kind of thing." *Id.*

to internalize human racism is an example of the malleable nature of machine learning in the face of the Internet's racism.

Similar to Tay's evolution in becoming biased against minority groups, ProPublica demonstrated that machine learning algorithms in the criminal justice context can learn to be "biased against blacks," even when such algorithms are not coded to rely on race as a variable.[78] ProPublica examined risk assessment machine learning algorithms, which various actors in the criminal justice system use to determine the "dangerousness" of particular defendants and "predict future criminals" in a manner similar to some predictive policing algorithms.[79] The developers of the crime risk assessment algorithms did not write the algorithms' source codes to rely on race as a variable.[80] ProPublica obtained seven thousand individuals' risk scores from 2013 and 2014 and investigated whether or not those individuals were charged with committing any new crimes in the subsequent two years.[81] ProPublica found that the predictive algorithm was "particularly likely to falsely flag black defendants as future criminals, wrongly labeling them [as greater crime risks] at almost twice the rate as white defendants."[82] Further, the researchers discovered that "[w]hite defendants were mislabeled as low risk" far more often than Black defendants.[83]

In order to dispel any concerns that this racist outcome could be attributed to Black defendants' prior crimes or the types of crimes for which police arrested Black defendants, ProPublica isolated the effect of race from criminal history, recidivism, age, and gender, and found that, "Black defendants were still 77 percent more likely to be pegged as at higher risk of committing a future violent crime."[84] Further, Black defendants were forty-five percent more likely to be identified as a perpetrator of any future crime at all.[85] The study concluded that the machine learning algorithm began to manifest bias against Black defendants because it learned to associate race with criminality.

---

[78] *See* Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, ProPublica (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (discussing the methodology and findings of ProPublica's COMPAS study).

[79] *See* Angwin et al., *supra* note 78.

[80] *Id.* (indicating that "[r]ace is not one of the questions" considered directly by the algorithm).

[81] *Id.* (describing the study).

[82] *Id.*

[83] *Id.*

[84] *Id.*

[85] *Id.*

Tay and the ProPublica study demonstrate that feeding an algorithm biased data sets—like historical policing data sets and Internet searches, which are riddled with bias[86]—into a machine learning algorithm can allow the algorithm to draw an association between race and criminality, even when the algorithm is not explicitly coded to use race as a variable. A recent *Science* publication suggests that input data used to train machine learning algorithms often reflect historic biases, leading the algorithm to make associations between terms such as race and criminality.[87] Researchers trained a machine learning algorithm by exposing it to roughly 840 billion words from the standard corpus of the Internet as the input data, and then subjected the algorithm to a version of the Implicit Association Test,[88] which measures whether a subject associates particular words or pictures with another set of concepts (like pleasantness or unpleasantness).[89] Instead, they were a sample of the vast array of different texts that exist on the web. In exposing the algorithm to approximately 840 billion words, the researchers were using word embeddings as training data for the algorithm. Word embeddings are the ways words appear together on web pages.[90] When the algorithm examines the manner in which words are strung together across billions of different web sites, the algorithm can perceive patterns in the ways that those words are used on the Internet, and the algorithm can then incorporate those patterns into its code.[91]

After the researchers trained the algorithm and subjected it to the Implicit Association Test, they found that the algorithm deemed European-sounding names to be more easily associated with "pleasant" terms, while it found Black-sounding names to be more easily associated with "unpleasant" terms.[92] To demonstrate the implications of this finding, a different group of researchers exposed the algorithm to text on the web and then used that algorithm to generate restaurant reviews. These researchers found that the algorithm "picked up . . . that the words Mexican and the phrase illegal immigrant often occur in proximity to each other . . . and so it picked up that the word Mexican is somehow related to illegal, and therefore

[86] *See supra* Sections I.C.1–2.

[87] *See generally* Aylin Caliskan et al., *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, 356 SCIENCE 183, 183 (2017).

[88] *See id.* (describing the contours of the study).

[89] *Id.*

[90] *See id.* at 184.

[91] *See id.* at 185; *see also Panel*, *supra* note 30 (interviewing Arvind Narayanan, one of the primary researchers involved in this project).

[92] *See* Caliskan et al., *supra* note 87, at 183; *see also Panel*, *supra* note 30 (explaining the results).

must have a negative connotation."[93] As a result, the algorithm incorporated this pattern into its code, and ranked Mexican restaurants lower than all other restaurants.[94]

Researchers have demonstrated that machine learning-based predictive policing algorithms that rely on biased data sources facially discriminate and lead to racially discriminatory outcomes. The Human Rights Data Analysis Group's (HRDAG) Dr. Kristian Lum and Dr. William Isaac examined precisely this premise as applied to predictive policing algorithms. HRDAG specifically examined the way that PredPol—a well-known predictive policing algorithm—targets Black populations in comparison to the way that it targets white populations.[95] In order to determine whether police data sets used in algorithms actually are biased against Black Americans, the team needed to compare crimes recorded by the police to a complete record of all crimes that occurred, whether or not they had been recorded by the police.[96] To investigate the effect of police-recorded data used as the input in PredPol, the team fed the algorithm recently published drug crime records from Oakland, California.[97]

The researchers found that the PredPol algorithm reinforces biases, "rather than correcting for the apparent biases in the police data."[98] In their study, the researchers demonstrated that PredPol sends police to Black neighborhoods, like West Oakland, far more than white neighborhoods, like Piedmont.[99] PredPol would ensure that Black people would be targeted by the algorithm at twice the rate

---

93 *Panel*, *supra* note 30.

94 *Id.*

95 *See* Lum & Isaac, *supra* note 48, at 17–18.

96 *Id.* at 16 ("How biased are police data sets? To answer this, we would need to compare the crimes recorded by police to a complete record of all crimes that occur . . . ."). The research team accomplished this goal by combining "a demographically representative *synthetic population* of Oakland, California," because "there is no 'ground truth' data set containing a representative sample of local crimes." *See id.* at 16–17. A synthetic population is "a demographically accurate individual-level representation of a real population . . . . Here, individuals . . . are labelled with their sex, household income, age, race, and the geo-coordinates of their home. These characteristics are assigned so that the . . . synthetic population [data] match data from the US Census at the highest geographic resolution possible." *See id.* at 16.

97 *Id.* at 17–18. The team estimated the number of drug users in the synthetic population by fitting a model to data from the 2011 National Survey on Drug Use and Health that would "predict[ ] an individual's probability of drug use within the past month based on their demographic characteristics (i.e. sex, household income, age, and race). Then, [they] appl[ied] this model to each individual in the synthetic population to obtain an estimated probability of drug use for every synthetic person in Oakland." *Id.* at 16.

98 *Id.* at 18.

