THE SOLUTION TO THE PERVASIVE BIAS AND DISCRIMINATION IN THE CRIMINAL JUSTICE SYSTEM: TRANSPARENT AND FAIR ARTIFICIAL INTELLIGENCE

Mirko Bagaric*, Jennifer Svilar**, Melissa Bull***, Dan Hunter****, and Nigel Stobbs*****

Abstract

Algorithms are increasingly used in the criminal justice system for a range of important matters, including determining the sentence that should be imposed on offenders; whether offenders should be released early from prison; and the locations where police should patrol. The use of algorithms in this domain has been severely criticized on a number of grounds, including that they are inaccurate and discriminate against minority groups. Algorithms are used widely in relation to many other social endeavors, including flying planes and assessing eligibility for loans and insurance. In fact, most people regularly use algorithms in their day-to-day lives. Google Maps is an algorithm, as are Siri, weather forecasts, and automatic pilots. The criminal justice system is one of the few human activities which has not substantially embraced the use of algorithms. This Article explains why the criticisms that have been leveled against the use of algorithms in the criminal justice domain are flawed. The manner in which algorithms operate is generally misunderstood. Algorithms are not autonomous machine applications or processes. Instead, they are developed and programmed by people and their efficacy is determined by the quality of the design process. Intelligently designed algorithms can replicate human cognitive processing, but they have a number of advantages, including the speed at which they process information. Also, because they do not have feelings, they are more objective and predictable than people in their decision-making. They are a core component of overcoming the pervasive bias and discrimination that exists in the criminal justice system.

* Dean of Law, Swinburne Law School, Melbourne.

** J.D., University of Tennessee College of Law. Jennifer is an attorney in the Commercial Litigation Group at Butler Snow LLP.

*** Professor, Law School, Queensland University of Technology Law School.

**** Dean, Faculty of Law, Queensland University of Technology.

***** Senior Lecturer, Law School, Queensland University of Technology.

INTRODUCTION

Algorithms are increasingly being used in the criminal justice system. They are used by police to predict where the next crime will happen and by courts to determine whether a defendant will commit a future crime. Pretrial risk assessment algorithms evaluate whether a criminal defendant poses a threat to public safety or will fail to show up to court for bail proceedings.[1] After adjudication, algorithms are relied on to predict whether a defendant will recidivate, which is an important consideration in sentencing determinations and parole decisions.[2] Risk scores provided by these tools may also be used during incarceration to determine what level of security a prisoner requires.[3] In all of these scenarios, risk assessment tools are used because they provide the objectivity and efficiency that humans cannot,[4] and

1. Sarah Brayne & Angèle Christin, *Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts*, 68 SOC. PROBS. 608, 611 (2021).

2. *Id.*

3. *Id.*

4. *Id.* at 615.

proponents believe that bringing objective regularity to the criminal justice system will lead to more accountability and transparency for decisions that impact people's lives.[5]

Today, more than sixty kinds of risk assessment tools are being used in the criminal justice system[6] because of a "focus toward evidence-based practices."[7] These systems, in theory, could lead to more uniformity in criminal sentencing, as "[m]odels and algorithms can process and review vast amounts of data in order to identify and trace these factors and signify their weight and force in causing criminal behavior."[8] Some risk assessment tools rely on machine learning algorithms, which "allow the machines to train themselves on how to best process the data."[9]

The use of algorithms in the criminal justice domain is, however, heavily criticized. The fact that these systems can rewrite their own processing structure indicates that "they might rely on assumptions about relationships between different categories of data that may remain hidden even to the systems' designers," creating a "black box."[10] Another key criticism is that they discriminate against already disadvantaged groups. A new coalition has formed[11] to argue that algorithms are neither accurate nor objective, and the burden of these shortcomings "is disproportionately borne by historically marginalized groups . . . ."[12] A 2019 survey showed that "nearly three in five Americans believe algorithms make bias inevitable."[13] According to the same survey, 56% of Americans disapproved of using an algorithm in decisions regarding parole.[14]

This is a remarkable observation given that it has been established that Black Americans have for decades been discriminated against by human decision-makers in the form of police officers, prosecutors, and judges.[15] Human decision-making

5. Matt Henry, *Risk Assessment: Explained*, APPEAL (Mar. 25, 2019), https://theappeal.org/risk-assessment-explained/.

6. Kia Rahnama, *Science and Ethics of Algorithms in the Courtroom*, 2019 J.L., TECH. & POL'Y 170, 174 (citing Alyssa M. Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L. REV. 303, 309 (2017)).

7. *Id.* at 173 (citing Carlson, *supra* note 6, at 305).

8. *Id.* (citing Carlson, *supra* note 6, at 309).

9. *Id.* at 174.

10. *Id.* (citing Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi, *The Ethics of Algorithms: Mapping the Debate*, BIG DATA & SOC'Y, July–Dec. 2016, at 1, 6).

11. According to Barabas, there is "an influential community of researchers from both academia and industry who have formed a new regulatory science under the rubric of 'fair, accountable, and transparent algorithms' ('FAccT algorithms')." Chelsea Barabas, *Beyond Bias: Re-Imagining the Terms of "Ethical AI" in Criminal Law*, 12 GEO. J. L. & MOD. CRITICAL RACE PERSP. 83, 96 (2020).

12. *Id.*

13. Jens Ludwig & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, BOS. GLOBE (Sept. 24, 2019, 5:00 AM), https://www.bostonglobe.com/opinion/2019/09/24/discrimination-age-algorithms/mfWUxRH8Odm6IRo3PZRLdI/story.html (citing Aaron Smith, *Public Attitudes Toward Computer Algorithms*, PEW RSCH. CTR. (Nov. 16, 2018), https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/11/PI_2018.11.19_algorithms_FINAL.pdf).

14. Smith, *supra* note 13, at 3.

15. *See infra* Part I.

in the context of the criminal justice system has resulted in appalling levels of mistreatment of disadvantaged groups. As noted by Jens Ludwig and Cass R. Sunstein, eliminating algorithms—or reducing their use—would cause "more, not less, discrimination."[16] Compared to algorithms, judges often "detain far more people than needed to achieve a given reduction in crime."[17] Algorithms, on the other hand, can reduce bias and discrimination by allowing us to detect and resolve these flaws.[18] The bias present in algorithms presents less of an obstacle than human bias because it can more easily "be observed, studied, and corrected in ways that human bias cannot."[19] Although it is helpful that algorithms display their biases in a way that we can observe and correct, "'one cannot expect any . . . tool to reverse centuries of racial injustice or gender inequality' in the criminal legal system."[20]

The argument that algorithms in the criminal justice system will lead to new problems in the form of unfairly targeting certain groups represents a deep misunderstanding of the nature of algorithms and the current workings of the system. In this Article, we refute the key criticisms of algorithms in the criminal justice system, and we make recommendations regarding how algorithms can be designed to operate optimally in this domain.

Previously, we have noted that human beings have an unjustified preference for human decision-making over computerized decision-making. This is a well-established attitude termed "algorithmic aversion."[21] In many areas, people have overcome this distrust and routinely use algorithmic processes to plan and coordinate their activities. This has not transferred to the criminal justice landscape.[22] Stakeholders in this domain are often resistant to the adoption of algorithms to make decisions that impact offenders. This is unfortunate. Algorithms synthesize and process information far more quickly and accurately than humans, and their outcomes are not distorted by emotional or subjective preferences.

The main reason that algorithms are not widely accepted in the criminal justice system is because stakeholders are not aware of the manner in which they operate. This Article seeks to overcome this knowledge deficit. Thus, we describe how algorithms in the criminal justice arena operate. We do this, in part, by comparing them to several other algorithms.

---

16. Ludwig & Sunstein, *supra* note 13.

17. *Id.*

18. *See id. But see* Sarah A. Seo, *What Cars Can Teach Us About New Policing Technologies*, N.Y. REV. (Oct. 12, 2019, 7:00 AM), https://www.nybooks.com/daily/2019/10/12/what-cars-can-teach-us-about-new-policing-technologies/ ("Once embedded in computer programs, these biases become much more difficult to challenge than those of human police, prosecutors, and judges.").

19. Henry, *supra* note 5.

20. Barabas, *supra* note 11, at 97 (quoting Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns & Aaron Roth, *Fairness in Criminal Justice Risk Assessments: The State of the Art*, 50 SOC. METHODS & RSCH. 3, 35 (2018)).

21. *See* Berkeley J. Dietvorst, Joseph P. Simmons & Cade Massey, *Algorithmic Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCH.: GEN. 114, 114 (2015).

22. *See infra* Part III.

The first is in relation to algorithms that millions of people use daily to plan and coordinate their lives. There are thousands of algorithms that we could have chosen to demonstrate this point, but we have chosen Google Maps because most readers are likely to be familiar with it. It also has many similar design features to common criminal justice algorithms. The analogy between Google Maps and the criminal justice system could be criticized because Google Maps does not involve predictions of future behavior and the decisions, arguably, do not affect important human rights. To negate this criticism, we analyze the workings of two other commonly used algorithms: weather predictions and airline automatic pilots. Weather forecasts involve complex predictions of future events. Automatic pilot instructions also are influenced by predictions of future events and, more importantly, involve life and death decisions; hence, it would be untenable to argue that algorithms cannot or should not be utilized regarding activities that affect important human rights or interests.

Another key pillar of our argument that (well-designed) algorithms should be used more extensively in the criminal justice system is that the alternative, in the form of human decision-making, has led to gross errors, inaccuracies, and mistreatment of minority groups. At the margins, algorithms have some scope for improvement, but they are immeasurably superior to the current outcomes achieved through decisions made by courts and other criminal justice institutions, such as parole boards.

In Part I of this Article, we provide an overview of the existing problem with decision-making in the criminal justice system. This is an important discussion because it transpires that major criticisms of algorithmic decision-making already exist far more acutely in the criminal justice landscape. This is followed by a brief discussion in Part II of the nature of Artificial Intelligence ("AI") systems, including an explanation of the manner in which often-used and widely-accepted AI systems operate. This is necessary in order to familiarize people engaged in the criminal justice system with the relevant technology, and it is also an essential step to acceptance of this technology. In Part III, we provide an overview of the existing use of AI in the criminal justice system. An analysis of the criticisms and advantages of these algorithms is undertaken in Part IV. The reform recommendations are set out in Part V and summarized in the concluding remarks.

I. CURRENT PROBLEMS WITH HUMAN-DECISION MAKING IN THE CRIMINAL JUSTICE SYSTEM

The contrast to algorithmic decision-making is unstructured human judgements. This remains the conventional form of decision-making in all aspects of the criminal justice system—from the policing phase to decisions relating to whether an offender should be released on parole. Hence, it is desirable to examine more carefully some of the drawbacks of this approach. Especially because, as we shall see, the drawbacks are of the same nature as the criticisms leveled against AI risk assessment decision-making. In addition to the fact that human decision-making is poor at determining reoffending, there are numerous other problems associated with this approach.

The principle of equality before the law is widely accepted as a fundamental tenet of justice. While sentencing law and criminal law do not expressly target or discriminate against certain groups, much critical research has demonstrated that, in practice, sentencing systems operate in discriminatory ways. This is evidenced in the over-representation of certain racial groups at every level of the criminal justice system: in police stops, in bail applications, and in the harshness of sentences.[23] Law disproportionality targets, for example, racial minorities and Black people in particular. Elizabeth Hinton and DeAnza Cook, in a recent historical review of the mass criminalization of Black Americans, note:

> From the late nineteenth century onward, the high rates of arrest and incarceration within African American communities served to create what historian Khalil Muhammad has called a "statistical discourse" about black crime in the popular and political imagination. Reinforced by data, this discourse cast black people as a uniquely dangerous and lawbreaking group and justified the perpetual expansion of the American prison system, sustained harsh sentencing practices, informed decisions surrounding capital punishment, and sanctioned racial profiling in general. In cities like New York and Chicago, local law enforcement policies and policing practices further strengthened common associations between black people and criminality by routing illegal activities and informal economies to police-patrolled vice districts in black neighborhoods . . . Considered an objective truth and a statistically irrefutable fact, notions of black criminality justified both structural and everyday racism.[24]

In many American states, urban, densely-populated, and largely Black neighborhoods are the target of more focused attention from law enforcement agencies than predominantly white, wealthy neighborhoods.[25]

As a consequence, police process more crimes, especially minor offenses, committed by racial minorities than crimes committed by white people.[26] Research has repeatedly demonstrated that police are more inclined to arrest suspects from racial

---

23. Margaret Bull Kovera, *Racial Disparities in the Criminal Justice System: Prevalence, Causes, and a Search for Solutions*, 75 J. SOC. ISSUES 1139, 1140 (2019); JACK GLASER, SUSPECT RACE: CAUSES AND CONSEQUENCES OF RACIAL PROFILING 8 (2015).

24. Elizabeth Hinton & DeAnza Cook, *The Mass Criminalization of Black Americans: A Historical Overview*, 4 ANN. REV. CRIMINOLOGY 261, 269–270 (2021).

25. K. Babe Howell, *Prosecutorial Discretion and the Duty to Seek Justice in an Overburdened Criminal Justice System*, 27 GEO. J. LEGAL ETHICS 285, 286, 290–91, 297–98 (2014); Kim Farbota, *Black Crime Rates: What Happens when Numbers Aren't Neutral*, HUFFINGTON POST (Sept. 2, 2015), http://www.huffingtonpost.com/kim-farbota/black-crime-rates-your-st_b_8078586.html.

26. Howell, *supra* note 25, at 286, 290–91; Rsch. Working Grp. & Task Force on Race and the Crim. Just. Sys., *Preliminary Report on Race and Washington's Criminal Justice System*, 35 SEATTLE U. L. REV. 623, 636–38 (2012).

minority groups than white suspects.[27] A recent report by the Bureau of Justice Statistics, for example, demonstrates that Black people were overrepresented by 30% as persons arrested for nonfatal violent crimes relative to their population size.[28] The over-policing of minority groups is the case even in an environment where decision-makers, including police officers, strenuously reject the use of group-based stereotypes to make judgements that affect others. Spencer, Charbonneau, and Glaser, argue that an important explanation lies in the implicit biases that operate outside of conscious awareness and control, but nevertheless influence behavior. Police often operate under conditions where uncertainty, high discretion, stress, and threats cause them to unconsciously rely on mental shortcuts and pernicious, but pervasive, stereotypes linking Black and Latinx people with violence, crime, and weapons. This response may occur even when officers disavow the use of category-based (e.g., racial) biases in deciding who to investigate or how to investigate them.[29]

Evidence also indicates a similar bias in prosecutorial actions. Prosecutors have largely unregulated and unreviewable discretion about whether to charge those arrested with crimes, as well as how many and which offenses to charge.[30] Mirroring disparities in policing, research shows that prosecutors are more likely to file and proceed with charges against Black suspects than white suspects, even where their criminal records are identical.[31] The explanations for this trend vary. Some argue that prosecutors prefer to avoid offending police with whom they must work,[32] or that they are time poor and lack resources to analyze relevant evidence.[33] Another possible explanation is that implicit racial bias influences

27. Task Force on Race and the Crim. Just. Sys., *supra* note 26, at 642 (quoting Tammy Rinehart Kochel, David B. Wilson & Stephen D. Mastrofski, *Effect of Suspect Race on Officers' Arrest Decisions*, 49 CRIMINOLOGY 473, 475 (2011)).

28. ALLEN J. BECK, RACE AND ETHNICITY OF VIOLENT CRIME OFFENDERS AND ARRESTEES, 2018 1 (2021), https://bjs.ojp.gov/content/pub/pdf/revcoa18.pdf; *see also*, Trevor Shoels, *The Color of Collateral Damage: The Mutilating Impact of Collateral Consequences on the Black Community and the Myth of Informed Consent*, 21 J. L. & Soc. Deviance 194, 223–226 (describing the disproportionate arrests of Black people for marijuana possession); Marisa Omori & Nick Petersen, *Institutionalizing Inequality in the Courts: Decomposing Racial and Ethnic Disparities in Detention, Conviction, and Sentencing*, 58 CRIMINOLOGY 678, 694 (2020) (noting Black people have a higher probability of detention).

29. Bull Kovera, *supra* note 23, at 1143; Katherine B. Spencer, Amanda K. Charbonneau, & Jack Glaser, *Implicit Bias and Policing*, 10/1 SOC. & PERSONALITY PSYCH. COMPASS 50, 51 (2016).

30. Angela J. Davis, *Racial Fairness in the Criminal Justice System: The Role of the Prosecutor*, 39 COLUM. HUM. RTS. L. REV. 202, 205–06 (2008); Robert J. Smith & Justin D. Levinson, *The Impact of Implicit Racial Bias on the Exercise of Prosecutorial Discretion*, 35 SEATTLE U. L. REV. 795, 806–08 (2012).

31. Task Force on Race and the Crim. Just. Sys., *supra* note 26, at 647 (quoting ROBERT D. CRUTCHFIELD, JOSEPH G. WEIS, RODNEY L. ENGEN & RANDY R. GAINEY, WASH. STATE MINORITY & JUSTICE COMM'N, RACIAL AND ETHNIC DISPARITIES IN THE PROSECUTION OF FELONY CASES IN KING COUNTY 4 (1995), http://www.courts.wa.gov/committee/pdf/November%201995%20Report.pdf); Smith & Levinson, *supra* note 30, at 806 (citing Michael L. Radelet & Glenn L. Pierce, *Race and Prosecutorial Discretion in Homicide Cases*, 19 LAW & SOC'Y REV. 587, 587 (1985)) (demonstrating that prosecutors charged White and Black defendants differently in homicide cases in Florida).