99 Jack Smith IV, *Crime-Prediction Tool May Be Reinforcing Discriminatory Policing*, Bus. Insider (Oct. 10, 2016, 7:02 PM), http://www.businessinsider.com/predictive-policing-discriminatory-police-crime-2016-10.

of white people.[100] This contrasted with the fact that patterns of drug use were equivalent across racial classifications in Oakland.[101] Crime is "everywhere, but police only find it where they're looking."[102] Thus, the team concluded, "predictive policing . . . results in increasingly disproportionate policing of historically over-policed communities."[103] Despite the fact that the group only focused on PredPol, their conclusions "are applicable to any predictive policing algorithm that uses unadjusted police records to predict future crime."[104]

Machine learning algorithms can pick up on racist patterns across wide swaths of data because their entire purpose is to recognize patterns and subsequently incorporate those patterns into their own decisionmaking.[105] Machine learning-based predictive policing algorithms that are not explicitly coded to seek out race as a factor can still learn that race is associated with criminality when they are exposed to historical crime data and big data mining when race is not redacted from these data sets. This can cause predictive policing algorithms to target Black neighborhoods and Black Americans by explicitly classifying, in part, on the basis of race. In incorporating a pattern of racial bias into their codes, predictive policing algorithms do not improve at what they purport to do—*objectively* predict where crime is most likely to happen next.[106] Further, the algorithm insulates racially biased practices because officers can truly claim they are making decisions based on a computer-generated output, not based on human biases. However, "if the data is biased to begin with and based on human judgment, then the results the algorithm is going to spit out will reflect those biases."[107]

[100] *See* Lum & Isaac, *supra* note 48, at 18.

[101] *See id.*

[102] Smith, *supra* note 99 (quoting Dr. Kristian Lum).

[103] Lum & Isaac, *supra* note 48, at 19.

[104] *Id.* at 18.

[105] *See Panel*, *supra* note 30 (featuring Cathy O'Neil, who said, "all machine learning algorithms do is recognize patterns, recognize patterns").

[106] Instead, by learning to associate race and criminality, a machine learning-based predictive policing algorithm "ends up being a self-fulfilling prophecy. . . . The algorithm is telling you exactly what you programmed it to tell you. 'Young [B]lack kids in [a predominantly Black neighborhood] are more likely to commit crimes.'" FERGUSON, *supra* note 6, at 47 (quoting Bryan Llenas, *Brave New World of 'Predictive Policing' Raises Specter of High-Tech Racial Profiling*, FOX NEWS (Feb. 25, 2014), http://www.foxnews.com/world/2014/02/24/brave-new-world-predictive-policing-raises-specter-high-tech-racial-profiling.html).

[107] *Id.*

4. *Racist Algorithms Will Exacerbate Racist Policing*

The preceding section established that machine learning algorithms can be racist, but the question remains: If algorithms discriminate because they simply learn to discern biases that already exist in humans, why are predictive policing algorithms any worse than the status quo of biased human police? The answer is that these algorithms can create feedback loops, whereby predictive policing becomes more virulent in its racial bias over time. Feedback loops exacerbate disparate policing of communities of color. In subsequent work to their initial PredPol study, HRDAG hypothesized that sending police to neighborhoods that the PredPol algorithm selected would lead to an increase in reported crime by twenty percent in those neighborhoods.[108] The researchers input the twenty percent increase in arrests in West Oakland back into the algorithm, which then "became orders of magnitude more confident that its predictions were correct. . . . '[This] creates a feedback loop, [in which] the algorithm becomes more certain about these places that are over-policed.'"[109] This is the ultimate danger of predictive policing. Via machine learning, algorithms learn how to get "better" at recreating the racially biased patterns that they discerned in the data.

The process by which predictive policing algorithms create feedback loops which reinforce deleterious and biased patterns happens over time. First, algorithms "go with the winner."[110] This means that even if crime rates in two neighborhoods are remarkably similar, "if region A has a crime rate of 10% and region B has a crime rate of 11%, the update process will settle on region B" with 100% probability.[111] When police are sent to a particular region repeatedly, they are more likely to see crime there.[112] This "predispos[es] [police] to collect more crime from one region than the other."[113] Since Black neighborhoods already face disproportionately higher reported crime rates, this means that the algorithm will, over time, consistently learn to send police only to Black neighborhoods. Further, the lack of observations about the under-policed region "prevents the system

---

108 *See* Smith, *supra* note 99 (describing the specifics of the subsequent study based on an interview with Kristian Lum and her team).

109 *See id.* (quoting Dr. Kristian Lum).

110 *See* Danielle Ensign et al., *Runaway Feedback Loops in Predictive Policing*, 81 Proc. Machine Learning Res. 160, 165 (2018), https://arxiv.org/pdf/1706.09847.pdf.

111 *Id.*

112 *See* The Conference on Fairness, Accountability, & Transparency (FAT*), *FAT*2018: Carlos Scheidegger – Runaway Feedback Loops in Predictive Policing*, YouTube (Mar. 7, 2018) [hereinafter FAT*], https://www.youtube.com/watch?v=9C6epG-Wyuw (discussing the study by Ensign et al.).

113 *Id.*

from learning" that the crime rates of two regions are actually very similar.[114] Under this analysis, predictive policing algorithms will learn less about crime in predominantly white areas and will report that there is less of a risk of future crime in those areas, while learning more about predominantly Black neighborhoods and indicating that more police personnel should be sent to those areas.[115]

II
Machine Learning-Based Predictive Policing Algorithms Can Be Challenged Under the Equal Protection Clause

Because machine learning-based predictive policing algorithms can discern race, learn that race is associated with an increased likelihood of future criminality, and incorporate race as such into future decisionmaking, they are facially discriminatory, and can be challenged on equal protection grounds. Section II.A provides an overview of the modern equal protection framework. Section II.B applies this framework to machine learning-based predictive policing algorithms. Section II.C examines the weaknesses in bringing such a claim.

A. The Modern Equal Protection Framework

The Equal Protection Clause of the Fourteenth Amendment guarantees every person "equal protection of the laws."[116] A litigant can only challenge a state actor's policy under the Equal Protection Clause when that policy intentionally discriminates on the basis of a "suspect" or "quasi-suspect classification."[117] Race, national origin, religion, and alienage are considered suspect classifications, while gender is considered a quasi-suspect classification.[118] State action that is facially neutral and lacks evidence of discriminatory intent against the class is insufficient for a litigant to make out an equal protection claim.[119] For this reason, the burden is initially on the plaintiff

114 Ensign et al., *supra* note 110, at 163 (emphasis omitted).

115 *See id.* at 160 (discussing the result of feedback loops on predictive policing algorithms); FAT*, *supra* note 112 (indicating that, because of the "go with the winner" phenomenon, the algorithms will continuously send officers to the areas with higher crime rates, thus under-policing the area with the lower crime rate and continuing to learn less about it, maintaining the cycle).

116 U.S. Const. amend. XIV, § 1.

117 *Equal Protection*, Legal Info. Inst., https://www.law.cornell.edu/wex/Equal_protection (last visited Mar. 12, 2019) (laying out the equal protection framework and distinguishing between intermediate and strict scrutiny).

118 *Id.*

119 The Court set out this constraining standard in a series of decisions, which began with *Washington v. Davis*, 426 U.S. 229, 245–48 (1976). *See* Kenji Yoshino, *The New Equal*

bringing the equal protection challenge to demonstrate either (a) that the state actor's policy is facially discriminatory in that it relies explicitly on race or (b) that the policy is facially neutral but has a disparate impact on an identifiable group and the policy discriminates "because of" its impact on the suspect or quasi-suspect class, not merely "in spite of" its effects upon that identifiable group.[120]

If the plaintiff can demonstrate that the policy intentionally discriminates based on the suspect or quasi-suspect classification, the court reviews the policy under heightened scrutiny.[121] If, more specifically, a litigant demonstrates that the policy intentionally discriminates on the basis of race, the policy is subjected to review under the most stringent form of heightened scrutiny: strict scrutiny.[122] Under strict scrutiny review, the government must first demonstrate that the policy carries out a government purpose that is "both constitutionally permissible and substantial."[123] The court may give "some, but not complete, judicial deference" to the state actor's experience and expertise in defining its purpose.[124]

The requirement that a plaintiff show that the policy either facially discriminates or embodies a discriminatory purpose does not mean that the plaintiff must prove that the challenged policy was motivated solely by a discriminatory purpose.[125] Additionally, the plaintiff does not need to demonstrate that the policy was motivated by the state actor's hostility or ill will towards the affected group.[126] For instance, in the context of affirmative action, a university's admis-

Protection, 124 HARV. L. REV. 747, 763–64 (2011) (interpreting *Davis* as the first case to "declare[ ] that facially neutral state action would draw only ordinary rational basis review so long as it was not enacted with discriminatory intent"). In *Davis*, the Court determined that disparate impact, though insufficient to trigger heightened scrutiny, could still be probative of discriminatory intent. *See Davis*, 426 U.S. at 242. However, the year after the Court handed down *Davis*, it issued *Personnel Administrator of Massachusetts v. Feeney*, 442 U.S. 256 (1979) and "made disparate impact almost irrelevant" by requiring that a plaintiff be able to demonstrate that the state action was taken "because of", not merely "in spite of" its discriminatory intent. Yoshino, *supra*, at 764 (quoting *Feeney*, 442 U.S. at 279).

120 *Feeney*, 442 U.S. at 279.

121 Korematsu v. United States, 323 U.S. 214, 216 (1944) (establishing that "all legal restrictions which curtail the civil rights of a single racial group are immediately suspect . . . [and] courts must subject them to the most rigid scrutiny").

122 *Id.*

123 Fisher v. Univ. of Tex. at Austin (*Fisher I*), 570 U.S. 297, 309 (2013) (quoting Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 305 (1978)).

124 *Id.* at 310 (citing Grutter v. Bollinger, 539 U.S. 306, 328 (2003)).

125 *See* Hayden v. Paterson, 594 F.3d 150, 163 (2d Cir. 2010) ("[A] plaintiff need not prove that the 'challenged action rested solely on racially discriminatory purposes.'" (quoting Vill. of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 265 (1977))).

126 Floyd v. City of New York, 959 F. Supp. 2d 540, 662 (S.D.N.Y. 2013) (quoting Ferrill v. Parker Grp., Inc., 168 F.3d 468, 473 & n.7 (11th Cir. 1999)).

sions policy is not motivated solely by considerations of race, nor is it motivated by hostility towards particular groups. A university that uses an affirmative action policy, of course, considers additional factors beyond race for each applicant.[127] Further, the university's purpose in relying on race in admissions is probably to achieve diversity in the classroom, not to keep out applicants because the university harbors animosity towards them based on the color of the applicants' skin.[128] Nonetheless, the Supreme Court has determined that affirmative action policies are subject to strict scrutiny because they facially utilize a suspect classification, and that a university's affirmative action policy that automatically assigns additional points to an applicant on the basis of race is a violation of the Equal Protection Clause.[129]

If the state actor can establish that the policy furthers a substantial government purpose, the government must subsequently demonstrate that the policy is narrowly tailored, meaning the policy is "'necessary . . . to the accomplishment' of the [articulated government] purpose."[130] On this issue of narrow tailoring, the state actor "receives no deference."[131] The state actor's expertise is not relevant in determining whether the means chosen to accomplish the purpose are as narrowly framed as possible.[132] In the context of affirmative action, narrow tailoring mandates "meaningful, individualized review" of applicants rather than the automatic attribution of additional points to an applicant's admission score on the basis of race.[133]

B. Generating an Equal Protection Claim Against Machine Learning-Based Predictive Policing Algorithms

Machine learning-based predictive policing algorithms are potential violations of the Fourteenth Amendment's Equal Protection Clause when they are trained on historical crime data or dragnet searches because this information allows the algorithms to classify and

127 Gratz v. Bollinger, 539 U.S. 244, 253 (2003) (indicating that the University of Michigan's admissions program considered grades, test scores, extracurriculars, and a host of other qualities in determining whether or not to admit an applicant).

128 *E.g.*, *id.* at 246 (detailing how the goal of the University of Michigan's consideration of race was to increase community diversity).

129 *Id.* at 269–70 (finding that the University of Michigan's admission criteria were not narrowly tailored as required by the strict scrutiny triggered by a "racial classification reviewable under the Equal Protection clause" (quoting Adarand Constructors, Inc. v. Peña, 515 U.S. 200, 224 (1995))).

130 *Id.* at 309 (quoting Regents of Univ. of Cal. v. Bakke, 438 U.S. 265, 305 (1978)).

131 *Id.* at 311.

132 *Id.* (citing Grutter v. Bollinger, 539 U.S. 306, 333 (2003)).