32. Howell, *supra* note 25, at 312.

33. *Id.* at 313.

prosecutors' exercise of discretion and results in the prosecution of a higher proportion of Black people.[34] Scholarship demonstrates how many Americans unconsciously assume that Black people are dangerous and associate them with criminality.[35] As members of the American population, prosecutors and police are susceptible to this bias, and they could be more inclined to charge Black suspects with violent offenses.[36] For the same reason, they might be more likely to deem offenses as more serious and proceed with charges if the suspect is Black and the victim is white than vice-versa.[37] Moreover, more rigorous policing and higher rates of arrests amongst Black people mean they are more likely to have prior convictions than white people, which makes it easier for prosecutors to substantiate charges at trial.[38]

This pattern of discrimination in relation to people from racial minorities is also reproduced in the courts and at sentencing, where they are more likely to be convicted and receive harsher sanctions than white people for the same offenses.[39] The empirical research is damning. Although Black people constitute only 13% of the U.S. population, they represent over 30% of those imprisoned, with one in four Black men incarcerated at some point in their lives.[40] Black people are imprisoned at more than five times the rate at which white Americans are imprisoned,[41] and they receive longer prison terms than white offenders who have committed the same crimes and have identical criminal histories.[42] A study undertaken in the federal jurisdiction between 2005 and 2012 found that judges imposed longer prison sentences on Black men than on white offenders convicted of similar offenses.[43]

---

34. Angela J. Davis, *Prosecution and Race: The Power and Privilege of Discretion*, 67 FORDHAM L. REV. 13, 34–35 (1998) [hereinafter Davis, *Prosecution and Race*]; Kristin N. Henning, *Criminalizing Normal Adolescent Behavior in Communities of Color: The Role of Prosecutors in Juvenile Justice Reform,* 98 CORNELL L. REV. 383, 429 (2013).

35. Smith & Levinson, *supra* note 30, at 798; Task Force on Race and the Crim. Just. Sys., *supra* note 26, at 665-66.

36. *See* Smith & Levinson, *supra* note 30, at 808 (explaining prosecutors are inclined to believe that Black suspects acted more violently than white suspects).

37. Davis, *Prosecution and Race*, *supra* note 34, at 35.

38. *Id.* at 36–37.

39. Bull Kovera, *supra* note 23, at 1145; Rose Matsui Ochi, *Racial Discrimination in Criminal Sentencing*, *in* CONTINUING THE STRUGGLE FOR JUSTICE: 100 YEARS OF THE NATIONAL COUNCIL ON CRIME AND DELINQUENCY 193, 193–195 (Barry Krisberg et al. eds., 2007).

40. Bull Kovera, *supra* note 23, at 1145; JENNIFER BRONSON & E. ANN CARSON, Bureau of Just. Stat., PRISONERS IN 2017 1, 17 (2019), https://bjs.ojp.gov/content/pub/pdf/p17.pdf.

41. Heather C. West, William J. Sabol & Sarah J. Greenman, Bureau of Just. Stat., Prisoners in 2009 1, 9 (2010). [hereinafter BUREAU OF JUST. STAT.. 2009]; *see also* Prison Reform Tr., *Prison Factfile,* BROMLEY BRIEFINGS, June 2012, at 1, 35, http://www.antoniocasella.eu/nume/Bromley_briefings_Factfile_June2012.pdf (showing that the over-representation of racial minorities in the United Kingdom is similar).

42. David Abrams, Mariannee Bertrand & Sendhil Mullainathan, *Do Judges Vary in Their Treatment of Race?*, 41 J. LEGAL STUD. 347, 356, 371 (2012); Ronald Everett & Roger Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANTITATIVE CRIMINOLOGY 189, 206–07 (2002).

43. William Rhodes, Ryan Kling, Jeremy Luallen & Christina Dyous, *Federal Sentencing Disparity: 2005-2012* 41 (BUREAU OF JUST. STAT., Working Paper No. 2015:01, 2015), https://www.bjs.gov/content/pub/pdf/fsd0512.pdf.

Other authors have found that heavier penalties have been imposed on Black offenders who harmed white victims than on offenders who harmed Black victims, accounting for this outcome by explaining that "the judges were also white, and their in-group or worldview was more threatened by criminal conduct against persons from their in-group."[44] Harsher penalties for certain groups within the community cannot be uncoupled from the reality that sentencing discretion, which is part of the judicial decision-making process, unavoidably leads to sentences based, at least in part, on the personal predispositions of judges. As a result, even when judges do their best to be fair and objective, they inevitably have preferences and biases that inform their decision-making. Research has shown that:

> [T]he influence of unconscious bias on judges is subtle. We know that judges harbor many of the same implicit associations as most adults. For example, in our study using the implicit association test, we found that 80 percent of white judges more strongly associated Black faces with negative words, and white faces with positive words. Black judges expressed a more complex pattern, with some judges showing the same white-good/Black-bad association as white judges, but an equal number showing the opposite preference. These results suggest that judges are no different than most adults in the United States.[45]

Research on judicial cognition suggests that judging is at least partly an intuitive or unconscious cognitive process like other human decision-making. This can sometimes assist in quick and efficient decision-making, but it can also produce systematic errors. Judges are generally not deliberate users of poor quality or inaccurate assumptions; rather, their decision-making can be influenced by the unconscious impact of factors like bounded rationality, which reflects the costs and difficulties of obtaining all information required for rational decision-making, the time constraints, and the physical limitations of human cognitive ability. When dealing with excess information in very short timeframes, judges draw on a range of cognitive heuristics, biases, and other cognitive factors including emotion, group identity, and cultural worldview.[46]

While judges are inclined to imbue their decisions with their subjective experiences, preferences, and values, this should not be accommodated by the general public, given that their decisions have resulted in unequal treatment of the people

44. Siegfried L. Sporer & Jane Goodman-Delahunty, *Disparities in Sentencing Decisions*, *in* SOCIAL PSYCHOLOGY OF PUNISHMENT ON CRIME 379, 390 (Margrit E. Oswald et al. eds., 2009) [hereinafter SOCIAL PSYCHOLOGY]; *see also* Scott Duxbury, *Who Controls Criminal Law? Racial Threat and the Adoption of State Sentencing Law, 1975 to 2012*, 86 AM. SOCIO. REV. 123, 144–46 (2021) (finding sentencing policies are uniquely responsive to white public policy opinion).

45. Bernice B. Donald, Jeffrey Rachlinski & Andrew J. Wistrich, *Getting Explicit About Implicit Bias*, 104 JUDICATURE 75, 76 (2021).

46. Kylie Burns, *Judges, 'Common Sense' and Judicial Cognition*, 25 GRIFFITH L. REV. 319 (2016).

affected.[47] The pervasive biases of police, prosecutors, and judges show that the current decision-making process in the criminal justice system is broken. We now discuss the benefits of artificial intelligence and how the system can be improved.

II. THE NATURE OF ARTIFICIAL INTELLIGENCE

A. *Overview of Algorithms and Artificial Intelligence*

AI systems utilize algorithms, which are capable of synthesizing large amounts of data involving prior actions or behaviors, to predict future behaviors. Algorithms are not new[48] and are simply "instructions for solving a problem or completing a task. Recipes are algorithms, as are math equations. Computer code is algorithmic."[49] Algorithms are increasingly used because "massive amounts of data are being created, captured and analyzed by businesses and governments."[50] They already have an important role in many aspects of society, from risk assessments for insurance premiums, to detection of tax fraud,[51] to determinations about credit risk.[52]

Because AI systems use algorithms to synthesize data and solve problems, there is an inextricable connection between algorithms and AI.[53] All AI systems are based on algorithms, but not all algorithms operate within an AI construct. The main advantage of AI systems that incorporate algorithms is that they are capable of providing accurate and efficient solutions to problems that often require computation or assessment of a large number of variables. AI can operate in an unlimited number of areas. One of the most commonly used forms of AI is Siri, which is a virtual assistant that uses voice recognition to provide information to iPhone users. Other common examples include ridesharing apps like Uber that use AI to anticipate driver demand;[54] plagiarism checkers such as "Turnitin";[55] and social media apps like Facebook that use AI to suggest connections.[56]

47. Mirko Bagaric, *Sentencing: From Vagueness to Arbitrariness: The Need to Abolish the Stain that Is the Instinctive Synthesis*, 38 U.N.S.W.L.J. 76, 110–11 (2015).

48. *See* Lee Rainie & Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age*, PEW RSCH. CTR. (Feb. 8, 2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

49. *Id.*

50. *Id.*

51. Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947, 952, 957 (2016).

52. *Id.*

53. Throughout this Article, algorithm and artificial intelligence, or AI, are used interchangeably.

54. Daniel Faggella, *Everyday Examples of Artificial Intelligence and Machine Learning*, EMERJ (Mar. 10, 2020), https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/.

55. *Id.*

56. *Id.*

Advances in neural network technology—especially deep learning systems—drive the recent explosion of interest in AI.[57] Deep learning is a statistical method for classifying patterns, based on large amounts of sample data, using neural networks with multiple layers. The networks are constructed with input nodes connected to output nodes via a series of "hidden" nodes arranged in layers. Input nodes can represent any data—such as pixels in image recognition and words in speech recognition—and the outputs involve the decision or coding that the researcher is looking for, e.g. the classification of a picture or the meaning of the sentence. All of the nodes (or "neurons") within the network have activation levels, so that a neuron will 'fire' if the nodes connected to it add up to a certain activation level or higher. All of the connections initially have a random weighting assigned to them, but by using a large training set and a process called back-propagation, eventually, the activation levels and weighting are adjusted to the point where any given input will produce the correct output.[58]

The best way to understand the manner in which algorithms work is to go from the general to the specific. To this end, we will discuss further the workings of Google Maps, weather forecasts, and automatic airline pilots.

B. Google Maps, Weather Forecasts, and Automatic Piloting

1. Google Maps

Google Maps is used by over one billion users annually, offering digitalized navigation for both indoor and outdoor environments.[59] Google Maps gathers "massive amounts of geospatial data," through its Base Map Partner Program[60] and vehicle patrols, which is provided to users via a mobile application.[61] To help users get to their destinations, Google Maps uses the Dijkstra algorithm,[62] which builds "up a solution piece by piece, always choosing the next piece that offers the

57. The field expanded in 2012 when Krizhevsky, Sutskever, and Hinton demonstrated remarkable results in image classification and object recognition using large scale multi-layer, deep networks. *See* Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, 60 COMMC'NS. ACM 84 (2017); RONALD ASHRI, THE AI-POWERED WORKPLACE HOW ARTIFICIAL INTELLIGENCE, DATA, AND MESSAGING PLATFORMS ARE DEFINING THE FUTURE OF WORK 36 (2019) (identifying the Krizhevsky, Sutskever & Hinton paper as a "seminal" work in the field of neural networks).

58. *See* GARY MARCUS, DEEP LEARNING: A CRITICAL APPRAISAL 3–4, 7 (2018), https://arxiv.org/pdf/1801.00631.pdf.

59. Harold Stark, *Since You Asked, Here's How Google Maps Really Works*, FORBES (Apr. 26, 2017, 8:57 AM), https://www.forbes.com/sites/haroldstark/2017/04/26/since-you-asked-heres-how-google-maps-really-works/#45b878d04dbe.

60. This program allows Google to collect data from organizations like "the US Geological Survey, Forest Service, city and state councils and so forth . . . ." *Id.*

61. *Id.*

62. The Dijkstra algorithm was created by Edsger W. Dijkstra in 1956 to "find the shortest route between two cities in the Netherlands." Michael Byrne, *The Simple, Elegant Algorithm that Makes Google Maps Possible*, VICE (Mar. 22, 2015, 7:00 AM), https://www.vice.com/en_us/article/4x3pp9/the-simple-elegant-algorithm-that-makes-google-maps-possible.

most obvious and immediate benefit."[63] Google Maps has made the lives of many easier by preventing individuals from getting lost, allowing drivers to beat traffic, revolutionizing the real estate market, and leading to the creation of many other helpful apps.[64] It has made detective work easier, for example, by allowing officers to expose child porn rings and helping to reunite a mother and son.[65] Most recently, Google Maps has added features allowing users in many countries to safely negotiate travel in light of the coronavirus (COVID-19) pandemic.[66] However, like many other systems that rely on algorithms, Google Maps is not without criticism. Real time location sharing has made it easier for abusers to monitor their partners, potentially exposing vulnerable individuals to further abuse.[67]

2. Weather Forecasting

Weather forecasts involve the prediction of future weather conditions. Forecasts can be made in a variety of ways, "from [a] relatively simple observation of the sky to highly complex computerized mathematical models."[68] Technological advances have led to improvements in weather forecasts. For instance, the error rate for high temperature forecasts in the United States for a five-day period has improved by two degrees since 1975, making the forecasts likely to be only about five degrees off.[69] Flash flood warnings and hurricane landfall predictions have also improved.[70] Accurate forecasts "can mean the difference between prosperity and ruin, life and death,"[71] but the perfect forecast is still unreachable.[72]

---

63. Yoon Mi Kim, *Dijkstra Algorithm: Key to Finding the Shortest Path, Google Map to Waze*, MEDIUM (June 12, 2019), https://medium.com/@yk392/dijkstra-algorithm-key-to-finding-the-shortest-path-google-map-to-waze-56ff3d9f92f0.

64. Hannah Francis, *10 Years of Google Maps: 10 Ways It Changed the World*, SYDNEY MORNING HERALD, https://www.smh.com.au/technology/10-years-of-google-maps-10-ways-it-changed-the-world-20150212-13d7wq.html (last updated Feb. 13, 2015, 1:55 PM).

65. *Id.*

66. *See* Alaa Elassar, *Google Maps Releases New Features to Help People Navigate Coronavirus Hot Spots*, CNN, https://www.cnn.com/2020/07/04/us/google-maps-coronavirus-warning-trnd/index.html (last updated July 4, 2020, 4:03 PM) (outlining efforts to provide users with restriction, testing, and public transit alerts).

67. *See id.*; Kevin Murnane, *The Good, the Bad, and the Ugly of Google Maps' New Real-Time Location Sharing*, FORBES (Mar. 22, 2017, 12:48 PM), https://www.forbes.com/sites/kevinmurnane/2017/03/22/the-good-the-bad-and-the-ugly-of-google-maps-new-real-time-location-sharing/#669a0cf95a86.

68. Piyush Kapoor & Sarabjeet Singh Bedi, *Weather Forecasting Using Sliding Window Algorithm*, INT'L SCHOLARLY RSCH. NOTICES, 2013, at 1, 1.

69. Tim Brookes, *Weather Forecasting*, NAT'L GEOGRAPHIC, https://www.nationalgeographic.com/science/earth/earths-atmosphere/weather-forecasting-science/#close (last visited July 28, 2020).

70. *Id.*; *see also* Kapoor & Bedi, *supra* note 68, at 1 (noting that accurate weather predictions are essential for "climate monitoring, drought detection, severe weather prediction, agriculture and production, planning in energy industry, aviation industry, communication, pollution dispersal, and so forth" and that "[i]n military operations, there is a considerable historical record of instances when weather conditions have altered the course of battles.").

71. Brookes, *supra* note 69.

72. *See id.*

Despite the difficulty in creating the perfect forecast, algorithms improve accuracy by processing large amounts of data. The National Weather Service takes in "192,000 observations from surface stations, 2,700 observations from ships, 18,000 from weather buoys, 115,000 from aircraft, about 250,000 from balloons, and 140 million from satellites" daily.[73] Modern weather forecasting is done through numerical weather prediction models.[74] A model is created by dividing the atmosphere into a 3D grid spanning the planet. After current conditions for each cell are inputted, the model runs using "the starting conditions to calculate how the weather is expected to change in each cell."[75] Then, the model creates predictions about when such changes might occur, thereby creating a weather forecast.[76] Once the models have calculated future conditions, meteorologists can verify the algorithms' accuracy.[77] Even if there are some errors, the use of computer algorithms, including machine learning algorithms, for forecasting future weather conditions is superior to other methods.[78]

3. Automatic Piloting

Today's aircraft use sophisticated autopilot systems that perform the duties of a highly-trained pilot, making flights safer and more efficient.[79] Autopilot, also known as the automatic flight control system ("AFCS"), is included as part of a plane's avionics system, which is the "electronic systems, equipment, and devices used to control key systems of the plane and its flight."[80] Although autopilot was originally created to help pilots "during tedious stages of flight," systems today conduct very precise maneuvers and control various parts of the plane.[81] An autopilot system consists of high-speed processors that gather data from plane sensors and other equipment, "including gyroscopes, accelerometers, altimeters, compasses and airspeed indicators."[82] The system's flight director component then compares the data—using a series of calculations—to a control mode entered by

73. *Id.*
74. Nikhil Chandwani, *How Are Weather Forecasts Made?*, NYK DAILY (Oct. 6, 2019), https://nykdaily.com/2019/10/how-are-weather-forecasts-made/.
75. *Id.*
76. *Id.*; *see* Brookes, *supra* note 69 (noting that meteorologists may use a time machine computer model, which uses equations to analyze information on present conditions and predict future conditions).
77. Brookes, *supra* note 69.
78. *See* Laurie Brenner, *Four Types of Forecasting*, SCIENCING, https://sciencing.com/four-types-forecasting-8155139.html (last updated Nov. 22, 2019); *see also* MARK HOLMSTROM, DYLAN LIU & CHRISTOPHER VO, MACHINE LEARNING APPLIED TO WEATHER FORECASTING 1 (2016), https://pdfs.semanticscholar.org/2761/8afb77c5081d942640333528943149a66edd.pdf; Chandwani, *supra* note 74 (noting that "[f]or a lot of the routine weather, the forecast models are so good now that there's really not that much that the human forecasters are going to add.").
79. William Harris, *How Autopilot Works*, HOWSTUFFWORKS, (Oct. 10, 2007), https://science.howstuffworks.com/transport/flight/modern/autopilot.htm/printable.
80. *Id.*
81. *Id.*
82. *Id.*

the pilot to ensure the plane is obeying flight parameters.[83] A servomechanism is then used to adjust the aircraft's control surfaces where necessary for proper performance.[84] This process is indicative of a negative feedback loop, which is "a self-regulating system that reacts to feedback in a way that maintains equilibrium."[85]

Much like other algorithms, autopilot systems are not fail-proof. Systems may stop working, but "no failure in the automatic pilot can prevent effective employment of manual override," and pilots must be able to fly an aircraft with and without autopilot systems.[86] Furthermore, pilots must understand how to program the systems, as "[a]utopilots are dumb and dutiful, meaning . . . if you program them incorrectly, they will kill you."[87] Autopilot algorithms have revolutionized the art of flying a plane, but human involvement is still necessary because "[a]s advanced as the technology is, an autopilot is not auto enough to think for itself, which means it's not smart enough to fly a plane by itself . . . ."[88] Algorithms and AI predominate our daily lives by helping ensure safety and flexibility. Now that we understand how they work in other contexts, we will turn to their use in the criminal justice system.