133 *Id.* (citing *Bakke*, 438 U.S. at 265).

target on the basis of race. Plaintiffs can argue that an algorithm facially discriminates on the basis of race to trigger review of an algorithm under strict scrutiny. Plaintiffs can argue that a racist predictive policing algorithm is similar to two different policies that courts have found to be facially discriminatory on the basis of race: affirmative action programs and racially motivated police officers. Plaintiffs can draw a connection between predictive policing algorithms and the affirmative action program that the Supreme Court struck down in *Gratz v. Bollinger*.[134] In *Gratz*, the University of Michigan automatically assigned additional points to underrepresented minority candidates.[135] Predictive policing algorithms learn to function in a similar manner. Through machine learning in conjunction with racist data inputs, these algorithms begin to automatically associate race with criminality.[136] Once they are programmed, these algorithms function like an admissions officer who has been told to automatically assign additional points to the applications of students of color; the algorithms automatically determine that people of color are more likely to perpetrate future crime and the neighborhoods they live in are more likely to experience future crime. This sort of automatic decisionmaking on the basis of race is precisely what the Court has forbidden.[137]

A litigant could also analogize predictive policing algorithms' discrimination based on race to a police chief directing officers to target people of color. In *Floyd v. City of New York*, the Southern District of New York reviewed the New York City Police Department's "unwritten policy of targeting 'the right people' for stops."[138] This unwritten policy was one which involved stopping young Black and Hispanic men "based on their prevalence in local crime complaints."[139] A group's overrepresentation in crime statistics does not mean that it is permissible to subject all members of that group to increased police focus; to do so, in fact, is impermissible "racial profiling."[140] Since these algorithms rely heavily on historical crime data, they rely on the overrepresentation of people of color in that data and

---

[134] 539 U.S. 244 (2003).

[135] *See id.* at 253–55 (discussing the intricacies of the affirmative action program).

[136] *See supra* Part I (discussing how the algorithm can learn this association from recognizing the patterns that exist in the human data).

[137] *See, e.g.*, *Gratz*, 539 U.S. at 270 (determining that the University's policy which "automatically distribute[d] 20 points, or one-fifth of the points needed to guarantee admission, to every single 'underrepresented minority' applicant solely because of race, is not narrowly tailored to achieve educational diversity").

[138] 959 F. Supp. 2d 540, 651 (2013).

[139] *Id.*

[140] *Id.*

use that overrepresentation to engage in a form of racial profiling similar to the unwritten policy of racial profiling that the court struck down on equal protection grounds in *Floyd*.

These algorithms do not rely on race alone in determining who and where ought to be policed;[141] however, as explained in Section II.A, race need not be the only factor that a policy uses in making a decision in order for a litigant to succeed in challenging that policy on equal protection grounds, nor does the use of race need to be motivated by racial animus or ill will.[142] Racial classifications of any form are prohibited, even if those classifications are not motivated by maliciousness.[143] To establish discrimination, the law does not require that claimants "prove that race was the sole, predominant, or determinative factor in a police enforcement action . . . [n]or [that] the discrimination [was] based on 'ill will, enmity, or hostility.'"[144] Given the foregoing analysis, demonstrating that algorithms classify, at least in part, on the basis of race to produce their outputs, they should be subjected to strict scrutiny.

If a court accepts that an algorithm ought to be evaluated under strict scrutiny, a police department will likely argue that there is a compelling government interest in efficient policing and that machine learning-based predictive policing algorithms allow a police department to achieve that interest by saving them money and time in maintaining a community's safety.[145] However, if the government framed its interest in this way, it would be unlikely to persuade the court. The

---

141 *See Panel*, *supra* note 30 (explaining that none of the algorithms discussed by the panel explicitly use race as a factor in their determinations).

142 *Floyd*, 959 F. Supp. 2d at 662.

143 For example, an interest in preserving diversity in higher education is a "benign" interest that the Court would still review under strict scrutiny. *See, e.g.*, Gratz v. Bollinger, 539 U.S. 244, 269–70 (2003); *cf. id.* at 302 (Ginsburg, J., dissenting) ("The mere assertion of a laudable purpose, of course, should not immunize a race-conscious measure from judicial inspection."). A government interest in preserving the safety of a community would also trigger strict scrutiny if it relied on facial discrimination, even though it is similarly a "benign" interest.

144 Floyd v. City of New York, 959 F. Supp. 2d 540, 662 (S.D.N.Y. 2013) (quoting Ferrill v. Parker Grp., Inc., 168 F.3d 468, 473 & n.7 (11th Cir. 1999)).

145 *See, e.g.*, PALANTIR LAW ENFORCEMENT, https://www.palantir.com/solutions/law-enforcement (last visited Jan. 15, 2019) ("Palantir Law Enforcement equips officers and agents with the tools they need to easily analyze intelligence, securely collaborate on investigations, . . . and respond to crime as it happens."); *About*, PREDPOL, *supra* note 42 (claiming that PredPol aims "to help law enforcement keep communities safer by reducing victimization" by allowing officers to "effectively allocate . . . resources and prevent crime").

Supreme Court has determined that "the Constitution recognizes higher values than speed and efficiency."[146]

The government interest the Court has accepted that is most similar to community safety would be the justification that the government presented in *Korematsu v. United States*: an interest in national security.[147] The Court only very recently overturned the anticanonical result it generated in *Korematsu*.[148] However, the question of whether national security is a viable compelling interest to justify facial discrimination remains open. Nonetheless, the Court would be unlikely to allow the government to base its claim on a connection to *Korematsu* for two reasons. First, the *Korematsu* decision is viewed as one of the worst decisions in Supreme Court history,[149] and it is unlikely the Court would want to rely on it in any future opinion. Second, *Korematsu* is symptomatic of the Court's equal protection analysis during wartime. The 1944 opinion is laden with fear of the "threatened danger" of "modern warfare."[150] Because predictive policing concerns are entirely domestic and not linked to war, there is not as strong a basis to rely on security as a justification for use of predictive policing algorithms.

Assuming that community safety may be a compelling government interest, machine learning-based predictive policing algorithms are not a narrowly tailored means by which to achieve that compelling interest. Machine learning-based predictive policing algorithms are not narrowly tailored to achieve efficient community safety, because police officers are not providing "meaningful, individualized review" of the people and places that the algorithms target.[151] Like *Gratz*, where administrators removed the human element from the admissions process and automatically added twenty points for every minority applicant, algorithms remove the human element from policing and enhance the risk scores for people or neighborhoods of color without regard for individual circumstances. By indiscriminately classifying on the basis of race—in a manner which programmers and

146 Frontiero v. Richardson, 411 U.S. 677, 690 (1973) (quoting Stanley v. Illinois, 405 U.S. 645, 656 (1972)).

147 *See* 323 U.S. 214, 220 (1944).

148 *See* Trump v. Hawaii, 138 S. Ct. 2392, 2423 (2018) (indicating that the case gave the Court the "opportunity to make express what is already obvious: *Korematsu* was gravely wrong the day it was decided [and] has been overruled in the court of history").

149 *See* Jamal Greene, *The Anticanon*, 125 Harv. L. Rev. 379, 380 (2011) (identifying, among others, *Korematsu* as "consistently cited in Supreme Court opinions, in constitutional law casebooks, and at confirmation hearings as [a] prime example[ ] of weak constitutional analysis").

150 *Korematsu*, 323 U.S. at 220.

151 *See* Gratz v. Bollinger, 539 U.S. 244, 276 (2003) (O'Connor, J., concurring) (citing Regents of Univ. of Cal. v. Bakke, 438 U.S. 265 (1978)).

officers cannot comprehend because of the black-box editing process of machine learning—machine learning-based predictive policing algorithms cannot be construed as being narrowly tailored to meet the state interest. Under this analysis, claimants could succeed in challenging machine learning-based predictive policing algorithms on equal protection grounds.

C. The Difficulties in Generating an Equal Protection Claim Against Machine Learning-Based Predictive Policing Algorithms

Although the argument that machine learning-based predictive policing algorithms constitute an equal protection violation is viable in theory, there are major hurdles that stand in the way of such a claim. The first major hurdle will be making the argument that an algorithm facially discriminates, and, relatedly, attaining evidence that the algorithm facially discriminates. The second problem will be attributing the privately developed algorithm's discrimination to the state actor.

1. Arguing that the Algorithm Facially Discriminates

The strongest route for litigants is to argue that a machine learning-based predictive policing algorithm facially discriminates. At first glance, the facial discrimination route may seem more difficult to prove than a disparate impact claim. If litigants pursued the disparate impact route, they could easily find proof of a disparate impact simply by pointing to changes in policing statistics when police rely on machine learning algorithms. If litigants pursue the facial discrimination route, they will struggle to find evidence that the algorithm is directly relying on race as a variable. The algorithm is not explicitly coded at the outset to rely on race, so the source code will not be a definitive indicator that race factored into the algorithm's output. The output itself also will not be a definitive indicator. As demonstrated in Figure 1 (which is a KNN algorithm) and Figure 2 (which is an artificial neural network), a machine learning-based algorithm's output does not explicitly articulate the new variables it has come to rely on; the algorithm simply spits out a decision based on those new variables.

FIGURE 1[152]:

```
In [73]:  knn = KNeighborsRegressor()
          knn.fit(X_16, y_16)
          knn.score(X_17,y_17)

Out[73]:  0.7129174735222423
```

FIGURE 2[153]:

```
In [107]:  mlp = MLPRegressor(hidden_layer_sizes = (100,100,100,100), random_state=444)
           mlp.fit(X_16,y_16)

Out[107]:  MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
                  beta_2=0.999, early_stopping=False, epsilon=1e-08,
                  hidden_layer_sizes=(100, 100, 100, 100), learning_rate='constant',
                  learning_rate_init=0.001, max_iter=200, momentum=0.9,
                  nesterovs_momentum=True, power_t=0.5, random_state=444,
                  shuffle=True, solver='adam', tol=0.0001, validation_fraction=0.1,
                  verbose=False, warm_start=False)

In [108]:  mlp.score(X_17, y_17)

Out[108]:  0.86872151340190129
```

Given this evidentiary conundrum, why is it wise for litigants to argue that algorithms facially discriminate, rather than arguing that they are facially neutral but produce a disparate impact? Even in the face of an obscene disparate impact, courts are unwilling to recognize an equal protection violation if there is no direct and obvious evidence of discriminatory intent.[154] In many ways, the Equal Protection Clause "has been shredded" and is now "functionally dead for people of color," because of the Court's reticence to deem policies that disparately impact communities of color a violation of the Equal Protection Clause.[155] The difficulty in proving an equal protection claim in the face of the defanged Equal Protection Clause has manifested itself in cases where the Court is reviewing a facially neutral law that objectively has a disparate impact, but no obvious discriminatory intent. For instance, in *McCleskey v. Kemp*, the Court found that statistical

[152] Torres, *supra* note 22. This Figure demonstrates the input and output as a programmer would see them. As seen here, the input does not point the algorithm towards a particular variable, but simply towards the nearest neighbor. The output does not explicitly state what it has found to be the nearest neighbor. Such information remains with the algorithm but is not revealed to the human programmer.

[153] *Id.* Similar to Figure 1, Figure 2 shows the inputs and outputs of a neural network. The output here is convoluted and cannot be unpacked by the programmer simply by looking at it. Instead, the variables on which the algorithm relies stay with the algorithm itself and are not revealed to its human creator.

[154] *See, e.g.*, United States v. Clary, 34 F.3d 709, 711, 713 (8th Cir. 1994) (determining that the mandatory minimums associated with crack cocaine convictions were not a violation of the Equal Protection Clause, even though 98.2% of defendants convicted of crack cocaine charges in the Eastern District of Missouri were Black).

[155] Brando Simeo Starkey, *A Failure of the Fourth Amendment & Equal Protection's Promise: How the Equal Protection Clause Can Change Discriminatory Stop and Frisk Policies*, 18 MICH. J. RACE & L. 131, 136–37 (2012).

proof of systematic racial disparities in the administration of the death penalty, in the absence of discriminatory intent, did not implicate the Equal Protection Clause.[156] The Court determined that there was merely a correlation between race and a death penalty sentence.[157] However, in the Court's view, such a correlation did not indicate that in Warren McCleskey's case, specifically, the jury or prosecutor discriminated against him as an individual.[158] Rather, the racial discrepancy in capital sentencing was merely "an inevitable part of our criminal justice system."[159] In light of courts' reticence to accept disparate impact claims, litigants wishing to challenge machine learning-based predictive policing algorithms ought to claim they are facially discriminatory on the basis of race.

Claimants will struggle to attain evidence of this facial discrimination. To generate a viable claim that an algorithm classified on the basis of race, plaintiffs will need access to the source code of the algorithm, its inputs, and its outputs because, unlike human discrimination, there will not be any statements or written policies that indicate that the classification existed.[160] Such information can be tremendously difficult to attain for anyone outside of the private company that created the product. Private developers "often assert that details about how their tools function are trade secrets."[161] Police departments claim that their hands are also tied when it comes to

[156] 481 U.S. 279, 298 (1987) ("For this [equal protection] claim to prevail, McCleskey would have to prove that the Georgia Legislature enacted or maintained the death penalty statute *because of* an anticipated racially discriminatory effect.").

[157] *See* David Rudovsky, *Litigating Civil Rights Cases to Reform Racially Biased Criminal Justice Practices*, 39 COLUM. HUM. RTS. L. REV. 97, 99 (2007) (citing *McCleskey*, 481 U.S. at 312) (describing the Court's determination in regards to the Baldus study, the statistical study that Mr. McCleskey provided to the Court as evidence of discrimination, as showing "only a 'correlation' between the victim's race and the death penalty sentence").