III. CURRENT USE OF ALGORITHMS AND AI IN THE CRIMINAL JUSTICE SYSTEM

We now discuss the manner in which AI has been incorporated into each stage of the criminal justice system, commencing with policing and crime detection, and then moving through the stages of parole determinations.

A. Policing and Detection of Crime

1. Predictive Policing

The detection and regulation of criminal activity has traditionally been reactive, meaning it occurs in response to real-time events rather than as a proactive analysis of historical and evolving evidence and data.[89] AI has the capacity to greatly increase the effectiveness of proactive, or predictive, policing. Algorithms, designed based on previous patterns of behavior, can predict the likelihood of

---

83. *Id.*

84. *Id.*

85. Alexandra Appolonia, Abby Tang & Uma Sharma, *How Autopilot on an Airplane Works*, BUS. INSIDER (Oct. 15, 2019, 9:30 AM), https://www.businessinsider.com/autopilot-how-airplane-automatic-flight-control-system-pilots-2019-10; *see also* Harris, *supra* note 79, (explaining what a negative feedback loop is to illustrate how automated flight control systems operate).

86. Harris, *supra* note 79.

87. Appolonia et al., *supra* note 85.

88. *Id.*; *see also* Morgan Cutolo, *We Finally Know How Autopilot Works on an Airplane*, READER'S DIGEST, https://www.rd.com/article/how-airplane-autopilot-works/ (last updated Jan. 14, 2020) (explaining various features of autopilot and how pilots use them when flying planes).

89. Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOCIO. REV. 977, 989 (2017).

crime in a certain geographical location, at a certain time, with a high degree of accuracy.

Expansions in technology have led to the use of algorithms in policing and detection of crime, and while many question the legitimacy of the use of these systems, others believe that they could meaningfully improve the criminal justice system. Predictive algorithms are used to make decisions about potential criminal activity, with the premise that "crime is not randomly distributed across people or places."[90] Predictive policing, as part of a larger move toward "algorithmic governance," is becoming "the new predominant method of policing, which thus impacts other methods of policing."[91] In fact, several jurisdictions have mandated the use of such technology, though use of "algorithmic forecasts alone do[es] not meet the threshold of reasonable suspicion or probable cause . . . ."[92]

As mentioned above, predictive policing has made its way into many jurisdictions across the United States. For example, the Arnold Foundation algorithm has been made available in twenty-one United States jurisdictions and pulls from 1.5 million criminal cases to predict behavior during the pretrial phase of a criminal case.[93] Police in Birmingham, Alabama, believe that using algorithms will "make the fight against crime smarter."[94] Accordingly, the police reinstituted the use of computerized statistics ("Compstat")[95] because "[t]his is what fighting crime looks like."[96] Compstat became highly regarded among the agencies using it because of the effectiveness and benefit provided as an information-sharing tool between agencies.[97] Although Compstat is proactive in terms of matching resources to needs, its algorithms identify existing trends instead of predicting them. When

---

90. Brayne & Christin, *supra* note 1, at 610.

91. Aleš Završnik, *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*, 18 EUR. J. CRIMINOLOGY 623, 624 (2021).

92. Brayne & Christin, *supra* note 1, at 610 (citing ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT X (2017)).

93. Završnik, *supra* note 91, at 625 (citing Shaila Dewan, *Judges Replacing Conjecture with Formula for Bail*, N.Y. TIMES (June 26, 2015), https://www.nytimes.com/2015/06/27/us/turning-the-granting-of-bail-into-a-science.html).

94. Carol Robinson, *Beat by Beat, Street by Street, Birmingham Police Use Algorithms to Fight Crime in Age of Analytics*, AL.COM (Sept. 6, 2019), https://www.al.com/news/birmingham/2019/09/beat-by-beat-street-by-street-birmingham-police-use-algorithms-to-fight-crime-in-age-of-analytics.html.

95. Compstat is a "decades-old policing concept first made popular in the 1990s." Robinson, *supra* note 94. At that time, crime data was collected primarily to provide statistics to the FBI, and anything like real-time trends in crime rates, types, or locations were basically anecdotal. *See* David Weisburd, Stephen D. Matrofski, James J. Willis & Rosann Greenspan, *Changing Everything so that Everything Can Remain the Same: CompStat and American Policing, in* POLICE INNOVATION: CONTRASTING PERSPECTIVES 284, 284–301 (David Weisburd & Anthony A. Braga eds., 2001).

96. Robinson, *supra* note 94.

97. In a survey of its members, the Police Executive Research Forum asked, "[w]hy is Compstat used by your agency?" The top five responses were: "To identify emerging problems; To coordinate the effective deployment of resources; To increase accountability of commanders/ managers; To identify community problems and develop police strategies; To foster information-sharing within the agency." POLICE EXEC. RSCH. FORUM, BUREAU OF JUST. ASSISTANCE, COMPSTAT: ITS ORIGINS, EVOLUTION, AND FUTURE IN LAW ENFORCEMENT AGENCIES 8 (2013).

initially used in Birmingham, "Compstat used weekly statistics reports and computer-generated pin maps to closely scrutinize crime trends and patterns," but now it "relies on a multi-faceted approach to crime reduction involving quality of life improvement and personnel and resource management."[98] Compstat, originally based on the Broken Windows Theory,[99] allows officers to identify crime spikes and use targeted enforcement to address such spikes.[100]

Birmingham—like many other jurisdictions—has come to rely on predictive policing. Several jurisdictions, including Birmingham, also adopted PredPol, a system that was first used in Los Angeles, California.[101] PredPol employs predictive policing by using "data technology and information about past crimes to predict future unlawful activity."[102] The program can predict, within a twelve-hour window, "where and when crimes [are] likely to occur" using an algorithm that reviews ten years of "data, including the types of crimes and the dates, times and locations where they occurred."[103] Every day, the program creates a grid of boxes "designated as zones for possible property crimes such as burglaries and car thefts."[104] Additionally, the police publish these zones on social media, which is meant to engage the public in helping deter crime.[105] However, as of March 2019, there was insufficient data available to determine whether the software had in fact decreased crime rates.[106] Other jurisdictions have complained about the usefulness of PredPol, with many dropping the service when it failed to perform as

98. Robinson, *supra* note 94.

99. The Broken Windows Theory proposes that policing lower-level crimes will prevent more serious crimes. *See* Sarah Childress, *The Problem with "Broken Windows" Policing*, PBS FRONTLINE (June 28, 2016), https://www.pbs.org/wgbh/frontline/article/the-problem-with-broken-windows-policing/.

100. *See* Robinson, *supra* note 94.

101. *See* Mark Puente, *LAPD Pioneered Predicting Crime with Data. Many Police Don't Think It Works*, L. A. TIMES (July 3, 2019), https://www.latimes.com/local/lanow/la-me-lapd-precision-policing-data-20190703-story.html. PredPol uses an algorithm that incorporates aspects that detect seismic activity:

> Just as earthquakes happen along fault lines. . .research has shown crime is often generated by structures in the environment, like a high school, mall parking lot or bar. Additional crimes tend to follow the initial event near in time and space, like an aftershock. PredPol uses years of crime data to establish these patterns and then the algorithm uses near real-time crime data to predict the next property crime. Other systems use even more esoteric data — from the weather to phases of the moon — to arrive at their crime forecasts.

Justin Jouvenal, *Police Are Using Software to Predict Crime. Is It a 'Holy Grail' or Biased Against Minorities?*, WASH. POST (Nov. 17, 2016), https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities/2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html?utm_term=.e0875d4113f8.

102. Puente, *supra* note 101.

103. *Id.*

104. *Id.*

105. *Id.*

106. *Id.*

expected.[107] As of July 2019, only "60 of the roughly 18,000 police departments across the United States used PredPol."[108]

While the use of the algorithms remains controversial, the limited data that is available suggests that predictive policing systems are statistically more likely to predict when and where some crimes will occur than human crime analysts.[109] Furthermore, while some studies have shown that they can target minority groups in certain contexts,[110] a recent study of PredPol found "no statistically significant difference between arrest rates by ethnic group."[111]

In addition to using AI to determine where crime is likely to occur next, some police departments use algorithms, which incorporate variables such as age and criminal history, to determine whether particular individuals are likely to commit a crime or have previously committed a crime.[112] Risk assessment tools are a common form of predictive policing and are defined as "statistical models used to predict the probability of a particular future outcome."[113] Such predictions are usually based on a person's features, which are given numerical value based on how they contribute to a specific outcome and further configured as a risk score:[114] In criminal law "[t]he data used to build these systems are typically administrative records collected by local police departments and administrations of the court."[115] Predictions address a number of issues, including whether a person will fail to appear at court and their likelihood of committing a crime.[116]

In Chicago, an algorithm is used to score every arrested person with a threat score from 1 to 500-plus, which then influences who police target for proactive intervention, as well as the manner in which they are handled.[117] Many police departments in other cities in the United States also use similar algorithms.[118]

107. *Id.*

108. *Id.*

109. *See* Jouvenal, *supra* note 101; *see also* James Ford, *Should CompStat be eliminated? A history and analysis of the crime tally provide answers*, June 29, 2020: https://pix11.com/news/local-news/should-compstat-be-eliminated-a-history-and-analysis-of-the-crime-tally-provide-answers/; Indiana University, *Field-Data Study Finds No Evidence of Racial Bias in Predictive Policing,* PHYS.ORG (Mar. 13, 2018), https://phys.org/news/2018-03-field-data-evidence-racial-bias-policing.html#nRlv. (finding that predictive policing in Los Angeles did not result in biased arrests).

110. Jouvenal, *supra* note 101.

111. *Id.*

112. *See* Will Douglas Heavan, Predictive Policing Algorithms are Racist. They Need to be Dismantled, MASS. INST. TECH. TECH. REV. (July 17, 2020) https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/.

113. P'SHIP ON AI, REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM 7 (2019), https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/ [hereinafter PARTNERSHIP ON AI].

114. *Id.*

115. Barabas, *supra* note 11, at 87 (citing Andrew Guthrie Ferguson, *The Police Are Using Computer Algorithms to Tell if You're a Threat*, TIME (Oct. 3, 2017), http://time.com/4966125/police-departments-algorithms-chicago/.).

116. PARTNERSHIP ON AI, *supra* note 113, at 7.

117. Ferguson, *supra* note 115.

118. *See generally id.*

Algorithms like this one are supported because they have accurately predicted locations with high rates of shooting victims based on their profiles. Critics argue, however, that high threat scores inappropriately distort police decisions relating to the use of force, leading to disproportionate police monitoring of minorities—a risk which is exacerbated by the fact that the algorithm is confidential.[119]

In the UK, scholars debate the use of algorithms to determine whether to charge an arrested individual and with what crime.[120] One tool used in this context is the Harm Assessment Risk Tool ("HART").[121] HART was created under the Turning Point program, which was primarily aimed at female offenders who were treated more harshly than other groups.[122] The success of Turning Point led to the creation of Checkpoint, which was aimed at even more groups.[123] Neither Turning Point nor Checkpoint required the use of algorithms, but the need for such tools was identified based on the fact that officers faced challenges in identifying people who posed a risk of recidivating.[124] HART uses a random forest algorithm and is meant to "reduce the number of people entering the justice system, and by doing so, hopefully reduc[es] the number of people re-entering it."[125] Despite its potential, HART was criticized for using postal code data and personal data that was purchased from Experian, which many viewed as an invasion of privacy.[126] Other police forces in the UK use forecasting systems that score individuals based on a number of attributes, including: potential for committing crimes, ranging from gang-related activities to sexual violence and burglary; potential for becoming a victim; and potential of police officers to engage in crime misclassifications or misconduct.[127]

The exponential rise in the collection and retention of information in both the public and private sectors has given police access to information about people who have never been offenders or have otherwise come into contact with the criminal justice system. This access, in conjunction with advances in coding techniques and big data analytics, has made it possible for policing algorithms to become truly predictive.

More sophisticated predictive policing systems use machine learning to learn how a much wider range of factors correlates with crime. Data is used to predict where and when crime will occur in the future. As the accuracy of those forecasts

---

119. *Id.*

120. THE LAW SOC'Y COMM'N ON THE USE OF ALGORITHMS IN THE JUST. SYS. & THE LAW SOC'Y OF ENGLAND AND WALES, ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM 45 (2019), https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/ [hereinafter ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM].

121. *Id.*

122. *Id.* (citing PIPPA COUTTS, ALL. FOR USEFUL EVIDENCE, TURNING POINT: THE POLICE'S PRODUCTION AND USE OF EVIDENCE TO REDUCE REOFFENDING (2018), https://www.alliance4usefulevidence.org/wp-content/uploads/2018/01/Turning-point-case-study-v1.pdf).

123. *Id.*

124. *Id.* at 45–46.

125. *Id.* at 46.

126. *Id.*

127. *Id.* at 47–48.

—and how they correlate with the factors used to make the forecasts—becomes clear, the algorithm then "learns" to make more accurate predictions. One such web-based system, Hunchlab, bases its forecasts on "records of public reports of crime and requests for police assistance, as well as weather patterns and Moon phases, geographical features such as bars or transport hubs, and schedules of major events or school cycles."[128] The provider claims that the system limits the potential harmful effects of over-policing by using a probabilistic model, which varies the locations to which officers are sent and provides them with timers to "encourage" them to avoid over- or under-policing a given area.[129] Only external evaluation will truly determine the success of these efforts.

Jurisdictions continue to move towards proactive policing because traditional reactive policing strategies of detection and investigation do not work.[130] However, as discussed below, there are well-documented problems of entrenched bias and potential for unethical and inequitable application and enforcement with algorithms. As a result, efforts to place too much value on the objective accuracy of computational models will continue to be a concern. A properly designed system, where emphasis is placed not only on crime prediction but also on a regulatory and ethical framework at the development and deployment stages, is surely not beyond us. This is examined further in Part V below.

2. Automated Visual Monitoring

Police increasingly rely on technology as part of predictive policing, and machine learning approaches can further assist crime reduction and detection by supplementing the use of existing criminal justice technological innovations. This is best illustrated by the use of CCTV cameras in millions of locations throughout the United States. These cameras have a two-fold role in the criminal justice sphere. First, they discourage the commission of crime in circumstances where offenders are aware of the location of the cameras. Empirical data establishes that the best way to reduce crime is to increase the perception in people's minds that if they offend, they will be detected and apprehended.[131] Research has shown the cameras achieve this goal.[132] The second role of cameras is to gather evidence that can be used by police and prosecutors for detecting crime, identifying offenders, and establishing their guilt in court.

---

128. Aaron Shapiro, *Reform Predictive Policing*, 541 NATURE NEWS 458, 459 (2017), https://www.nature.com/news/reform-predictive-policing-1.21338.

129. *See Strategically Plan Patrol and Task Force Missions for Maximum Crime Deterrence*, SHOTSPOTTER, https://www.hunchlab.com/resources/ (last visited Mar. 26, 2020).

130. *See, e.g.*, Lawrence W. Sherman, *The Rise of Evidence Based Policing: Targeting, Testing and Tracking Police Services*, 42 CRIME & JUST. 377, 388 (2013).

131. *See* Mirko Bagaric & Theo Alexander, *(Marginal) General Deterrence Doesn't Work – and What it Means for Sentencing*, 35 CRIM L.J. 269, 276–77 (2010).

132. *See* AUSTRALIAN INST. OF CRIMINOLOGY, EFFECTIVENESS OF PUBLIC SPACE CCTV SYSTEMS 20 (2017).

A major limitation associated with the use of CCTV cameras is that real-time monitoring of feeds is extremely labor intensive (given that individuals can only focus on a limited number of cameras at any given time), which makes them less effective at stopping crime and apprehending criminals. This process can be made far more cost-effective and efficient through computer monitoring of CCTV footage, and recent advances in machine learning visual processing now allow for large-scale automated monitoring of locations, as well as the flagging of problematic behavior within that space.

The use of AI to monitor CCTV and alert law enforcement officers to suspicious behavior is already occurring in a number of locations, including the Swinburne University of Technology in Australia. The system used at this location is called iCetana, which is one of the leading manufacturers of real-time, AI-assisted video monitoring.[133] The tool learns by monitoring the relevant area for a period of time and then flags unusual behavior. The system is constantly re-calibrating movement patterns in order to classify the types of behavior that are normal, while also using irregular behavior as a proxy for activity that is potentially criminal activity. It detects behavior such as running, loitering, falling, and punching; it can even recognize pre-aggression stances that occur due to differences in posture that coincide with hostility, though the system is not sufficiently nuanced to pick up all forms of criminal conduct.[134] In addition to self-learning automated CCTV algorithms, like iCetana, there are also rule-based systems in which a computer is pre-programmed to raise an alert whenever certain events occur, even if they are not classified as unusual. As discussed further below, these systems could be programmed to detect more subtle forms of offending.

AI-based visual processing is also used extensively in facial recognition, which is commonly used in the criminal justice system. Machine learning techniques have advanced quickly in this area, and their accuracy has consistently improved.[135] For policing, "it can be envisaged that facial recognition be deployed in attempting to detect either suspects, victims, or vulnerable persons."[136] Recently, facial recognition systems have hit the headlines for a range of reasons: the potential misuse of the technology by commercial operators to discriminate against certain groups,[137] its privacy-

133. *Product*, ICETANA, https://icetana.com/icetana-product-overview/ (last visited Mar. 26, 2020). The technology has been in use on the Swinburne campus for over seven years. *See* GRAEME WOODS, ICETANA, VIDEO SURVEILLANCE FOR UNIVERSITY & COLLEGE CAMPUSES (2019), https://s.icetana.com/wp-content/uploads/2019/02/Video-surveillance-for-university-and-college-campuses-iCetana-White-Paper.pdf.

134. *Product*, ICETANA, *supra* note 133.

135. Kate Kaye, *This Little-Known Facial-Recognition Accuracy Test Has Big Influence*, IAPP, Jan. 7, 2019, https://iapp.org/news/a/this-little-known-facial-recognition-accuracy-test-has-big-influence/ (reporting on NIST tests, reporting facial recognition accuracy rates as high as 99.8%).

136. ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 36.