[158] *See McCleskey*, 481 U.S. at 292, 297 ("[T]o prevail under the Equal Protection Clause, McCleskey must prove that the decisionmakers in *his* case acted with discriminatory purpose. . . . [T]he [statistical] study is clearly insufficient to support an inference that any of the decisionmakers in McCleskey's case acted with discriminatory purpose.").

[159] *Id.* at 312.

[160] This type of information was available and relied upon in *Floyd* to prove that the NYPD had classified on the basis of race in enacting its stop-question-and-frisk policy. *See* Floyd v. City of New York, 959 F. Supp. 2d 540, 663 ("When an officer is directed to target 'male blacks 14 to 21' for stops *in general* based on local crime suspect data—a practice that the City has defended throughout this litigation—the reference to 'blacks' is an express racial classification subject to strict scrutiny.").

[161] Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1349–50 (2018) (discussing contested ownership in the context of data-driven criminal justice technologies).

releasing information about the predictive policing algorithms that they employ.[162]

It is possible that plaintiffs could seek a subpoena to access an algorithm's source code, input data, and outputs, but, when developers and police departments claim that they are entitled to withhold these "trade secrets" from claimants, they typically refuse to "comply even with those subpoenas that seek information under a protective order and under seal."[163] In refusing to comply with subpoenas, developers and police departments assert that information regarding trade secrets is privileged.[164] Trade secret evidentiary privilege is, indeed, recognized by twenty-one states.[165] The remaining jurisdictions "recognize some common law variation of it."[166] The "general view among legislators, judges, and scholars alike is that some form of trade secret evidentiary privilege both does and should exist,"[167] making it very difficult to challenge trade secret evidentiary privilege. If a court accepts the assertion of trade secret protection and its accompanying privilege, an algorithm's input data and source code will be inaccessible to a claimant, thereby thwarting any claim that the algorithm violates the Equal Protection Clause.

However, as predictive policing algorithms draw more negative attention,[168] some developers have become more transparent about their source codes and data and have even opted to make this information accessible to the public.[169] Further, as advocates better edu-

---

162 *See id.* at 1367 (citing letters from the New York City Police Department, the Nebraska Police Department, and the Iowa Police Department, all of which are on file with Professor Wexler, in which the police departments "cite[ ] trade secrets as reason to deny open records requests").

163 *Id.* at 1350 (discussing cases invoking trade secret privilege in response to subpoenas).

164 *See id.* at 1360–62 (citing the increasing prevalence of assertions that trade secrets are privileged).

165 *Id.* at 1352 (listing twenty-one states that have "codified a trade secret privilege in their evidence rules").

166 *Id.*

167 *Id.* (noting however, that the consensus is limited to civil proceedings).

168 *See, e.g.*, Alisha Jarwala, *Minority Report: Why We Should Question Predictive Policing*, HARV. C.R.-C.L. L. REV. AMICUS BLOG (Mar. 7, 2018), http://harvardcrcl.org/minority-report-why-we-should-question-predictive-policing (exemplifying one of the many pieces that have analogized these technologies to the 2002 dystopian film *Minority Report*).

169 For example, the source code for CivicScape and HunchLab, two predictive policing technologies, is available to the public. *See* Andrew Guthrie Ferguson, *Policing Risk: Predicting the Litigation Risk in Predictive Policing Tech*, HUFFINGTON POST (July 11, 2017), https://www.huffingtonpost.com/entry/policing-risk-predicting-the-litigation-risk-in-predictive_us_5965377fe4b0911162fc2f83 (discussing CivicScape's "embrace [of] uber-transparency" and its release of its code and strategy). In addition to embracing transparency, some firms seek to use their technology for other social ends. *See* Maurice

cate judges about machine learning-based predictive policing algorithms, judges may be more stringent in evaluating what falls under the protection of trade secret privilege. For instance, advocates recently succeeded in obtaining a court order to compel the New York City Police Department to produce historical output data from any existing predictive policing systems used in New York City as well as the NYPD's correspondence with Palantir, a known partner in NYPD's predictive policing program.[170] The New York state trial court stated that such disclosures are "premised on the public's inherent right to know . . . and [are] intended to expose government abuses and hold it accountable."[171] Thus, it ordered the disclosures, given the absence of "expert evidence that the disclosure of the output data . . . would jeopardize the NYPD's capacity to guarantee the security of its information technology assets. . . ."[172]

Even if litigants are able to get their hands on the source code, inputs, and outputs, a court still may not be convinced that facial discrimination is present. No source code is originally coded to rely on race. Instead, as explained in Part I, machine learning algorithms develop the ability to discern patterns in the data and incorporate those patterns into their decisionmaking.[173] However, the algorithms do not make clear what new variables they are relying on based on

---

Chammah, *Policing the Future*, MARSHALL PROJECT (Feb. 3, 2016), https://www.themarshallproject.org/2016/02/03/policing-the-future (calling HunchLab's rhetoric "civic-minded" and discussing the other progressive projects that its developer is involved with, including "tools to analyze legislative districts, as well as an app that helps city residents map the locations of trees in order to study their environmental impact"). PredPol's algorithm was made public for a short period of time after an attack by a "hacktivist" group. *See* Lee Johnstone, *Police Crime Prediction Software, PredPol Source Code Leaked by Anonymous*, CYBER WAR NEWS (Mar. 21, 2013), https://www.cyberwarnews.info/2013/03/21/police-crime-prediction-software-predpol-source-code-leaked-by-anonymous (reporting on the leak and the firm's response). This leak enabled the HRDAG team to perform the study referenced in Section II.B of this Note.

170 *See* Rachel Levinson-Waldman & Erica Posey, *Court: Public Deserves to Know How NYPD Uses Predictive Policing Software*, BRENNAN CTR. FOR JUST. (Jan. 26, 2018), https://www.brennancenter.org/blog/court-rejects-nypd-attempts-shield-predictive-policing-disclosure (discussing the documents requested by the Center and ordered released by the court). Although the Brennan Center withdrew its request for access to the source code prior to the judge's order, the same logic requiring release of such output data may apply to future cases. However, to avoid trade secret protections, future litigants should learn from the Center's strategic approach. *See id.* (noting that "the Brennan Center narrowed its request to exclude the algorithm itself").

171 Brennan Ctr. for Justice v. New York City Police Dep't, No. 160541/2016, slip op. at 9 (Sup. Ct. N.Y. Cty. Dec. 27, 2017) (citing NLRB v. Robbins Tire & Rubber Co., 437 U.S. 214, 242 (1978) and Matter of Fink v. Lefkowitz, 47 N.Y.2d 567, 571 (1979)), http://www.nycourts.gov/reporter/pdfs/2017/2017_32716.pdf (discussing the rationale of the Freedom of Information Law).

172 *Id.* at 13.