137. Jieshu Wang, What's in Your Face? Discrimination in Facial Recognition Technology (2018) (MA thesis, Georgetown University), http://hdl.handle.net/10822/1050752.

invading nature,[138] limitations in datasets that have led to misidentification of people in certain groups,[139] and the way authoritarian governments can use the technology to control dissidents or ethnic groups.[140]

In the UK, there are approximately 12.5 million "biometric, searchable faces" available within the Police National Database. Police have been able to obtain photos with or without consent, as well as to share or keep them as needed for crime prevention, though there is no power for police "to photograph and retain or repurpose images of individuals in public spaces more generally."[141] There have been rulings in favor of individuals challenging the police practice of retaining photographs, particularly of individuals who were in police custody.[142] However, the rule appears to have not been implemented generally.[143] There are consistent challenges to the use and retention of such photographs, but no concerted efforts have been made to resolve gaps in legislation that allows the use of these photographs to continue.[144] Even though there are concerns about the technology that must be overcome, there are also manifest benefits that can come from using AI to discourage crime and apprehend criminals.

As noted above, algorithms are used in many parts of our everyday lives,[145] and just as these programs enhance our lives, many believe that these programs will enhance the criminal justice system.[146] In both the United States and the UK, the focus in "criminal justice has been moving from the crime and the criminal to the victim and victimization, with an emphasis on vulnerability placing greater demands on police forces to act in a generally anticipatory and preventative, rather

138. *See, e.g.*, I. Bennett Capers, *Crime, Surveillance, and Communities*, 40 FORDHAM URB. L.J. 959, 963–64 (2012); Sahil Chinoy, *We Built an 'Unbelievable' (but Legal) Facial Recognition Machine*, N.Y. TIMES (Apr. 16, 2019), https://www.nytimes.com/interactive/2019/04/16/opinion/facial-recognition-new-york-city.html; Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion,* 163 U. PA. L. REV. 327 (2015)*;* Wayne A. Logan, *Policing Identity*, 92 B.U. L. REV. 1561 (2012).

139. Victoria Burton-Harris & Philip Mayor, *Wrongfully Arrested Because Face Recognition Can't Tell Black People Apart*, ACLU (June 24, 2020), https://www.aclu.org/news/privacy-technology/wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart/; Steve Lohr, *Facial Recognition Is Accurate, if You're a White Guy,* N.Y. TIMES (Feb. 9, 2018), https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

140. *See, e.g.*, Paul Mozur, *One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority*, N.Y. TIMES (Apr. 14, 2019), https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

141. ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 39.

142. *Id.* at 40. (citing R (RMC and FJ) v. Metropolitan Police Service [2012] EWHC 1681, para. 58 (noting that the "existing policy concerning the retention of custody photographs . . . is unlawful," after which the High Court gave the police a reasonable period of time to revise the policy)).

143. *Id.* A Biometrics Commissioner was appointed to address similar issues in 2012, but the role's mandate did not include photographs. *Id.*

144. *Id.* at 40–41.

145. Derek Thompson, *Should We Be Afraid of AI in the Criminal-Justice System?*, THE ATLANTIC (June 20, 2019), https://www.theatlantic.com/ideas/archive/2019/06/should-we-be-afraid-of-ai-in-the-criminal-justice-system/592084/.

146. ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 13.

than specifically reactive manner."[147] This renewed focus on preventative justice leaves space for the greater use of algorithms[148] and, indeed, in the United Kingdom, "[p]olicy makers have increasingly focused their attention on how technology may be able to assist in bridging the need to access justice."[149] There is evidence that predictive policing is more effective at reducing crime than traditional policing approaches. However, few studies have been undertaken, and the results of these studies are not definitive. Thus, there is a need to better evaluate current predictive policing systems, including computer monitoring of CCTV and the use of facial recognition. This is discussed further in Part V below.

B. Bail

Bail decisions are made following the charging of a suspect and prior to the determination of guilt or innocence. At this point, the suspect has not been convicted of an offense, and a decision must be made as to whether the offender should be released into the community until criminal liability is determined.[150] The main determinants are 1) the risk that a suspect will reoffend if they are released into the community, and 2) whether the offender is a flight risk; both are usually determined by judges.

Bail decisions are made in courtrooms every day across the United States, and most judges make these decisions by relying on their own intuition.[151] By doing so, however, the system becomes subjective. To prevent this from occurring, many courts have begun to rely on algorithms to make these decisions.[152] Algorithms, unlike humans, "bring a measure of consistency and evenhandedness to the process."[153] A large amount of data is available to identify characteristics that are most indicative of a risk of absconding or reoffending, and this data can be readily collated and used to develop an algorithm "to assess the likelihood of the recidivism or flight of those awaiting trial or offenders in bail and parole procedures."[154]

The use of algorithms in bail decisions can lead to benefits for communities, as seen in recent efforts to incorporate algorithms into the criminal justice systems in Virginia and New Jersey.[155] In Virginia, random agencies incorporated an algorithm

---

147. *Id.* (citing Emma Williams, Jennifer Norman & Daniela Wünsch, *Too Little Too Late: Assessing Vulnerability*, 3 POLICING 355 (2009)).

148. *See id.*

149. *Id.* at 14.

150. Forms of release that are permitted include payment of a full cash bond, grant of an unsecured bond or conditional release, or bail guaranteed by way of surety, that is, by a third party (in some states, a commercial bail bondsman).

151. Sam Corbett-Davies, Sharad Goel & Sandra Gonzålez-Bailón, *Even Imperfect Algorithms Can Improve the Criminal Justice System*, N.Y. TIMES (Dec. 20, 2017), https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html.

152. *Id.*

153. *Id.*

154. Završnik, *supra* note 91, at 3.

155. Corbett-Davies et al., *supra* note 151.

rating defendants' "likelihood of skipping trial and their likelihood of being arrested if released."[156] Based on use of the algorithm, "[n]early twice as many defendants were released and there was no increase in pretrial crime."[157] The use of algorithms in New Jersey led to a "16 percent drop in its pretrial jail population, again with no increase in crime."[158] Similarly, a study conducted in New York City determined that for pretrial decisions, "an algorithm's assessment of risk would far outperform judges' track record."[159]

Some states have mandated algorithm usage when setting bail without much success.[160] In Kentucky, for instance, a law requiring judges to consult an algorithm to determine who would be eligible for cash bail led to Black defendants being dealt with more harshly than white defendants.[161] Kentucky was no stranger to the use of scoring systems to make decisions,[162] and the system "employed a point system to produce a score estimating the risk that a defendant [would] skip their court date or reoffend before trial."[163] Judges demonstrated their mistrust of the system by overruling its recommendation a substantial proportion of the time, and although this led to more defendants being released (for a short period of time), white defendants were the ones to benefit.[164]

In 2017, New Jersey installed an algorithm, the Public Safety Assessment ("PSA"), to "mathematically assess the risk of defendants fleeing or committing a crime—particularly a violent one—before their trial date."[165] Creators of this system believe it is different, as it looks across cases for traits that re-offenders share and also uses conviction records rather than arrest records, "which are less likely to tip the scales against individuals in heavily policed neighborhoods."[166] It weighs nine factors to predict the likelihood of three pretrial outcomes for a given offender, two of which relate to Failure to Appear (absconding).[167]

---

156. *Id.*

157. *Id.*

158. *Id.*

159. Adam Neufield, *In Defense of Risk-Assessment Tools*, MARSHALL PROJECT (Oct. 22, 2017, 10:00 PM), https://www.themarshallproject.org/2017/10/22/in-defense-of-risk-assessment-tools.

160. Završnik, *supra* note 91, at 3 (citing Laurel Eckhouse, *Big Data May be Reinforcing Racial Bias in the Criminal Justice System*, WASH. POST (Feb. 10, 2017), https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d_story.html); *See, e.g.*, Tom Simonite, *Algorithms Should've Made Courts More Fair. What Went Wrong?*, WIRED (Sept. 5, 2019, 7:00 AM), https://www.wired.com/story/algorithms-shouldve-made-courts-more-fair-what-went-wrong/ (discussing a 2011 Kentucky law mandating the use of an algorithm to decide whether a defendant should be jailed before trial).

161. Simonite, *supra* note 160.

162. *See id.* (noting that Kentucky began using pretrial risk scores in 1976).

163. *Id.*

164. *Id.*

165. Sarah Kramer, *An Algorithm Is Replacing Bail Hearings in New Jersey*, VICE (Feb. 23, 2017, 7:00 AM), https://www.vice.com/en_us/article/78mngy/an-algorithm-is-replacing-bail-hearings-in-new-jersey.

166. *Id.*

167. *See Risk Factors and Formula*, PUB. SAFETY ASSESSMENT, https://advancingpretrial.org/psa/factors/#psa-factors (last visited Oct. 7, 2021). The data used to determine the correlates was contained in approximately

Two outcomes are worth noting following the introduction of PSA in New Jersey. First, there was a substantial decline in the number of arrests in the year after implementation[168] and there was a reduction in the time defendants spend in jail the month after arrest.[169]

California also planned, but failed, to reform its cash bail system by relying on a similar algorithm that would determine who should be detained based on "statistical predictions of what other people with similar characteristics have done in the past . . . ."[170] The use of such an algorithm was legally required to have oversight, but some judges were attempting to use it without the mandated oversight, leading many to believe they would not use it to lower incarceration rates and would have "complete and unaccountable power" over such life-impacting decisions.[171]

C. Sentencing

Sentencing involves the deliberate infliction of sanctions on citizens, including the imposition of financial penalties, the deprivation of liberty and, in extreme cases, the death penalty. Sentencing decisions can have a major impact on people's lives, and it is unjustifiable for courts to make inconsistent, arbitrary, or opaque decisions, as such decisions would fundamentally violate the rule of law.[172] The rule of law is both a legal doctrine and normative concept of modern liberal democratic countries, which constitutes "an ideal towards which a legal order should move if it is . . . to secure certainty in human relations."[173] Although it is important that the law remains flexible and changes in response to shifting "public opinion," there is a crucial "need for certainty and stability in the law so that people will be able to plan and organize their arrangements in accordance with it."[174] Consistent, predictable, and transparent sentencing decisions constitute a crucial safeguard of the rule of law. As Maria Jean J. Hall and others also put it, "it is desirable that like cases be treated alike,"[175] and "there is universal acceptance that consistency of approach should be an essential feature of sentencing decision-making."[176]

750,000 bail decisions from 300 U.S. jurisdictions. *See PSA Background*, PUB. SAFETY ASSESSMENT, https://www.psapretrial.org/about/background (last visited Mar. 26, 2020).

168. CHLOE ANDERSON, CINDY REDCROSS, ERIN VALENTINE & LUKE MIRATRIX, PRETRIAL JUSTICE REFORM STUDY: EVALUATION OF PRETRIAL JUSTICE SYSTEM REFORMS THAT USE THE PUBLIC SAFETY ASSESSMENT 2 (2019), https://www.mdrc.org/sites/default/files/PSA_New_Jersey_Report_%231.pdf

169. *Id.*

170. Jeff Adachi, *Don't Let Judges Hijack California Bail Reform*, SACRAMENTO BEE (Nov. 27, 2017, 12:00 PM), https://www.sacbee.com/opinion/op-ed/soapbox/article186040863.html.

171. Adachi, *supra* note 170.

172. *See, e.g.*, JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 270–76 (1st ed. 1980); JOSEPH RAZ, THE AUTHORITY OF LAW 211, 214–16 (2d ed. 1979).

173. GEOFFREY DE Q. WALKER, THE RULE OF LAW 1 (1st ed. 1988).

174. *Id.* at 42.

175. Maria Jean J. Hall, Domenico Calabro, Tania Sourdin & Andrew Stranieri, *Supporting Discretionary Decision-Making with Information Technology: A Case Study in the Criminal Sentencing Jurisdiction*, 2 UNIV. OTTAWA L. & TECH. J. 1, 3 (2005).

176. *Id.* at 31.

One of the main reasons for the move from indeterminate to prescriptive sentencing in the United States over the past forty years is that inconsistencies previously plagued sentencing law and practice.[177] However, even prescriptive sentencing models have failed to achieve a reasonable level of consistency. Studies have demonstrated a wide-ranging sentencing disparity among judges applying the Federal Sentencing Guidelines.[178] A key reason for these inconsistencies is that judges' implicit biases and deep-rooted values and beliefs often affect their decision-making.[179]

An ideal algorithmic sentencing program would overcome the ingrained latent problems with sentencing decision-making. In Hutton's words, the program would comprise "a set of rules describing the criteria which should be taken into account and the method through which account is to be taken," as well as "an unambiguous, formally specified aim or set of aims for punishment, and a rational set of rules determining how appropriate punishments are to be allocated to particular cases."[180] In developing automated sentencing systems, it is important that a constant, unvarying suite of factors be used to inform the penalty—including aggravating and mitigating considerations that increase or decrease the penalty, respectively—and machine learning techniques should be used to determine the weight attached to the factors. Underpinning those elements would be clear aims that sentence are designed to achieve, namely: rehabilitation, community protection, incapacitation of serious sexual and violent offenders, and punishment that is commensurate with the seriousness of an offense.[181] Hutton envisaged a model sentencing system in which "any sentencer presented with the same case would reach the same decision as to the appropriate sentence. Thus, the sentence for any case would be predictable providing the correct rules and procedures had been followed."[182]

---

177. MARVIN E. FRANKEL, CRIMINAL SENTENCES: LAW WITHOUT ORDER 8 (Hill and Wang Press, 1st ed. 1972). For a critique of his impact, see Lynn Adelman & Jon Deitrich, *Marvin Frankel's Mistakes and the Need to Rethink Federal Sentencing*, 13 BERKELEY J. CRIM. L. 239 (2009).

178. *See, e.g.*, Joshua M. Divine, Booker *Disparity and Data-Driven Sentencing*, 69 HASTINGS L.J. 771 (2018) (noting disparate impacts of judicial discretion in the federal sentencing system); Nancy Gertner, *A Short History of American Sentencing: Too Little Law, Too Much Law, or Just Right*, 100 J. CRIM. L. & CRIMINOLOGY 691 (2010) (discussing various stages of American sentencing and issues faced throughout the evolution). In relation to the Federal Guidelines, see Alan Ellis & Mark Allenbaugh, *Unwarranted Disparity: Effectively Using Statistics in Federal Sentencing*, 12 WHITE COLLAR CRIME REP. 63 (2017), https://alanellis.com/wp-content/uploads/2017/04/effectively-using-statistics-in-federal-sentencing.pdf.

179. *See supra* Part II.

180. Neil Hutton, *Sentencing, Rationality, and Computer Technology*, 22 J.L. & SOC'Y 549, 558 (1995).

181. For a discussion regarding the contours of a principled sentencing system, see Mirko Bagaric & Sandeep Gopalan, *Saving the United States from Lurching to Another Sentencing Crisis: Taking Proportionality Seriously and Implementing Fair Fixed Penalties*, 60 ST. LOUIS U. L.J. 169 (2016).

182. Hutton, *supra* note 180, at 552.

Sentencing has a number of objectives, including deterrence and rehabilitation, but community protection has been the paramount aim in the United States for the past few decades.[183] Accordingly, the key consideration that informs the in-or-out-of-prison decision and the imposed prison term length is an assessment of the likelihood that the defendant will commit a serious offense.

A number of different techniques have been used to determine a defendant's level of risk of reoffending. The first is unstructured clinical assessment, where an individual assessor determines the offender's risk of reoffending according to impressionistic criteria without empirical validation.[184] This approach is the least reliable, and, because of the subjectivity associated with this approach, it cannot be built into an algorithmic system. The second technique for predicting risk of reoffending involves actuarial-based assessments.[185] Such mechanisms are often termed "risk assessment" tools[186] because they measure an individual's chances of endangering public safety generally by using actuarial methodologies to identify variables that contribute to recidivism.[187] Developers of "actuarial instruments manipulate existing data in an empirical way to create rules. These rules combine the more significant factors, assign applicable weights, and create final mechanistic rankings."[188] Although these sorts of tools are relatively new, their design concepts are well-established. As Berk and Hyatt note:

> Forecasting has been an integral part of the criminal justice system in the United States since its inception. Judges, as well as law enforcement and correctional personnel, have long used projections of relative and absolute risk to help inform their decisions. Assessing the likelihood of future crime is not a new idea, although it has enjoyed a recent resurgence: an increasing number of jurisdictions mandate the explicit consideration of risk at sentencing.[189]

---

183. *See, e.g.*, NAT'L RSCH. COUNCIL, THE GROWTH OF INCARCERATION IN THE UNITED STATES: EXPLORING CAUSES AND CONSEQUENCES 7 (Jeremy Travis, Bruce Western & Steve Redburn eds., 2014) (discussing incapacitation of criminals as a guiding principle for policymakers since the 1970s).

184. Christopher Slobogin, *Risk Assessment*, *in* THE OXFORD HANDBOOK OF SENTENCING AND CORRECTIONS 196, 198 (Joan Petersilia & Kevin R. Reitz eds., 2012); *see also* Jordan M. Hyatt & Steven L. Chanenson, *The Use of Risk Assessment at Sentencing: Implications for Research and Policy* (Vill. Univ. Charles Widger Sch. of L., Working Paper No. 193, 2016), http://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps (identifying the subjective, clinical approach to risk assessment as the "first generation" of risk assessment instruments).

185. Shawn Bushway & Jeffrey Smith, *Sentencing Using Statistical Treatment Rules: What We Don't Know Can Hurt Us*, 23 J. QUANTITATIVE CRIMINOLOGY 377, 378, 384 (2007).

186. Michael R. Davis & James R. P. Ogloff, *Key Considerations and Problems in Assessing Risk for Violence*, *in* PSYCHOLOGY AND LAW: BRIDGING THE GAP 194–96 (David Canter & Rita Žukauskienė eds., 2008); Pari McGarraugh, *Up or Out: Why "Sufficiently Reliable" Statistical Risk Assessment Is Appropriate at Sentencing and Inappropriate at Parole*, 97 MINN. L. REV. 1079, 1093–94 (2013).

187. McGarraugh, *supra* note 186, at 1091.

188. Melissa Hamilton, *Back to the Future: The Influence of Criminal History on Risk Assessments*, 20 BERKELEY J. CRIM. L. 76, 92 (2015).