173 *See supra* notes 21–37 and accompanying text.

those patterns.[174] For instance, when an algorithm learns to associate race and criminality, the algorithm's output does not bluntly articulate "I am now relying on race." Instead, as is shown in Figures 1 and 2, it simply generates an output based on its ever-adapting understanding of new patterns. Litigants will have to ask the court to unpack the "black-box" of algorithmic decisionmaking. They will need to argue that, based on the source code, the data inputs, and the algorithm's outputs, the algorithm must be facially discriminating because historical crime data and dragnet data searches are racist, thereby making the algorithm rely on race. The algorithm relies on race in a manner that is unlike a possibly racist human whose intentions are unknown in that the algorithm's mind is only comprised of what it is exposed to through dragnet data searches or historical crime data. Unlike a free-thinking human who could (one would hope) choose to not engage in racist thinking in a particular decision, the machine has no choice. The algorithm's mind is only populated with racist data, and for that reason every output it generates will be informed, at least in part, by using race as a variable.

If a litigant fails to demonstrate that the algorithm facially discriminates—either because the litigant cannot obtain evidence to show that an algorithm explicitly relied on race or the court is not willing to unpack the "black-box" and find that an algorithm relies on race as a variable and facially discriminates—a litigant may not be able to make out a claim of facial discrimination. Instead, she will be left with a disparate impact challenge, arguing that the algorithm is a neutral policy that leads to a disparate impact and discriminates *because of* that disparate impact rather than in spite of it. If litigants tried to levy a disparate impact challenge against machine learning-based predictive policing algorithms, there is a chance their claim would meet the same dismal fate met by the litigants in *McCleskey*.[175] However, making a disparate impact argument will not necessarily debilitate claimants' success. Unlike a prosecutor or set of jurors who, according to the Court, only speculatively holds the prejudicial views borne out in the statistics that Mr. McCleskey and his counsel presented, predictive policing algorithms actually rely on racist statistics and racist data inputs in producing their outputs and make decisions *based on* individuals' race and neighborhoods' racial composition.[176] The disparate impacts created by racially biased policing, prosecuting, and policymaking are no longer viable subjects

[174] *See supra* notes 34–35.

[175] *See supra* notes 154–59 and accompanying text.

[176] *See supra* Section I.C (discussing the racialized implications for algorithms of biased historical crime data and dragnet data searches).

of an equal protection claim. Instead, the disparate treatment of communities of color is now the input on which machine learning-based predictive policing algorithms rely. It is for this reason that claimants can argue that these algorithms constitute a neutral policy that creates a disparate impact and does so *because of* its disparate impact on a particular group, rather than in spite of it, and can evade the dismissal that would likely meet a disparate impact challenge.

2. *Attributing Algorithms' Facially Discriminatory Race-Based Classification to State Actors*

An equal protection claim against a machine learning-based predictive policing algorithm may fail because claimants must demonstrate that a state actor, not a private actor, adopted a discriminatory policy.[177] The federal government can only regulate a state actor under the Fourteenth Amendment.[178] Private contractors typically develop machine learning-based predictive policing algorithms and sell them to police departments,[179] so a claimant might face difficulty in attributing an algorithm's discriminatory facial race-based classification to the state actor: the police department.

Plaintiffs could argue, however, that the privately developed algorithm's facially discriminatory race-based classification ought to be attributed to the police department because of the "pervasive entwinement" between the algorithm's classification and the police

177 *See generally Developments in the Law – State Action and the Public/Private Distinction*, 123 HARV. L. REV. 1248, 1250 (2010) [hereinafter *State Action*] (examining how the state action doctrine is applied such that there is a "bright-line" rule between state action and private action).

178 *See id.* at 1256–57 (stating that absent some state action, "section 5 of the Fourteenth Amendment confers no authority on Congress to regulate individual conduct"). As the law review piece highlights, during the twentieth century, courts "expanded the concept of state action, stretching it to cover a wide spectrum of government involvement." *Id.* at 1251; *see, e.g.*, Burton v. Wilmington Parking Auth., 365 U.S. 715, 716–17 (1961) (finding that equal protection applied to a privately leased restaurant in a publicly owned and operated garage); Shelley v. Kraemer, 334 U.S. 1 (1948) (reviewing and deeming unconstitutional the judicial enforcement by state courts of a community's privately developed, racially restrictive covenant). However, the Rehnquist Court stringently pushed back on such an expansion, attempting to clearly demarcate the line between state and private action. *See State Action*, *supra* note 177, at 1251; *see, e.g.*, United States v. Morrison, 529 U.S. 598, 621 (2000) ("[T]he action inhibited by the first section of the Fourteenth Amendment is only such action as may fairly be said to be that of the States. That Amendment erects no shield against merely private conduct, however discriminatory or wrongful." (quoting *Shelley*, 334 U.S. at 13)).

179 Machine learning-based predictive policing algorithms like PredPol and Beware are all developed and owned by private companies. *See* DAVID ROBINSON & LOGAN KOEPKE, UPTURN, STUCK IN A PATTERN 3–4 (2016), https://www.teamupturn.org/static/reports/2016/stuck-in-a-pattern/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf (describing the origins of each known predictive policing algorithm).

department's subsequent actions in reliance on the algorithm's outputs.[180] In *Brentwood Academy v. Tennessee Secondary School Athletic Ass'n*, the seminal case on the intertwining of private and state action, the Court recognized state action may exist when there is "such a 'close nexus between the State and the challenged action' that seemingly private behavior 'may be fairly treated as that of the State itself.'"[181]

In the case of a privately developed predictive policing algorithm, there is a close nexus between the action of the state and the action of the private entity. Although the algorithm generates the discriminatory output, it is the police officer who acts on the algorithm's discriminatory output and makes it a reality. In this way, the logic undergirding *Brentwood Academy* is analytically analogous to the logic of attributing a privately developed algorithm's discrimination to a police officer.[182] In *Brentwood Academy*, the Tennessee Secondary School Athletic Association was a not-for-profit corporation organized to regulate interscholastic sports among private and public high schools.[183] Public school officials were intimately involved in its regulation.[184] Because of state actors' part in the Association, the Court determined that the "nominally private character of the Association [was] overborne by the pervasive entwinement of public institutions and public officials in its composition and workings, and [thus] there [was] no substantial reason to claim unfairness in applying constitutional standards to it."[185] Similarly, predictive policing algorithms are "overborne by the pervasive entwinement of public institutions,"[186] because, without the state actor, the "will" of the algorithm would never become a reality. Without the state agent, the algorithm's output would be lifeless. With the state agent, the algorithm's output becomes the arbiter of who is policed, and where.