189. Richard Berk & Jordan Hyatt, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, 27 FED. SENT'G REP. 222, 222 (2015).

The third system that is used to predict offenders' recidivism are termed "risk and needs assessments" tools.[190] These sorts of instruments are more fulsome than risk assessment tools. Risk assessments measure an offender's chances of reoffending and thereby endangering the public.[191] On the other hand, risk and needs assessments seek to reduce an offender's risk of recidivism by determining strategies and steps that would make reoffending less likely. Risk and needs assessment tools rely on a technique called "structured professional judgment,"[192] which aims to develop a needs assessment and a risk management plan.[193]

Many jurisdictions today rely on risk and needs assessment tools when determining "an individual's likelihood of recidivism for bail, sentencing, and parole decisions."[194] The extent of the use of such tools is noted by Amy Cypher:

> As of 2015, over sixty different risk assessment tools were used in the sentencing context alone, and more were used for bail determinations and by corrections officials. A 2010 Vera Institute for Justice survey found that "[a]lmost every state uses an assessment tool at one or more points in the criminal justice system," and that "over 60 community supervision agencies in 41 states reported using an actuarial assessment tool."[195]

The main differences among the various assessment tools are the variables that are incorporated into the programs and the emphasis accorded to the respective variables. These systems catalog "life facts" that can be used to determine an individual's propensity to commit crimes, which directly impacts the sentence a person may receive.[196] An offender's criminal history is a key determinant,[197] and other important variables include an offender's criminal associates, pro-criminal attitudes, and antisocial personality.[198] For example, one of the most sophisticated tools of this sort is the Post-Conviction Risk Assessment ("PCRA"), an instrument currently used for probation assessments in the United States federal jurisdiction.[199]

190. NATHAN JAMES, CONG. RSCH. SERV., R44087, RISK AND NEEDS ASSESSMENT IN THE CRIMINAL JUSTICE SYSTEM 1–2 (2015).

191. McGarraugh, *supra* note 186, at 1091.

192. Slobogin, *supra* note 184, at 199.

193. *Id.*

194. Seo, *supra* note 18; *see also* Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans* 7 (IZA Inst. of Lab. Econ., Discussion Paper No. 12853 Dec. 2019), https://ftp.iza.org/dp12853.pdf (noting that risk assessment tools are used for sentencing in at least twenty-eight states).

195. Amy Cyphert, *Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk,* 51 Seton Hall L. Rev. 331, 338 (2020).

196. Seo, *supra* note 18.

197. Hamilton, *supra* note 188, at 89.

198. *Id.* at 90.

199. Other assessment tools are: COMPAS – Correctional Offender Management Profiling for Alternative Sanctions; LSI-R – Level of Service Inventory – Revised; LSI/CMI – Level of Service/Case Management Inventory; LS/RNR – Level of Service/Risk, Need, Responsivity; ORAS – Ohio Risk Assessment System; Static-99 (for sex offenders/ offenses only); STRONG – Static Risk and Offender Needs Guide; and, Wisconsin

The PCRA is the latest (fourth generation) predictive tool;[200] it factors in not only static factors such as prior criminal history, but also dynamic variables such as employment status and history, education, and family relationships.[201]

As the Brennan Center for Justice noted, "states are increasingly turning toward risk assessment tools to help decide how much time people should spend behind bars."[202] Sentencing is an area where the use of risk assessment is controversial "because of the high stakes of criminal sentencing and because sentencing decisions are informed by many other goals beyond simply incapacitating those at high risk of reoffending . . . ."[203] This has not dented the use of such tools in sentencing decisions,[204] although judges have generally applied the tools in a non-systematic manner that does not strongly drive sentencing outcomes.[205]

Prosecutors and defense attorneys have a significant role in incorporating risk assessments into the sentencing process because "heavy caseloads often restrict a judge's ability to spend time finding appropriate alternatives to incarceration on a case-by-case basis."[206] A study examined the attitudes of prosecutors and defense attorneys in Virginia toward the use of risk assessment tools.[207] According to the study, prosecutors and defense attorneys each view the use of risk assessments for sentencing differently because each one has different obligations and different conceptions of justice.[208] Prosecutors, for example, favored the use of risk assessment tools for sentencing, arguing they "were likely a more consistent and fair way than relying on intuition or personal experience."[209] On the other hand, defense attorneys "were consistently opposed to using future recidivism risk as a factor in sentencing," as tools measuring future recidivism were based on "group means" rather than individual ones.[210] Even though both sides had differing views on the use of

---

State Risk Assessment Instrument. Most of these are used for assessing post-sentencing correctional populations. Hyatt & Chanenson, *supra* note 184, at 4.

200. *See Id.*

201. Hamilton, *supra* note 188, at 91–92.

202. JAMES AUSTIN, LAUREN-BROOKE EISEN, JAMES CULLEN & JONATHAN FRANK, BRENNAN CTR. FOR JUST., HOW MANY AMERICANS ARE UNNECESSARILY INCARCERATED? 18–19 (2016). Judges often pay little regard to the results of risk assessment tools. As noted by Slobogin, in Virginia, 59% of defendants who were considered to be at low risk of reoffending by a risk assessment tool were still sentenced to prison. Slobogin, *supra* note 184, at 202; *see also* Simmons, *supra* note 51, at 966 (discussing the increasing use of formal risk assessment instruments by judges in various states).

203. Stevenson & Doleac, *supra* note 194, at 7.

204. Cade Metz & Adam Satariano, *An Algorithm that Grants Freedom, or Takes It Away*, N.Y. TIMES (Feb. 7, 2020), https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html.

205. They are most commonly used in Virginia, Missouri, and Oregon. Slobogin, *supra* note 184, at 202–03.

206. Anne Metz, John Monahan, Luke Siebert & Brandon Garrett, *Valid or Voodoo: A Qualitative Study of Attorney Attitudes Towards Risk Assessment in Sentencing and Plea Bargaining* 8 (Univ. of Va. Sch. of Law, Public Law and Legal Theory Paper Series 2020–25, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552018.

207. *See id.*

208. *See id.* at 4, 13.

209. *Id.* at 15.

210. *Id.* at 15–16.

risk assessment tools, both groups had "skepticism towards the NVRA instrument both in its validity as a predictive tool and in the factors it considers pertinent to recidivism risk . . . ."[211]

One of the newest and most sophisticated predictive algorithms is used in the context of the Formerly Incarcerated Re-enter Society Transformed Safely Transitioning Every Person ("FIRST STEP") Act. This legislation was enacted in December 2019 and aims to lower federal prison numbers[212] by reducing penalties for many cohorts of non-violent offenders—including drug offenders[213]—and making provision for the early release of such offenders.[214]

The FIRST STEP Act required the Attorney General to create a "Risk and Needs Assessment System," which evaluates prisoners' risk of reoffending and the programs that will assist them most to lower that risk.[215] To satisfy this requirement, the Department of Justice ("DOJ") has developed the Prisoner Assessment Tool Targeting Estimated Risk and Needs ("PATTERN"),[216] which has a number of important features. The risk assessment includes dynamic considerations which are susceptible to change, including conduct while incarcerated, rather than solely immutable matters such as the nature of offenses.[217] The algorithm expressly aims to be racially neutral and is routinely updated and re-validated.[218] The instrument incorporates fifteen factors; four are static and eleven are dynamic.[219]

The DOJ has already started implementing the FIRST STEP Act.[220] As of September 30, 2020, 3,705 offenders had received a sentence reduction, with the average sentence decrease being six years.[221] Further, over 1,000 inmates qualified

211. *Id.* at 18.

212. Press Release, Donald J. Trump, President, President Donald J. Trump Secures Landmark Legislation to Make Our Federal Justice System Fairer and Our Communities Safer (Dec. 21, 2018), https://www.presidency.ucsb.edu/documents/press-release-president-donald-j-trump-secures-landmark-legislation-make-our-federal.

213. Gina Martinez, *The Bipartisan Criminal-Justice Bill Will Affect Thousands of Prisoners. Here's How Their Lives Will Change*, TIME (Dec. 20, 2018, 4:21 PM), http://time.com/5483066/congress-passes-bipartisan-criminal-justice-reform-effort/.

214. Brandon Sample, *The First Step Act Bill Summary Explained: A Comprehensive Analysis*, SENTENCING.NET (Dec. 19, 2018), https://sentencing.net/legislation/the-first-step-act-2018-summary.

215. *Id.*

216. OFF. OF THE ATT'Y GEN., THE FIRST STEP ACT OF 2018: RISK AND NEEDS ASSESSMENT SYSTEM 43 (2019), https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/the-first-step-act-of-2018-risk-and-needs-assessment-system_1.pdf.

217. *Id.*

218. *Id.* at 63, 77.

219. OFF. OF THE ATT'Y GEN., THE FIRST STEP ACT OF 2018: RISK AND NEEDS ASSESSMENT SYSTEM – UPDATE 10 (2020) [hereinafter RISK AND NEEDS ASSESSMENT SYSTEM – UPDATE], https://www.bop.gov/inmates/fsa/docs/the-first-step-act-of-2018-risk-and-needs-assessment-system-updated.pdf. Dynamic factors include an offender's education score and convictions while in custody and number of work programs that have been completed. Static factors include the offender's criminal history and age.

Id. at 10–11.

220. *Id.* at 1.

221. U.S. SENTENCING COMM'N, FIRST STEP ACT OF 2018 RESENTENCING PROVISIONS RETROACTIVITY DATA REPORT 4 (2021).

for enrollment in a pilot program that aims to transition elderly and terminally ill offenders to home confinement.[222]

Thus, although many jurisdictions use risk assessment tools, others have long resisted using such tools in their trials. Roughly ten years ago, Pennsylvania's legislature decided there was a problem with the impact of judicial discretion and decided to start working on an algorithm that would predict whether a person was likely to commit a crime and whether a judge should obtain additional information in a given case.[223] The algorithm was implemented in 2020, but still many believe that biases on account of race, for instance, cannot be eliminated as desired because there are many proxies for race that make their way into the algorithm.[224]

D. Parole

Risk and needs assessment tools are used extensively for parole determinations in many states. The use of these instruments has increased rapidly over the past three decades.[225] Like bail, the main determinant for parole is the risk that a suspect will reoffend if they are released into the community. AI is used in a variety of ways for parole decisions. In some cases, algorithms are relied on to decide which offenders should be released from jail and which ones should remain incarcerated.[226] For instance, some jurisdictions use the Level of Services Inventory-Revised (LSI-R), "which predicts parole and supervised release success."[227] This program uses an algorithm that assigns predictive scores to help determine whether an offender should be released.[228] Questions asked may cover topics such as the "individual's criminal history, education, employment, financial problems, family or marital situation, housing, hobbies, friends, alcohol and drug use, emotional or mental health issues, and attitudes about crime and supervision."[229] In other cases, GPS monitoring may be used to track parolees.[230]

---

222. *See* DEP'T OF JUSTICE, NCJ 254799, FIRST STEP ACT IMPLEMENTATION FISCAL YEAR 2020 90-DAY REPORT 4 (2020), https://www.ncjrs.gov/pdffiles1/nij/254799.pdf.

223. Alan Yu, *Can Algorithms Help Judges Make Fair Decisions?*, WHYY (Feb. 20, 2020), https://whyy.org/segments/can-algorithms-help-judges-make-fair-decisions/.

224. *Id.*

225. *See* Robert Werth, *Theorizing the Performative Effects of Penal Risk Technologies: (Re)producing the Subject Who Must Be Dangerous*, 28 SOC. & LEGAL STUD. 327 (2019). In 1970, only Illinois used an actuarial instrument to determine illegibility for parole, but this has increased to use in 28 of the 32 states that had a parole system by 2004. *History of Risk Assessment*, BUREAU OF JUST. ADMIN., https://bja.ojp.gov/program/psrac/basics/history-risk-assessment (last visited Oct. 17, 2021).

226. *See* Meghan J. Ryan, *Secret Conviction Programs*, 77 WASH. & LEE L. REV. 269, 342 (2020).

227. Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1119 (2017).

228. *Id.*

229. *Id.* (citing BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 79–81 (2007)).

230. *See, e.g.*, James Baimbridge, *My GPS-Tracked Life on Parole*, MARSHALL PROJECT (Oct. 28, 2019, 10:00 AM), https://www.themarshallproject.org/2019/10/28/my-gps-tracked-life-on-parole.

IV. CRITICISMS OF ALGORITHMS AND AI IN THE CRIMINAL JUSTICE SYSTEM AND RESPONSES TO THE CRITICISMS

A. *Policing and Detection of Crime*

As we have seen, algorithms relating to policing generally focus on enhancing public safety. In some locations, predictive hotspot tools are used to determine "where future crime may take place or where future police interventions may have a positive impact."[231] Many of these tools rely on arrest and conviction data to predict future criminal acts, but it has been noted that the data fed to the systems "directly reflect[s] the allocation of law enforcement resources and priorities, rather than rates of criminal activity across the population."[232] Some of these tools are theoretical in nature, looking to understand repeat victimization, while others consider data that does not necessarily have a connection to criminal law, "such as the weather, or sociodemographic features of the geographic area" for which crime levels are predicted.[233] The use of such tools has raised concerns, including the danger of discriminatory policing, the exacerbation of the bias that accompanies these systems, the lack of transparency of these systems, and the potential invasion of privacy associated with relying on these systems, which are discussed below.[234]

1. Validity and Accuracy Concerns

In order to consolidate the use of predictive policing and potentially increase reliance on it, it is necessary to first establish the validity of the system and, if possible, to improve the system's accuracy.

The Partnership on AI ("PAI") issued a report outlining shortcomings associated with the use of risk assessment tools primarily within the pretrial context, though many of these shortcomings are not limited solely to that context.[235] According to PAI, challenges to using these tools effectively fall into three categories: "1. Concerns about the validity, accuracy, and bias in the tools themselves; 2. Issues with the interface between the tools and the humans who interact with them; and 3. Questions of governance, transparency, and accountability."[236] PAI noted that in many cases, toolmakers are primarily concerned with whether the tool functions accurately, i.e. whether the tool performs according to "an accepted baseline of correctness."[237] By focusing on accuracy, toolmakers do not determine whether the tool is performing reasonably, causing validity and bias to become much more

231. ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 34.
232. Barabas, *supra* note 11, at 85 (citing DELBERT S. ELLIOTT, LIES, DAMN LIES AND ARREST STATISTICS 8 (1995)).
233. ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 34.
234. *Id.* at 35–36.
235. PARTNERSHIP ON AI, *supra* note 113, at 3.
236. *Id.*
237. *Id.* at 13.

pressing issues.[238] Validity is determined based on context and incorporates broader ideals such as fairness, so determining a tool's validity in one context does not automatically mean it is valid in others.[239] In fact, the majority of experts consulted in the drafting of PAI's report indicated that "current risk assessment tools are not ready for use in helping to make decisions to detain or continue to detain criminal defendants without the use of an individualized hearing."[240] However, there was no consensus on whether such tools can be improved upon to allow for use in detainment or continued detainment.[241] Many civil rights organizations have also expressed particular concern over predictive policing through the use of algorithms and risk assessment tools. For example, seventeen such organizations joined together and listed several risks associated with such systems, including "a lack of transparency, ignoring community needs, failing to monitor the racial impact of predictive policing and the use of predictive policing primarily to intensify enforcement rather than to meet human needs."[242] Lack of transparency must be addressed, but this can be difficult in some cases—particularly if an algorithm is provided by a private company,[243] which occurs frequently with facial recognition systems.[244]

Another overarching concern with the use of these systems is that they will be viewed as automatically objective and infallible.[245] This is a challenge in the use of statistical decision-making tools, as "information presented by a machine is viewed as inherently trustworthy and above skepticism," which leads officials to "over-rely" on them.[246] In reality, these systems are created with some human input, and these inputs can allow biases to creep into the systems and further perpetuate existing unfairness and discrimination.[247] Even so, some are of the opinion

---

238. *Id.*

239. *Id.* at 14. As PAI noted, "[r]isk assessment tools should only be deployed in the specific context for which they were intended, including at the specific stage of a criminal proceeding and to the specific population for which they were meant to predict risk." *Id.* at 22.

240. *Id.* at 11.

241. *Id.*

242. *See* Završnik, *supra* note 91, at 624–25 (citing *Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations*, ACLU (Aug. 2016), https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice).

243. See *infra* Section C for a discussion on the lack of transparency in sentencing and suggestions for how to counteract the problem.

244. *See* ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 43–44 (noting that unlike other algorithms, "facial recognition technologies have not been developed in-house or in academic consortia . . . but effectively outsourced to the private sector," making oversight and transparency more difficult to achieve).

245. *See* Ryan, *supra* note 226, at 277, 302 (noting that "science and technology are not infallible" and that prosecutors frequently present "conviction program evidence as virtually infallible").

246. PARTNERSHIP ON AI, *supra* note 113, at 23 (citing M.L. Cummings, *Automation Bias in Intelligent Time Critical Decision Support Systems* (Am Inst. Aeronautics & Astronautics Meeting Paper No. 2004-6313, 2004), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.2634&rep=rep1&type=pdf). For evidence of the opposite phenomenon, see *id.* at 243 n.44.

247. *See id.* at 3.

that humans have little say over the outcomes produced by machines.[248]

Providers within the criminal justice system, such as police officers and judges, have other reasons for disliking such tools. Some believe that these tools "devalue experiential knowledge . . . ."[249] Some judges disagree with the factors considered when calculating risk scores, while others "find that education on the instrument is not currently sufficient."[250] According to a study on the Los Angeles Police Department ("LAPD"), patrol officers, detectives, judges, and prosecutors alike were skeptical of using such tools, "suggesting the tools failed to tell them anything they did not already know."[251] In some cases, judges had logistical trouble relying on risk assessments. At least one judge noted "'there is always a lot of delays in getting the reports from Pretrial Services. . . . Usually I only get them 15 minutes before the beginning of the hearings,'" which apparently was a well-known problem.[252] The same study showed that judges also "worried that predictive algorithms could be used by administrators and defense attorneys to directly compare their sentencing decisions, incarceration rates, and productivity."[253] Because of such skepticism, many in policing and in criminal courts have resisted the use of algorithms where possible.[254] In any case, the same study determined that the "implementation of predictive algorithms comes with unintended consequences in terms of discretionary power."[255] These algorithms are intended to diminish bias and improve oversight by eliminating discretionary power but, admittedly, this is not always the outcome.[256]

2. Privacy and Liberty Concerns

Another criticism of the increased use of AI in policing—especially relating to facial recognition and the automated monitoring of CCTV cameras—is that it will violate the right to privacy. It has been recently argued that non-carceral forms of

248. *See* ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 28 (citing campaign group Liberty's assertion that "'[w]hile the idea of having human involvement or oversight of an algorithmic decision-making process may sound reassuring, there is a lack of evidence as to our ability as humans to provide meaningful intervention over algorithms and decisions made by machines.'").

249. Brayne & Christin, *supra* note 1, at 617.

250. Metz et al., *supra* note 206, at 7.

251. Brayne & Christin, *supra* note 1, at 617; *see also* Puente, *supra* note 101 (noting that "numerous departments dump[ed] the software because it did not help them reduce crime and essentially provided information already being gathered by officers patrolling the streets.").

252. Brayne & Christin, *supra* note 1, at 618.

253. *Id.* at 616.; *see also* Metz et al., *supra* note 207, at 8 (noting that "[a] judge's adherence to the sentencing guidelines is reviewed upon the end of their term and can factor into their re-appointment.").

254. Brayne & Christin, *supra* note 1, at 618; *see also* Metz et al., *supra* note 206, at 7 (noting skepticism among judges considering the Non-Violent Risk Assessment (NVRA) in Virginia). Despite these concerns about algorithms, Brayne & Christin concluded that "the implementation of predictive algorithms was more strictly implemented in policing compared to the criminal courts," which may be the result of the fact that judges often have more discretion than police officers. Brayne & Christin, *supra* note 1, at 621.

255. Brayne & Christin, *supra* note 1, at 621–22.

256. *Id.* at 620–21.

monitoring constitute "punitive surveillance."[257] This is not an overwhelming obstacle to the greater usage of CCTV, particularly since the right to privacy itself is a contentious interest. Robert Post has lamented that "[p]rivacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all."[258] The basis for privacy is typically linked to the ideals of autonomy and dignity[259] but lacks a concrete justification.

The United States Supreme Court has acknowledged privacy as a legitimate interest.[260] However, it is far from an absolute right and is often subordinated to other competing interests, including in the context of the imposition of criminal punishment. In *Hudson v. Palmer*, the Court noted that it would not be possible to achieve many of the security objectives of prisons—such as prohibiting the introduction of drugs and weapons into prisons—if prisoners retained the right to privacy.[261] Thus, the right to privacy receives some legal recognition, but it is often impinged upon, frequently, without the need for a formal or established legal justification. This is demonstrated by the massive intrusions into privacy that have occurred over the past decade or so.

One popular subject in AI policing is AI monitoring of people. Using AI to monitor CCTV cameras or to operate facial recognition systems may increase monitoring of people, but it would only be a small increase to the amount of monitoring that already occurs.[262] Currently, CCTV technology is used to attempt to prevent the commission of crime and as an evidence-gathering tool when a crime is committed. Thus, if CCTV is being monitored in real time and a crime is occurring, the operator will typically do whatever is reasonably possible to prevent the crime, including notifying police or, where the technology is available, notifying the offender that the event is being viewed and recorded with the purpose of discouraging continuation of the conduct. When an offense is recorded by CCTV, this can also be used to assist in the detection and prosecution of the offender. Countless instances of crimes have been solved by police viewing CCTV footage of an event,

---

257. Kate Weisburd, *Punitive Surveillance*, 108 VA. L. REV. (forthcoming 2022), https://ssrn.com/abstract=3808657.

258. Robert C. Post, *Three Concepts of Privacy*, 89 GEO. L.J. 2087, 2087 (2001).

259. VICTORIAN LAW REFORM COMM'N, WORKPLACE PRIVACY: ISSUES PAPER 17 (2002), www.lawreform.vic.gov.au/sites/default/files/IssuesPaperfinal.pdf.

260. The right to privacy, at least so far as personal autonomy is concerned, has been mainly acknowledged in contexts relating to procreation and family relationships. *See generally* Lawrence v. Texas, 539 U.S. 558 (2003); Roe v. Wade, 410 U.S. 113 (1973); Griswold v. Connecticut, 381 U.S. 479 (1965).

261. 468 U.S. 517, 517–18 (1984); *see also* Williams v. Kyler, 680 F. Supp. 172, 172 n1 (M.D. Pa. 1986) (noting "[t]he law is clear that prisoners do not possess any privacy right in their cells under the Fourth Amendment").

262. There are already approximately 50 million CCTV cameras in use in America. *See The U.S. Has More Surveillance Cameras Per Person Than China, New Study Shows*, INVERSE (Dec. 9, 2019), https://www.inverse.com/article/61552-united-states-china-surveillance-cameras; Liza Lin & Newley Purnell, *A World With a Billion Cameras Watching You Is Just Around the Corner*, WALL ST. J. (Dec. 6, 2019, 1:00 AM), https://www.wsj.com/articles/a-billion-surveillance-cameras-forecast-to-be-watching-within-two-years-11575565402.

generally when the offender was unaware of CCTV filming, and identifying the offender after their image was screened in the mainstream media. Therefore, the incursion into the right to privacy that will stem from the *increased* monitoring of certain locations (which is likely to occur if automated CCTV is demonstrated as a means of significantly reducing crime) is no different in nature from existing limitations of this right. To the extent that the incursions are more frequent and targeted, the change could be readily justified by the common good that is achieved in reducing crime and the increased rate of detecting and prosecuting offenders.

Moreover, both facial recognition and automated CCTV observance are less intrusive, to some extent, than real-time viewing by a human being because the computer does not evaluate behavior. In the automated context, law enforcement officers will only observe the footage or view the facial image when a computer detects something suggesting that a crime is being committed or that an offender has been recognized. Thus, for the most part, individuals will be *potentially* observable, not constantly observed or monitored by law enforcement.

The rights to liberty, property, and bodily integrity (which are often violated by criminal acts) are, however, more powerful and have far stronger legal protections than the right to privacy. AI-directed policing likely will result in certain cohorts of people being more frequently arrested, searched, and stripped of their property as a result of searches following arrests. There is also concern that predictive policing instruments will increase "police presence in poor and minority communities by creating a ratchet effect [i.e., entrenching and perpetuating the practice]."[263] As we have seen, there have already been claims of unfairness, discrimination, and persecution leveled at this form of policing, and some suggest that it violates the Equal Protection Clause and the Fourth Amendment.[264] These are potentially strong objections, but they are not decisive if the algorithms are developed appropriately.[265]

Thus, it is important to put this objection into perspective. Police—without resort to algorithms—have been heavily criticized for targeting neighborhoods predominately occupied by lower-socio economic and racial minority groups.[266] This criticism has been forcefully leveled at many police departments and is suggested as one reason for the grossly disproportionate rate of arrest and imprisonment of Hispanic and Black offenders.[267] Thus, even if crime prevention and detection algorithms do result in police disproportionality policing non-white or lower socio-economic groups, this is unlikely to result in the advent of a new problem.

Further, and more importantly, a significant advantage of AI-directed policing compared to current practices is that every integer informing the algorithm is

263. Jouvenal, *supra* note 101.
264. Simmons, *supra* note 51, at 947, 972.
265. See *infra* Part V.
266. Simmons, *supra* note 51, at 976.
267. Paul Butler, *Starr Is to Clinton as Regular Prosecutors Are to Blacks*, 40 B.C. L. REV. 705, 708–09 & n.16 (1999); Davis, *Prosecution and Race*, *supra* note 34, at 30; Farbota, *supra* note 25.

consciously and deliberately prescribed, which allows the opportunity to evaluate the algorithms for group profiling to ensure that such disparity is not a design feature. This, of course, assumes that the workings of the algorithm are transparent or can be independently tested to demonstrate that they are not biased in their selection of suspects. Although algorithms are not always transparent,[268] it may still be possible to test the algorithm to ensure bias does not exist. Therefore, our view is that for AI to gain acceptance and legitimacy in the criminal justice sector, it must be established that it does not result in the discriminatory targeting of certain groups in the community. This is discussed further in Part V below.

B. Bail

Just as the use of risk assessment tools faces criticism in other areas of the criminal justice system, similar criticisms exist when it comes to their use in determining bail and bond. For instance, United States Senator John Kennedy authored a commentary stating that the use of such tools when determining bail and bond is "[r]eally . . . a dangerous collision of the poorly vetted cost cuts and socialist agendas that are sweeping this country," and equated them to using a "Magic 8 ball" in courtrooms.[269] In response, Douglas A. Berman took to his *Sentencing Law and Policy Blog*, calling Kennedy's commentary a parody, while admitting that the use of risk assessments is "an important modern criminal justice development that justifies much scrutiny . . . ."[270] Despite grounds for criticism, Berman noted that risk assessments are appealing because judicial discretion—without the use of data—may lead to questionable decisions.[271] Reliance on data could, in theory, help lead to more reasoned decisions.

A study reviewing Kentucky bail decisions found that judges responded to algorithmically-produced risk assessment scores differently, depending on where they were located.[272] Another study on Kentucky found that "judges were more likely to overrule the default recommendation to waive a financial bond for moderate-risk defendants if the defendants were black."[273] These results demonstrate the impact of the AI in the use of bail and the need to better educate individuals who

268. Secrecy relating to the algorithms has been defended on the basis that: "Police officials say they can't release some information about their predictive programs because of citizen privacy and safety concerns and because some data is proprietary. The programs are helping to reduce crime and better deploy officers in a time of declining budgets and staffing, they argue." Dave Collins, *Should Police Use Computers to Predict Crimes and Criminals?*, PHYS.ORG (July 5, 2018), https://phys.org/news/2018-07-police-crimes-criminals.html.

269. John Kennedy, *Sen. John Kennedy: Bail, bond decisions are being made today with algorithms – That puts your safety at risk*, FOX NEWS (Aug. 5, 2019), https://www.foxnews.com/opinion/bail-bond-decisions-algorithms-safety-risk-john-kennedy.

270. Douglas A. Berman, *Are pretrial risk assessment algorithms really part of "socialist agendas that are sweeping the country"?*, SENT'G L. & POL'Y BLOG (Aug. 5, 2019, 7:07 PM), https://perma.cc/FU8H-9V34.

271. *Id.*

272. *See* Simonite, *supra* note 160. For instance, judges in rural areas—with mostly white defendants—granted release without bond more frequently, while judges in urban areas—with a mixed group of defendants—did not significantly change their habits of release without bond. *See id.*

273. *Id.*

are interpreting the results of risk assessment tools, which should also help lessen the role of bias in the system.[274] While algorithmic tools for bail can better predict risk than judicial insight alone, the tools need to be regularly assessed against local data and scaffolded on properly resourced data infrastructure.

Even when algorithms are deemed successful, [275] there is still concern that they will "perpetuate racial disparities within the criminal-justice system."[276] In 2018, more than one hundred civil rights groups petitioned several jurisdictions to stop using such tools to make bail decisions.[277] Critics of such tools believe they are "'racial profiling 2.0,'" and they are particularly dangerous because they are in such widespread use.[278] However, there is no validation for such claims and certainly no evidence of bias greater than what already exists in the context of judicial bail decisions.

C. Sentencing

Although algorithms are increasingly used at the sentencing stage of proceedings, they do not have widespread support. There are numerous reasons for this, including skepticism regarding their utility. Megan T. Stevenson and Jennifer L. Doleac conducted a study examining the use of risk assessment tools in felony sentencing.[279] Their study looked at how judicial decisions are influenced by risk scores, primarily focusing on the use of nonviolent risk assessments in Virginia and the impact on individuals based on race and age.[280] Age is a focus because it "is one of the most important predictors of criminal activity and, accordingly, has large weight in almost every risk assessment currently in use."[281] Even though race is not an explicit factor considered, many factors correlate to race, and it impacts sentencing decisions whether or not a risk assessment tool is applied.[282] The study showed that young defendants were adversely impacted by the use of risk assessments, and, in some cases, judges relied on risk assessment tools during decision-making, leading to "an increase in racial disparities relative to judicial discretion alone."[283] The study set out to determine why use of a risk assessment in

274. *See id.* According to Harvard researcher Alex Albright, "[w]e should put as much effort into how we train people to use predictions as we do into the predictions." *Id.*

275. *See generally* Kramer, *supra* note 165 (examining a risk assessment algorithm used by New Jersey courts for bail determinations and its purported ability to reduce racial bias).

276. Madeleine Carlisle, *The Bail-Reform Tool that Activists Want Abolished*, ATLANTIC (Sept. 21, 2018), https://www.theatlantic.com/politics/archive/2018/09/the-bail-reform-tool-that-activists-want-abolished/570913/.

277. *Id.*; *see also* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1750 (2018) (describing opposition to risk assessment tools by other advocacy groups).

278. Carlisle, *supra* note 276. According to Carlisle, "[f]orty jurisdictions use the PSA," and other jurisdictions employ tools created by state governments or private companies. *Id.*

279. *See* Stevenson & Doleac, *supra* note 194, at 2.

280. *See id.*

281. *Id.*

282. *Id.* at 3, 4, 14.

283. *Id.* at 3–4.

nonviolent cases did not lead to a decline in recidivism.[284] The study also found that risk assessment in nonviolent cases did not lead to a decline in recidivism because judges often ignored the tool's sentencing recommendations if they believed the outcome should be different.[285] Judges also appeared to be less inclined to incarcerate young defendants, which could "curtail risk assessment's expected benefits."[286]

A key concern underpinning the use of these algorithms is their accuracy. Despite this, research suggests that risk assessment and risk and needs assessment tools are more accurate in predicting re-offending than evaluations made without the use of such tools.[287] The authors of a recent study, Julia Dressel and Hany Farid, cast doubt about the rigor of the COMPAS assessment tool, concluding that it was no more accurate than an online poll of people with no criminal or legal training.[288] This study utilized Amazon Mechanical Turk and recruited about 400 participants to predict recidivism using a sample of 1000 real defendants.[289] A

---

284. *Id.* at 4. The researchers proposed two hypotheses for the lack of decline. *Id.* The first hypothesis was that judges made "fewer prediction errors than previously believed." *Id.* The second hypothesis was that "criminogenic effects of incarceration for higher risk defendants may have effectively canceled out its incapacitative effects." *Id.*

285. *See id.* at 24–25. A statewide survey showed that of Virginia judges, "only half 'always' or 'almost always' consider the results of the nonviolent risk assessment in sentencing." *Id.* at 29.

286. *Id.* at 31, 36. Although some judges may be less interested in incarcerating young defendants, some programs are in place to create a risk score specifically for young individuals. In Bristol, England, algorithms are used "to identify the most at-risk youths in the city and review caseloads." Metz & Satariano, *supra* note 204. This effort was in response to budget cuts and an uptick in youth violence and crime. *Id.* The risk assessment tool compiles "crime data, housing information and any known links to others with high risk scores, and if the youth's parents were involved in a domestic incident. Schools feed in attendance records." *Id.* While community officials relying on this information have determined that the algorithm identifies at-risk youth correctly, the community is unsure how to use the data because "people can't be arrested simply because of the algorithm." *Id.*

The UK also relies on the Offender Assessment System ("OASys") to provide an assessment of risks and needs across the prison systems in England and Wales. ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM, *supra* note 120, at 48. OASys provides indicators of reoffending that have been statistically validated. Scoring of these indicators is conducted based on information about individuals, including their age, gender, and criminal history. *See id.* at 49.

Similar to their use in the United States, these scores are used by UK police systems to "inform the sentencing process," often via pre-sentencing reports. *See id.* at 52. The system appears to have "a strong research and reflection base that has long been considering issues such a[s] bias and on-the-ground deployment, but the longevity of these efforts is unclear." *Id.* at 55.

287. Cyphert, *supra* note 197, at 379; Edward Latessa & Brian Lovins, *The Role of Offender Risk Assessment: A Policy Maker Guide*, 5 VICTIMS & OFFENDERS 203, 212 (2010). Moreover, risk assessment tools are generally more accurate than predictions based solely on clinical judgment. *See* D.A. Andrews, James Bonta & J. Stephen Wormwith, *The Recent Past and Near Future of Risk and/or Need Assessment*, 52 CRIME & DELINQ. 7, 12–13 (2006); William M. Grove, David H. Zald, Boyd S. Lebow, Beth E. Snitz & Chad Nelson, *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 12 PSYCHOL. ASSESSMENT 19, 25 (2000). For a more skeptical view regarding the accuracy of such tools, see Erin Collins, *Punishing Risk*, 107 Geo. L.J. 57, 75 (2019). *But cf.* Christopher Slobogin, *A Defense of Modern Risk-Based Sentencing* (Vanderbilt Law Research Paper No. 18-52, 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3242257 (responding to various criticisms of risk-based sentencing, including that it is inaccurate).

288. *See* Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES 1 (2018), https://advances.sciencemag.org/content/4/1/eaao5580.

289. *See id.* at 4–5.

similar study by Zhiyuan "Jerry" Lin and his colleagues, however, found that there are situations in which an algorithm may outperform a human. For example, when the participants did not receive immediate feedback on how accurate their response was,[290] or they had to consider a significant number of useful factors,[291] humans did not perform as well. Although participants were not "competitive with the existing tools,"[292] the study uncovered two situations where humans could be as successful as a risk assessment tool, both of which involve providing more information to humans, either through feedback on their success or through enhanced predictive factors.[293] Overall, the study supported the conclusion that "algorithmic risk assessments can often outperform human predictions of reoffending."[294]

A different study involving 1.36 million pretrial detention cases showed that "a computer could predict whether a suspect would flee or re-offend better than a human judge,"[295] though there are debates within scholarly circles as to what constitutes "success"[296] and fairness[297] in the context of using AI in the criminal justice system.