Despite this close nexus, there is a difference between *Brentwood Academy* and a police department's use of a predictive policing algorithm. In *Brentwood Academy*, the public officials, "overwhelmingly perform[ed] all but the purely ministerial acts by which the

[180] *See generally* Brentwood Acad. v. Tenn. Secondary Sch. Athletic Ass'n, 531 U.S. 288, 291 (2001) (establishing the pervasive entwinement standard).

[181] *Id.* at 295 (quoting Jackson v. Metro. Edison Co., 419 U.S. 345, 351 (1974)) (noting that the inquiry is a context specific one).

[182] *Id.* at 291.

[183] *See id.* at 291–93 (outlining the facts of the case).

[184] *See id.* (noting that the Association was designated by the State Board of Education as the organization in charge of supervising interscholastic athletics).

[185] *Id.* at 298 (noting that the Association's members were primarily public schools and its board was comprised primarily of public school officials).

[186] *Id.*

Association exist[ed] and function[ed] . . . ."[187] When an officer uses an algorithm's output, he does not have a direct role in coding the algorithm and thereby generating the algorithm's output. The police department could attempt to argue that the officer's role in relying on the algorithm's output is akin to him relying on a racist resident who calls the police when she claims she sees a Black man behaving "suspiciously" in her neighborhood. The Supreme Court would probably not attribute the racist neighbor's discrimination to the officer just because the officer relied on her description.[188]

To combat this argument, litigants could highlight the unique entanglement between predictive policing algorithms and police officers. Predictive policing algorithms involve the state actor at both the front and the back end of the process. Unlike a free-thinking, racist neighbor, programmers create predictive policing algorithms specifically for a state actor and feed those algorithms data that is generated by that very same state actor. In this way, machine learning-based predictive policing algorithms are even more entwined with the state actor than the Association in *Brentwood Academy*.

Conclusion

The purpose of this Note has been to highlight the facially discriminatory nature of machine learning-based predictive policing algorithms and the potential for equal protection claims against the government for relying on such algorithms. As the preceding analysis suggests, machine learning endows an algorithm with the ability to learn, mimic, and refine patterns that exist in the real world. In the context of policing, machine learning allows an algorithm to associate race and criminality, and thereby discriminate via race-based facial classifications. The Equal Protection Clause is the obvious remedy for facial discrimination. However, claimants will face significant barriers to success because of the difficulties of attributing private action to state actors and the difficulties of gathering proof of the algorithms' classifications on the basis of race.

187 *Id.* at 300 (discussing the "integral" role played by public officials in the Association).

188 For example, in *Brown v. City of Oneonta*, 221 F.3d 329, 334 (2d Cir. 2000), an elderly woman who was robbed saw only a Black person's hand, which had a cut on it. She gave the police a description for the purposes of tracking down the suspect. *Id.* She did not give any description of the person's face or body. *Id.* The police subsequently stopped over two hundred Black men based on the description. *Id.* The court determined that the police department's "policy was to investigate crimes by interviewing the victim, getting a description of the assailant, and seeking out persons who matched that description," which is "race-neutral on its face." *Id.* at 337.

If courts view these barriers as insurmountable, they will render algorithms immune from equal protection review and will (yet again) fail to deliver on the promise of the Equal Protection Clause. The slow deterioration of the Equal Protection Clause has occurred because "the presumption—despite staggering evidence—seems now to be that nondiscrimination is the norm . . . ."[189] However, if the courts permit this presumption to eviscerate equal protection challenges to algorithms, a burgeoning number of state policies will be deemed unreviewable, given that government reliance on algorithms is becoming more pervasive across the board, including in decisions of who gets access to healthcare,[190] who teaches American children,[191] and who receives loans.[192]

As municipalities begin to rely on black-box artificial intelligence to determine where to dispatch police officers, American jurisprudence must navigate ways to hold those machines accountable. Trust in police is already at a low ebb in the United States.[193] If systematic racism is not just perpetuated but exacerbated by the facially discriminatory nature of machine learning, trust in police might be eroded

---

189 Derrick A. Bell, Race, Racism and American Law 185 (6th ed. 2008).

190 *See* Colin Lecher, *What Happens When an Algorithm Cuts Your Health Care*, Verge (Mar. 21, 2018), https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy (describing the story of Tammy Dobbs, a woman whose Medicaid was cut because of an algorithmic output).

191 *See* Alvin Roth, *Why New York City's High School Admissions Process Only Works Most of the Time*, Chalkbeat (July 2, 2015), https://www.chalkbeat.org/posts/ny/2015/07/02/why-new-york-citys-high-school-admissions-process-only-works-most-of-the-time (describing New York City's use of the deferred acceptance algorithm in high school admissions since 2003).

192 *See* Kaveh Waddell, *How Algorithms Can Bring Down Minorities' Credit Scores*, Atlantic (Dec. 2, 2016), https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333 (examining algorithmic decisions in generating credit scores and how this impacts minorities' abilities to attain loans from the government).

193 *See* Jeffrey M. Jones, *In U.S., Confidence in Police Lowest in 22 Years*, Gallup (June 19, 2015), http://news.gallup.com/poll/183704/confidence-police-lowest-years.aspx (noting the record number of people who had very little or no confidence in the police). In part, this low confidence can be attributed to the increased publicity surrounding police brutality and killings of Black Americans. *See, e.g.*, *Freddie Gray's Death in Police Custody – What We Know*, BBC (May 23, 2016), http://www.bbc.com/news/world-us-canada-32400497 (discussing the public backlash after Freddie Gray's death in police custody); Jasmine C. Lee & Haeyoun Park, *15 Black Lives Ended in Confrontations with Police. 3 Officers Convicted.*, N.Y. Times (Oct. 5, 2018), https://www.nytimes.com/interactive/2017/05/17/us/black-deaths-police.html (discussing the low conviction rates for officers who killed Black men); Eliott C. McLaughlin, *What We Know About Michael Brown's Shooting*, CNN (Aug. 15, 2014), https://www.cnn.com/2014/08/11/us/missouri-ferguson-michael-brown-what-we-know/index.html (discussing protests following Michael Brown's shooting); Debbie Nathan, *What Happened to Sandra Bland?*, Nation (Apr. 21, 2016), https://www.thenation.com/article/what-happened-to-sandra-bland (discussing the mysterious death of Sandra Bland in police custody).

entirely, particularly for communities of color. Further, this generation will be yet another in the history of the United States that has maintained racial discrimination in the criminal justice system. If police departments employ these algorithms for a sustained period of time, the algorithms' feedback loops could exacerbate disparate policing practices in the United States. If the Equal Protection Clause is truly meant to ensure that no state denies to "any person within its jurisdiction the equal protection of the laws,"[194] then the Equal Protection Clause should protect against the facial discrimination of machine learning-based predictive policing algorithms.

194 U.S. Const. amend. XIV, § 1.