Thus, risk assessment and risk and needs assessment tools are more accurate than unstructured judgments. Moreover, the rate of recidivism even amongst offenders who were deemed to have a high risk of reoffending was reduced when they participated in treatment programs recommended by risk and needs assessments.[298]

The main concern that has been levelled at risk and needs assessment tools is that they are biased in their design and application. Although algorithms are designed to discriminate, or discern information, they are not necessarily in tune with what is socially acceptable[299] and hence it is understandable that there would be some continued reluctance regarding reliance on them. Things that would normally be considered protected characteristics, such as gender, race, pregnancy status, religion, sexuality, and disability, all play a part in human decision-making processes. This suggests that algorithms trained using "past biased data" are likely to recreate the same biases in their decision-making processes, further exacerbating discrimination and unfairness.[300]

---

290. Zhiyuan Lin, Jongbin Jung, Sharad Goel & Jennifer Skeem, *The Limits of Human Predictions of Recidivism*, 6 SCI. ADVANCES 1 (2020).

291. *Id.* at 4.

292. *Id.*

293. *Id.* at 5.

294. *Id.*

295. Završnik, *supra* note 91, at 3 (citing Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendil Mullainathan, *Human Decisions and Machine Problems* (Nat'l Bureau of Econ. Rsch., Working Paper No. 23180, 2017)).

296. *Id.* at 4.

297. *See* Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 811 (2020).

298. JAMES, *supra* note 190, at 8.

299. *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 28 (2017).

300. *Id.*

In *State v. Loomis*,[301] the Supreme Court of Wisconsin held that as long as it is used properly and in combination with other factors, a circuit court was allowed to consider a risk assessment that accounted for immutable characteristics. Importantly, the Supreme Court of Wisconsin held that the COMPAS risk assessment should not be conclusive: "Risk scores may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community."[302] The court also recognized that sentencing meets a variety of goals, while COMPAS is only designed to address the risk of recidivism.[303] Lastly, the court included certain cautions that must be included to advise sentencing courts.[304]

Corbett-Davies et al. also take the view that the criticisms of algorithms stemming from the perceived problems (relating to accuracy and bias) associated with COMPAS are overstated. They draw a line between poorly-designed algorithms and well-designed algorithms, the latter of which "can mitigate pernicious problems with unaided human decisions."[305] Algorithms are not perfect, but "[i]t is misleading and counterproductive to blame the algorithm for uncovering real statistical patterns."[306]

Further, the private company that created COMPAS and some scholars indicated that "the overall predictive accuracy of COMPAS is similar across racial groups, making the algorithm itself ostensibly unbiased, even where *outcomes* based on the tool—such as who gets detained pretrial—differ systematically by race."[307] Additionally, even if the tool itself were unbiased, individuals may be faced with "high bail amounts and pretrial detention" because "[a]ll risk assessments make mistakes" that may lead to disproportionate application across individuals.[308]

Even though risk assessment tools may perform similarly across races, there is still evidence of disparity, particularly with systems that rely heavily on criminal history.[309] Despite racial disparities, the Center for Court Innovation did not argue that pretrial risk assessment tools should be abandoned.[310] Instead, the Center argued that taking on a "strategy of reserving pretrial detention only for defendants

301. State v. Loomis, 881 N.W.2d 749, 770–72 (2016).

302. *Id.* at 760 (citing PAMELA M. CASEY ET AL., NAT'L CTR. FOR STATE COURTS, USING OFFENDER RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: GUIDANCE FOR COURTS FROM A NATIONAL WORKING GROUP 14 (2011), https://www.ncsc.org/~/media/Microsites/Files/CSI/RNA%20Guide%20Final.ashx).

303. *Id.* at 768.

304. *See id.* at 769–70 (listing cautions).

305. Corbett-Davies et al., *supra* note 151.

306. *Id.*

307. SARAH PICARD, MATT WATKINS, MICHAEL REMPEL & ASHMINI KERODAL, CTR. FOR COURT INNOVATION, BEYOND THE ALGORITHM 3 (2019), https://www.courtinnovation.org/publications/beyond-algorithm.

308. *Id.* at 4.

309. *See id.* at 8 (noting that "people of color are likely to average longer criminal histories, increasing their average risk score").

310. *Id.* at 13.

facing serious, violent charges and using risk-based decision-making only with those charges" will help reduce "both unnecessary detention and . . . racial disparities."[311]

By introducing new technology into the system, risk assessment tools, though deemed useful in many situations,[312] may "strengthen preexisting societal biases and perpetuate arbitrary and subjective decisions . . . ."[313] Although risk assessment tools may be perceived as more objective than judges, questions arise regarding whether "tools of scientific discourse created by a value-laden human culture could ever be purely objective, dispassionate, and disassociated from the values endorsed by social process."[314] Some risk assessment tools may include questions "that take a presumptive and subjective attitude toward serious ethnic and social dilemmas," thereby reducing the objectivity normally associated with risk assessment tools.[315] Objectivity, as promised by the use of such algorithms, then becomes subjective and furthers injustice that already exists in society.[316]

As is further discussed in the reform proposals section of this Article, to prevent the recreation of existing biases, tools should be developed carefully with a focus on preventing the operation of factors that lead to indirect discrimination. This will minimize the potential for race and other immutable factors to influence the outcomes of risk assessment algorithms. The key to achieving this is to understand fully the types of considerations that can act as proxies for the coding of inappropriate traits and to exclude them from the design of the algorithm. Further, as noted by Christopher Slobogin, "enhancing the punishment of an offender because of gender, age, or any other immutable characteristic strikes some as grossly unfair."[317] However, if immutable traits are included, a coherent rationale needs to exist for how the trait justifiably influences sentencing decisions. This, too, has been noted by Slobogin:

> If sentences can be enhanced in response to risk, then neither society's nor the offender's interests are advanced by prohibiting consideration of factors that might aggravate or mitigate that risk simply because they consist of immutable characteristics. In any event, risk-based sentences are ultimately based on a prediction of what a person will do, not what he is; immutable risk factors

311. *Id.* at 14.

312. *See, e.g.*, Rahnama, *supra* note 6, at 175 (noting that empirical evidence "shows a positive trend in the contributions these tools can make to human decision-making.").

313. *Id.* at 177.

314. *Id.* at 175 (citing THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 128 (Chi. Univ. Press, 2d ed. 1962)).

315. *Id.* at 178 (citing Carlson, *supra* note 6, at 311). For instance, an algorithm may ask defendants whether they agree with a statement, such as: "A hungry person has a right to steal." *Id.* The concern with such questions is that they are not easily resolved using algorithms, and they make the algorithm "a mere embodiment of already 'deeply held cultural values and beliefs.'" *Id.* at 179 (citing Sheila Jasanoff, *The Songlines of Risk*, 8 ENV'T VALUES 135, 135 (1999)).

316. *Id.* at 178–79.

317. Slobogin, *supra* note 184, at 205.

> are merely *evidence* of future conduct, in the same way that various pieces of circumstantial evidence are not blameworthy in themselves.[318]

In addition to potential problems associated with accuracy and bias, another criticism of algorithms, including COMPAS, is their lack of transparency.[319] With regard to COMPAS specifically, the parent company is not required to share details of its programs with courts, and it has taken efforts to "'seal some of the details of their algorithm, and you don't know exactly how those scores are computed.'"[320]

Although algorithms are not perfect, criticisms of algorithms should not just compare them against "some perfect ideal, but also against the very imperfect status quo."[321] Accordingly, "[p]ublic officials have a social responsibility to pursue the opportunities that algorithms present, but to do so thoughtfully and rigorously."[322] As part of this effort, public officials must know how algorithms work and must expand research to improve accuracy and fairness.[323] Algorithms that "inappropriately combine data for all defendants" or use inaccurate measurements —i.e. comparing arrest for an offense to committing the offense—will produce bias in the system.[324] Many jurisdictions tackle these issues by focusing on the likelihood of arrest for committing a violent crime and the flight risk, as well as by using gender-specific risk models.[325] Although not all jurisdictions have taken these steps, this is more indicative of a problem "rooted in poor policy rather than the use of algorithms more generally."[326] Algorithms can help reform the criminal justice system, but they "must be carefully applied and regularly tested to confirm that they perform as intended."[327]

Although there are debates over whether risk assessment tools are useful in the criminal justice system, one area where they would almost certainly be beneficial is the drug court. According to a recent study,[328] the use of a decision support system ("DSS") in drug courts could lead to a variety of benefits, including

318. *Id.* Likewise, the court in *Malenchik v. State*, 928 N.E.2d 564, 574 (Ind. 2010), held that risk assessment tools used in sentencing could take into account offenders' immutable traits: Indiana Code § 35-38-1-9(b)(2) "mandates that pre-sentence investigation reports include 'the convicted person's history of delinquency or criminality, social history, employment history, family situation, economic status, education, and personal habits.' Furthermore, supporting research convincingly shows that offender risk assessment instruments, which are substantially based on such personal and sociological data, are effective in predicting the risk of recidivism and the amenability to rehabilitative treatment." *Id.*

319. Thompson, *supra* note 145.

320. *Id.*

321. Neufield, *supra* note 159.

322. *Id.*

323. *See id.*

324. Corbett-Davies et al., *supra* note 151.

325. *Id.*

326. *Id.*

327. *Id.*

328. Hamed M. Zolbanin, Dursun Delen, Durand Crosby & David Wright, *A Predictive Analytics-Based Decision Support System for Drug Courts*, 22 INFO. SYS. FRONTIERS 1323 (2019). This study relied on "over 11 years of real-world data obtained from 56 adult drug courts serving over 65 counties of a state in the south central region of the United States" and was geared toward finding a way to predict recidivism in drug courts. *Id.* at

"efficiency, transparency, consistency of decisions, and a reduction in the number of errors," while helping the "governments fight against the pandemic abuse of drugs more effectively."[329] DSS systems are valuable tools for making decisions in situations where unstructured decisions are needed.[330] Drug courts offer up a unique set of challenges, as involvement with an offender typically lasts for at least one year, and limitations on decision makers' ability to make rational decisions exist due to "constraints on information, cognition, memory, and time."[331] Further limitations exist due to sample size, failure to use cross-sampling to increase the sample size, and the fact that there are no set standards for determining how drug courts should handle cases.[332] As a result, "drug courts lack an easy-to-use and generalizable model to support decision making."[333]

Use of a DSS in the drug court setting also improves decisions and decreases the time needed to make such decisions by "relax[ing] cognitive, temporal, spatial, and/or economic limits on the decision maker."[334] Additionally, the DSS can "identify both relapsing and non-relapsing participants equally well . . . allow[ing] the officials to make decisions that are more effective for each individual."[335] More effective decisions, in turn, will lead to substantial cost savings and improvement of overall public safety and mental health.[336] Even with these improvements, those leading the study were clear that "DSS are not supposed to replace humans," and as a result, the DSS must be transparent because users will still be able to use their discretion when relying on the system.[337]

Thus, in the sentencing arena, there is increased use of AI. The main challenges to its use stem from perceived problems relating to the potential inbuilt biases of these applications. Hence, some judges still believe that algorithms are not the answer to resolving racial disparity in the courtroom; rather, racial disparity will be resolved by judges who are transparent and accountable for their biases and decisions.[338] However, the weight of evidence suggests that despite concerns about bias and racial disparity, the use of AI may lead to "immediate and tangible benefits" within the sentencing process—such as shorter sentences for those who have

---

1323. For the purposes of this study, recidivism was "defined as the violation of the treatment program requirements within three years after admission." *Id.*

329. *Id.* at 1323, 1328.

330. *See id.* at 1326.

331. *Id.* at 1325

332. *See id.* at 1326

333. *Id.*

334. *Id.* at 1328 (citing Clyde W. Holsapple, *DSS Architecture and Types*, *in* HANDBOOK ON DECISION SUPPORT SYSTEMS 1: BASIC THEMES 163 (Frada Burstein & Clyde W. Holsapple eds., 2008)).

335. *Id.*

336. *See id.*

337. *Id.*

338. *See* Yu, *supra* note 223.

committed nonviolent crimes and have low risk scores—with a goal of improving fairness across sentencing decisions.[339]

D. Parole and Probation

Concerns regarding bias embedded within risk assessment systems exist at all levels, including parole and probation. It has been suggested that these instruments increase prison numbers due to faulty design and user error:

> [Risk assessment instruments] establish an ontological order that precludes the possibility of a parolee who is not risky . . . . Acts of assessment disperse risk to everyone on parole; they produce all paroled subjects as risky of reoffending to some degree. In this way, it could be said that parole evaluation is somewhat of a false act of evaluation, or at least a predetermined and delimited one. Rather than querying whether or not someone is risky, assessments ask *how* risky is this person . . . . [340]

The ACLU and other groups protested the use of such tools in sentencing in Pennsylvania because of a fear that such use "would expand the power of predictive algorithms used for probation," leading to increased levels of racial bias.[341] Not only was bias a concern, but tools used during the probation process were negatively impacting probationers.[342] For example, in Philadelphia, if an individual on probation was charged with a new crime and was deemed "high risk," the probation office would notify the jail not to release the person.[343] Whether or not this activity actually occurred,[344] the alleged use of the system in this way is a problem according to the designer of the risk assessment tool, who noted that "'[probation officers] are hand tailored to a particular decision,' . . . . 'If you move them to another decision, the warranty doesn't apply anymore.'"[345] Many individuals caught in the probation system worry how the use of such risk assessment tools will continue to affect them, particularly when their life is being dictated by algorithms.[346]

The design of some of these instruments is less than optimal, and the results are contingent on the manner in which the instrument is used.[347] These disadvantages can be overcome by programming AI to combine formal definitions from risk

339. Corbett-Davies et al., *supra* note 151; *see also* Ryan, *supra* note 226, at 286 (arguing AI creates system-wide fairness by taking decision-making authority out of the hands of individual judges).

340. Werth, *supra* note 225, at 329.

341. Metz & Satariano, *supra* note 204.

342. *Id.*

343. *Id.*

344. The information regarding probation practices in Philadelphia came from an affidavit filed in a lawsuit and was denied by a spokesman for Philadelphia County. *See id.*

345. *Id.*

346. *See id.*

347. Sarah L. Desmarais, Kiersten L. Johnson & Jay P. Singh, *Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings*, 13 PSYCH. SERVS. 206, 207–8 (2016).

assessment tools (leaving no scope for user error or interpretation) and, where necessary, adjusting the weightings of the relevant variables based on machine learning. The integers that currently inform risk and needs assessment tools can be used as inputs to a supervised machine learning neural network. Then, using data from actual defendants who reoffended during or after the parole period, it is possible to use the machine learning system to build an accurate model showing which offenders will reoffend. This approach marries the benefit of assessment based on clear and specific factors together with the fast, statistical modeling that machine learning promises.

Jennifer Skeem and Christopher Lowenkamp undertook a study, published in 2016, analyzing probation risk assessments calibrated using PCRA in relation to nearly 35,000 offenders.[348] It was observed that the tool was accurate more than seventy percent of the time[349] and there was little evidence of bias:

> The instrument strongly predicts re-arrest for both Black and White offenders. Regardless of group membership, a PCRA score has essentially the same meaning, i.e., same probability of recidivism. . . . Second, Black offenders tend to obtain higher scores on the PCRA than White offenders (*d* = .34; 13.5% nonoverlap). So some applications of the PCRA might create disparate impact—which is defined by moral rather than empirical criteria. Third, most (66%) of the racial difference in PCRA scores is attributable to criminal history—which strongly predicts recidivism for both groups, is embedded in current sentencing guidelines, and has been shown to contribute to disparities in incarceration (Frase et al., 2015). Finally, criminal history is *not* a proxy for race. Instead, criminal history partially mediates the weak relationship between race and a future violent arrest.[350]

In a study reviewing the use of machine learning algorithms by the Pennsylvania Board of Probation and Parole ("Board"), the use of forecasts on recidivism "apparently had no effect on the overall parole release rate but did appear to alter the mix of inmates released."[351] According to the results, fourty-eight percent "of the inmates who had not been paroled were projected to be paroled had the forecasts been available."[352] As part of the study, an effort was also made to explain the importance of forecasts and how they can aid in making better parole decisions, which "might have encouraged board members to make better use of the usual

348. JENNIFER SKEEM & CHRISTOPHER T. LOWENKAMP, RISK, RACE, & RECIDIVISM: PREDICTIVE BIAS AND DISPARATE IMPACT 2 (2016). Because risk assessments have no impact on sentencing decisions in the federal system, Skeem and Lowenkamp did not examine the application of the PCRA on sentencing. *Id.* at 13.

349. *Id.* at 20.

350. *Id.* at 29; *see also* Sam Corbett-Davies, Emma Pierson, Avi Feller & Sharad Goel, *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not that Clear.*, WASH. POST (Oct. 17, 2016), https://www.washingtonpost.com/news/monkeycage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/.

351. Richard Berk, *An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism*, 13 J. EXPERIMENTAL CRIMINOLOGY 193, 193 (2017).

352. *Id.* at 205.

information provided as well as the new forecasts and reliabilities."[353] The use of forecasts led to differences in how the Board handled nonviolent crimes, but even with the addition of machine learning, forecasts did not change for violent crimes.[354]

Despite the successes of the study, the use of machine learning forecasts did not lead to a "substantial change in the overall proportion of individuals paroled," as the "overall proportion paroled dropped from 62% to 58%."[355] The forecasts, like many other algorithms, were only meant to supplement the Board's decision, and there was statistical evidence that "the Board made more accurate distinctions between inmates likely to be arrested for a *nonviolent* crime and inmates unlikely to be arrested for any crime."[356] Therefore, the use of the forecasts could be said to actually improve public safety by allowing the Board to make better decisions about inmates.[357]

V. REFORM RECOMMENDATIONS

As we have seen, the use of algorithms in the criminal justice system is increasing, but it also remains very controversial. This is in large part because of a fundamental misunderstanding of the workings of algorithms. It is also in part because many of the algorithms used in the criminal justice system are poorly designed. It is clear that algorithms can enhance criminal justice outcomes; however, in order to achieve this result algorithms must be transparent, fair, and effective. We now address these ideals in greater detail.

A. Transparency and Accountability

In order to inject greater confidence into the use of AI, it is necessary to produce publicly accessible algorithms that set out the factors considered in the algorithms and the weight that is attached to them. In the context of court-based decision-making, Eric Engle observed that "courts generally 'duck' the question of exactly how they weight the [varying] interests."[358] In principle, this is certainly achievable given the manner in which algorithms are designed. For example, introducing "modeling law by computer" could eliminate judicial discretion and discrimination while also articulating precisely how various interests are balanced in the decision-making process.[359] Indeed, Richard Susskind observed that AI-based systems are

---

353. *Id.*
354. *See id.* at 206.
355. *Id.* at 211–12.
356. *Id.* at 212.
357. *See id.* at 213.
358. Eric Engle, *Legal Interpretation By Computer: A Survey of Interpretive Rules*, 5 AKRON INTELL. PROP. J. 71, 93 (2011).
359. *Id.* at 73, 92–93.

"usually . . . transparent" by nature because they "can generate explanations of the lines of reasoning that lead them to their conclusions."[360]

In its report, PAI set out multiple requirements that would help make risk assessment tools more useful in the criminal justice system,[361] including that risk assessment tools and the people who use them should be transparent and accountable for decisions made,[362] adding further that the tools should undergo independent review as well as continuous evaluation and monitoring.[363] As part of this effort, each system must be accompanied by some indicia of reliability so that users can accurately predict the results.[364] Because there is potential for great uncertainty in these systems, the tools should disclose their measure of certainty—both generally and for specific individuals—as well as sources of potential uncertainty.[365] Risk assessment tools must also "remain transparent to citizens and accountable to the policy-making process."[366]

It has been suggested that as part of transparency, tool designs, data, and architecture should be made publicly available with redactions to protect individuals' privacy.[367] If models are created based on public data and by public companies, the public will know more about how they work within the criminal justice system.[368] Public access would also allow for more transparency and would make it easier to know when the algorithm makes mistakes, thereby injecting greater confidence in the use of such algorithms.[369] To this end, in late 2017, New York City passed the Algorithmic Accountability Bill to allow for more scrutiny of algorithms and

360. RICHARD SUSSKIND, TRANSFORMING LAW: ESSAYS ON TECHNOLOGY, JUSTICE AND THE LEGAL MARKETPLACE 183 (2000).

361. The requirements are: "Training datasets must measure the intended principles; Bias in statistical models must be measured and mitigated; Tools must not conflate multiple distinct predictions; Predictions and how they are made must be easily interpretable; Tools should produce confidence estimates for their predictions; Users of risk assessment tools must attend trainings on the nature and limitations of the tools; Policymakers must ensure that public policy goals are appropriately reflected in these tools; Tool designs, architectures, and training data must be open to research, review, and criticism; Tools must support data retention and reproducibility to enable meaningful contestation and challenges; Jurisdictions must take responsibility for the post-deployment evaluation, monitoring, and auditing of these tools": PARTNERSHIP ON AI, *supra* note 113, at 5.

362. *Id.*; *see also* Završnik, *supra* note 91, at 10 (noting that "[j]ust as biases can slip into the design of a statistical model, they can also slip into the interpretation—the interpreter imposes political orientations, values and framings when interpreting results.").

363. PARTNERSHIP ON AI, *supra* note 113, at 4.

364. *Id.* at 25.

365. *Id.* Potential sources of uncertainty include: "[u]ncertainty due to sample size and the presence of outliers in datasets"; "[u]ncertainty about the most appropriate mitigation for model bias"; and "[u]ncertainty as a result of sampling bias and other fundamental dataset problems." *Id.* at 25–26. The methods used to convey uncertainty to the user are also insufficient in many cases. *Id.* at 26.

366. *Id.* at 27.

367. *Id.* at 29.

368. Cynthia Rudin, *Algorithms and Justice: Scrapping the 'Black Box,'* CRIME REP. (Jan. 26, 2018), https://thecrimereport.org/2018/01/26/algorithms-and-justice-scrapping-the-Black-box/.

369. *Id.*

create a task force to evaluate how algorithms were used.[370] Other jurisdictions may use the New York Bill as a blueprint to increase algorithm transparency.

Another option for transparency and accountability is to create an audit trail—provided by an independent body—so that defendants who are impacted by risk assessment tools can contest the results.[371] Criminal defendants have the right to face their accusers, and for many, that includes the algorithm that made life-impacting decisions.[372]

A key impediment to transparency is the use of algorithms developed by private institutions.[373] There is a fear that the criminal justice system may actually increase reliance on algorithms produced by private companies in an effort to avoid disclosure of the uncertainties that exist within these tools.[374] Introduction of science and technology into the courtroom also exposes the government to a higher risk of liability, primarily because these tools have a real impact on the lives of defendants.[375] If private companies are able to successfully argue for trade secret protection, calls for transparency will be of limited success. To prevent this from occurring, it is imperative that the rights of defendants prevail over the interests of private companies, who will continue to contend that the protections for their trade secrets preclude disclosure.[376]

It is imperative that the program that drives recidivism-prediction risk assessment tools is fully transparent. This is the only manner in which we can have confidence that immutable characteristics are only included if it they have a demonstrable impact on risk and that some factors are not acting as proxies for other considerations, such as deprived social and economic background. This objective can only be achieved if the interests of defendants are placed above those of institutions that create algorithms.

It is possible to eliminate bias in the criminal justice system. But it requires a concerted focus on reducing discrimination in sentencing decisions and the development of integers which can achieve this result. This is demonstrated by recently documented reductions in racial inequality in the federal sentencing arena. Michael Light in a recently published survey notes that:

> Racial inequality in sentencing has decreased substantially over the last decade. In 2009, the average sentencing difference between black and white defendants in federal court was nearly 3 yrs. By 2018, this difference was less

370. *Id.*

371. PARTNERSHIP ON AI, *supra* note 113, at 30, 31; *see also* Hannah Sassaman, *Artificial Intelligence Is Racist yet Computer Algorithms Are Deciding Who Goes to Prison*, NEWSWEEK (Jan. 24, 2018, 9:49 AM), https://www.newsweek.com/ai-racist-yet-computer-algorithms-are-helping-decide-court-cases-789296 (suggesting that results of these algorithms should be audited).

372. *See* Sassaman, *supra* note 371.

373. *See* Rahnama, *supra* note 6, at 174, 184.

374. *See id.* at 183.

375. *See id.*

376. *Id.*

> than 6 mos. Among drug offenders over this same period, the black–white gap went from 47 mos. down to zero. Yet, despite the fact that racial inequality in the legal system remains at the fore of sociological discourse, these developments remain conspicuously underevaluated and the underlying processes driving these changes remain unknown.[377]

The need for transparent and valid risk and needs assessment tools is now gaining currency. The DOJ is already moving toward this objective with its PATTERN tool, which is discussed further below. In relation to the PATTERN tool, the DOJ noted its efforts to identify a group "to independently test and validate PATTERN in response to requests to publish the underlying PATTERN data to allow stakeholders to independently test its validity."[378] While the DOJ agreed with the need of stakeholders and advocates to "independently review and validate PATTERN . . . [and] has encouraged thoughtful criticism and input on how to improve" the program, the distribution of such data "is restricted because the data includes arrest and conviction information provided directly to DOJ by state and local jurisdictions, who have a significant interest in protecting their data."[379] Therefore, "[t]he retrieval, disclosure, and redistribution of that criminal history data is prohibited by the sharing agreements used to acquire the underlying data."[380]

Although agreements prohibit such disclosure, "DOJ arranged for external experts to evaluate and validate PATTERN."[381] Initially, the DOJ worked with Drs. Grant Duwe and Zachary Hamilton to perform statistical analysis of anonymized data which it used to develop the tool.[382] An independent review committee was also granted access to results of this analysis, as well as "reanalysis of PATTERN to help verify the assumptions, analysis, and conclusions incorporated."[383]

Researchers, on the other hand, could only access the information after completing required background investigations, and "[s]ince background checks are expensive and lengthy, in the absence of a contractual relationship, sharing the data widely is thus not a feasible solution to increasing transparency with outside stakeholders."[384] The National Institute of Justice ("NIJ"), in an effort to show commitment to addressing transparency concerns, intended to "solicit proposals for a five-year project to review and revalidate the new version of PATTERN."[385] Once background investigations are complete and relevant agencies grant permission to

377. Michael Light, *The Declining Significance of Race in Criminal Sentencing: Evidence from US Federal Courts*, SOC. FORCES, Mar. 2021, at 1.

378. RISK AND NEEDS ASSESSMENT SYSTEM – UPDATE, *supra* note 219, at 14.

379. *Id.*

380. *Id.*

381. *Id.*

382. *Id.* at 15. According to the report, DOJ continued to engage with Drs. Duwe and Hamilton, as well as stakeholders interested in improving the PATTERN tool. *Id.* at 1.

383. *Id.* at 15.

384. *Id.*

385. *Id.*

access the data, experts will "access and analyze data to validate PATTERN's reliability."[386]

B. Algorithmic Fairness

In addition to the need for transparency, the criminal court system must be fair, even when using algorithms. Some commentators have argued that algorithms carry a natural presumption of fairness. Thus, it has been observed that many people turn to these tools uncritically—that is, they rely on algorithms as providers of "objective facts, a neutral 'view from nowhere' that stands in stark contrast to the flawed, fickle, and opaque subjectivity of human decision makers."[387] It seems to follow that because these tools are objective, they can resolve any problems or inequities seen in human decision-making. Then, social issues become "data processing challenges, in which key decision makers . . . would benefit from tools that help them to distinguish 'signal from noise' when making time-sensitive decisions about potentially dangerous individuals."[388] Given the criticisms leveled at the use of algorithms in the criminal justice system, it is not clear that this sentiment has mainstream currency. In any event, the fairness necessary in the context of algorithms is substantive, not perceived, fairness.

In order for an algorithm to be fair, the programmer has "to define what [they] think is fair, then decide and code for the kind of fairness [they] want."[389] Algorithmic fairness is embodied by two approaches: "1) the development of formal fairness criteria that illustrate the trade-offs of different algorithmic interventions and 2) the development of managerialist 'best practices' for maintaining a baseline of accuracy, transparency and validity in algorithmic systems."[390] Many believe that "algorithmic fairness" is just an idea and that algorithms are fundamentally flawed because they cannot be as fair as humans.[391]

On the other hand, some believe that algorithmic fairness can be measured as long as the appropriate factors are considered.[392] For instance, Deborah Hellman argued that two primary measures of fairness are generally considered.[393] For one measure, "the score an algorithm produces should be equally accurate for members of legally protected groups."[394] This measure concerns one's beliefs, which are not the best way to gauge fairness.[395] For the second, "algorithmic fairness requires that the algorithm produce the same percentage of false positives or false negatives

---

386. *Id.* According to the report, NIJ expected to fund such research in 2020. *Id.*
387. Barabas, *supra* note 11, at 95 (quoting Donna Haraway, *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, 14 FEM. STUD. 575–99 (1988)).
388. *Id.* (citation omitted).
389. Yu, *supra* note 223.
390. Barabas, *supra* note 11, at 96.
391. Yu, *supra* note 223.
392. *See, e.g.*, Hellman, *supra* note 297, at 811, 812.
393. *Id.* at 811.
394. *Id.*
395. *Id.* at 830–31.

for each of the groups at issue."[396] If there is no parity between error ratios for false positives and false negatives, then groups that "have suffered disadvantage in the past" are particularly vulnerable to unfairness.[397] Hellman suggested that unfairness between groups may be mitigated by making algorithms more accurate, which could be accomplished by "us[ing] protected traits in a limited way to determine which other traits to consider within the algorithm."[398] Hellman acknowledged the "common assumption that antidiscrimination law prohibits the use of racial and other protected classifications in all contexts," but suggested that there may be "less of a barrier" to using this method to achieve accuracy and fairness because constitutional law that precludes disparate treatment on account of race does not explicitly rule out this technique.[399]

There are also those who believe that movements to reform the use of AI within the criminal justice system miss the point because they frequently focus on the narrow issue of bias.[400] The key question is not whether these tools themselves are biased, but rather "whether predictive tools reflect and reinforce punitive practices that drive disparate outcomes, and how data regimes interact with the penal ideology to naturalize these practices."[401] As Chelsea Barabas argued, this issue really comes down to how data is interpreted. With current interpretations through AI systems, "a specific penal ideology that justifies unwarranted punishment and harm" is ultimately reproduced.[402] Fairness is of course a complex notion but designing algorithms requires a clear understanding of the objectives that the instruments seek to achieve and the variables that are relevant to pursuing these aims. This provides the optimal occasion and environment for assessing, reflecting upon, and improving the substantive fairness of the instruments and minimally involves developing systems that do not discriminate against individuals. Recent advances in this area suggest that this is achievable.

A good illustration is the approach taken by the DOJ to eliminate bias in the PATTERN instrument. Its recent report on the tool noted agreement that neither inmates nor the community are well served by a risk assessment tool "that is racially biased or does not accurately reflect inmate risk."[403] Despite such agreement, the tool is unable to "correct for any outside biases that lead to higher recidivism."[404] Even so, it is important that the tool "avoid adding or exacerbating any bias that may exist. Critically, it must predict recidivism fairly, accurately, and objectively with the available information."[405]

---

396. *Id.* at 811.
397. *Id.* at 865.
398. *Id.*
399. *Id.* at 812, 846, 850–52.
400. *See, e.g.*, Barabas, *supra* note 11, at 84.
401. *Id.* at 83.
402. *Id.* at 85.
403. RISK AND NEEDS ASSESSMENT SYSTEM – UPDATE, *supra* note 219, at 8.
404. *Id.* at 9.
405. *Id.*

Accordingly, DOJ removed or changed scientific factors "that might be associated with bias, especially racial bias, in order to implement the most fair and predictive tool possible" to determine an inmate's risk of recidivism.[406] Two factors were removed from PATTERN: (1) age of first arrest/conviction; and (2) voluntary surrender.[407]

The change decreases the tool's predictive accuracy by approximately one percent, which was acceptable as long as it "prevents the actual or perceived perpetuation of any bias."[408] The DOJ also faced criticism that including "supervised release violations" as a factor "would disproportionately affect African Americans and Hispanics because . . . they are more likely to have their supervised release revoked due to biases in the criminal justice system."[409] The DOJ researched the issue, finding "that the opposite was actually true—removal of supervised release violations from the data would actually increase the potential racial disparity and have a negative impact on the predictability of the tool."[410]

A recent assessment of the PATTERN tool concluded that:

> The PATTERN tool includes certain best practices in recidivism prediction, and its developers have made a good faith effort to engage advocates and scholars about the tool's development. But much remains to be done if PATTERN is to truly represent an advance in algorithmic recidivism prediction. The DOJ must release more information about PATTERN and its underlying datasets so that scholars, advocates, and community members can truly assess the tool and offer responses for improving it.[411]

The ongoing analysis and presumably refinement of this instrument has the potential to provide the framework for further enhancing algorithmic fairness.

C. Enhanced Predictability and Consistency

Another advantage of computerized decision-making is that it necessarily makes decision-making more consistent and predictable—assuming that relevant integers are transparent. Hutton has noted that "one of the main aims of using computer technology to support sentencing has been to make the sentencing process more formal and more rational," and thereby to "reduce disparities" by ensuring that sentencing decisions are consistent with one another.[412] Computerized decision-making has the potential to achieve consistency between sentences imposed on offenders for similar crimes. Feelings, emotions and subjective preferences cannot influence computerized decision-making. And, as Susskind observes, "computer

406. *Id.*
407. *Id.*
408. *Id.*
409. *Id.*
410. *Id.*
411. Cyphert, *supra* note 197, at 381.
412. Hutton, *supra* note 182, at 558.

systems will not suffer from 'off-days' that so often inhibit the performance of human beings."[413] Indeed, lacking human irrationality, there is no reason for computers to deviate from a consistent approach to decision-making. This sentiment applies not only to sentencing decisions but also to algorithms at every stage of the criminal justice system.

Moreover, if algorithms are to have a greater role in the criminal justice system, it is important that they assist, not replace, human-decision making. The PAI correctly warned against using such tools alone "to make decisions to detain or to continue detention," (though the organization appears to support potentially using such tools "to facilitate the automatic pretrial release of more individuals").[414] Thus, the results of algorithmic decisions should operate in a recommendatory manner, as indeed does the automatic pilot function. However, as confidence grows in the accuracy and efficacy of criminal justice algorithms, it is most likely that as is the case with automatic pilots, it would be rare for human decision makers to make a different decision.

CONCLUSION

The criminal justice system interacts heavily with fundamental human rights and interests. Mistakes in the administration and operation of the criminal justice system can have profoundly adverse effects on defendants, victims, and the wider community. It is vital that decisions in this area be transparent, accurate, and effective.

Decision-making in all areas of the criminal justice system is sub-optimal. It is marked by a high level of subjectivity and lack of accountability. Thus, at the policing stage, there is no proven methodology for allocating resources and efficiently preventing crime. Bail, sentencing, and parole decisions involve crude, subjective judgments regarding the likely future behavior of defendants. In all of these areas, the subconscious sentiments of decision-makers result in discrimination against already disadvantaged groups—especially Black people.

Hence, it is not surprising that artificial intelligence is being used in some parts of the criminal justice system. The roll out of this technology has been ad hoc and/or systematic, and its use varies greatly from jurisdiction to jurisdiction. However, it is sufficiently widespread for considered analysis to be undertaken regarding its efficacy. The empirical testing of AI systems in the criminal justice system, on balance, establishes that they enhance the integrity and rectitude of decision-making. At the same time, most commentators have not been positive towards the use of AI in this realm. There are several reasons for this, including: the fact that there are some poorly designed algorithms in use; the lack of transparency of some programs; the instinctive aversion that many people have towards computer decision-making; and, a misunderstanding of the fundamental workings of algorithms.

413. SUSSKIND, *supra* note 360, at 173.
414. PARTNERSHIP ON AI, *supra* note 113.

The Article rebuts these criticisms and suggests a model for the adoption of a greater use of AI in the criminal justice system. In particular, it attempts to break down the resistance to AI in the criminal justice system by making comparisons to other widely-accepted algorithms many of us use in our lives. This is an important ingredient to securing greater acceptance of algorithms in the criminal justice system. The structural design of criminal justice algorithms is identical to those used in other fields. Algorithms in these other fields are not perfect but they are superior to judgments made by humans. The same situation applies in the criminal justice system. In particular, the most common and compelling criticisms leveled at algorithms in this arena, bias, is far more prevalent in decisions made by humans. The key to securing greater receptivity and efficacy of the AI in all areas of the criminal justice system is ensuring greater transparency regarding the design of the algorithms and explaining their operation to users and the public.