Florida State University Law Review

Volume 47 | Issue 4 Article 7

Summer 2020

A Framework for the Efficient and Ethical Use of Artificial Intelligence in the Criminal Justice System

Dan Hunter

Mirko Bagaric

Nigel Stobbs

A FRAMEWORK FOR THE EFFICIENT AND ETHICAL USE OF ARTIFICIAL INTELLIGENCE IN THE CRIMINAL JUSTICE SYSTEM

DAN HUNTER, MIRKO BAGARIC, AND NIGEL STOBBS*

Machine learning techniques are transforming the manner in which much of the legal system works, and criminal justice is the area which will be most fundamentally changed. Given the fundamental rights and interests at stake in the criminal justice system, this is also the field where the unthinking application of artificial intelligence ("AI") is most troubling, and where there is the greatest threat to individual rights and the likelihood of unanticipated damage to the rule of law. These problems will occur (and are occurring) throughout the criminal justice system: from data-driven predictive policing systems in the criminal investigation process, through to recidivism prediction for parole applications and sentencing recommendation systems post-trial. The risks presented by AI to the proper functioning of the criminal justice system will be exacerbated by commercial pressures on law enforcement and the criminal justice system, partisan political interests, and a lack of technological understanding by the judiciary and the legal profession more generally. Notwithstanding this dystopian vision, there is an opportunity to use AI techniques to improve the detection of crime, prosecute and sentence criminal offenders, help uncover discrimination, ensure parity of treatment across the system, and identify unfair and unjust treatment. The thoughtful and appropriate use of "ethical" AI systems can greatly assist in the administration of justice and the rule of law. In this Article, we propose a framework for systematically implementing AI into the criminal justice system in order to ensure that the system operates in a normatively enhanced and more effective and efficient manner. In proposing this framework we grapple with the reality that humans have an intrinsic emotional dislike of computers making decisions that have an important impact on peoples' lives.

* Professor Dan Hunter, Executive Dean, Faculty of Law, Queensland University of Technology, Australia; Professor Mirko Bagaric, Director, Evidence-Based Sentencing and Criminal Justice Project, Swinburne Law School, Australia; Dr. Nigel Stobbs, Senior Lecturer, QUT Law School, Australia.

I. INTRODUCTION

Advances in artificial intelligence, machine learning, and big data promise to transform the legal and judicial process. Over the last five years, machine-learning-based AI methods have made it possible to build autonomous decision-making systems that are derived from, and mimic, human behavior.[1] This is most obvious in the development of self-driving cars—which are in essence autonomous systems that pilot large hunks of metal around at great speed, based on billions of past decisions made by human drivers. These technologies are now finding their way into all areas where there are large datasets of previous decisions, and law is, of course, one of those fields.

Scholarship in AI and law is well established, stretching back to seminal work in automating US taxation law decisions by Thorne McCarty in 1972.[2] However, the initial AI and law research, and the dominant paradigm up until as recently as five or ten years ago, was in symbolic systems. These approaches represent law as rules, cases, or arguments within the computer, and decisions from these systems are understandable by humans. The more recent work in deep learning systems, also known as layered or convolutional neural

1. *See, e.g.*, Tad Friend, *How Frightened Should We Be of A.I.?*, THE NEW YORKER (May 7, 2018), https://www.newyorker.com/magazine/2018/05/14/how-frightened-should-we-be-of-ai [https://perma.cc/53H4-EJSV]; Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977, 977 (2017); Sam Corbett-Davies et al., *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 [https://perma.cc/5VE8-QH6Q].

2. *See* Ric Simmons, *Quantifying Criminal Procedure: How To Unlock The Potential Of Big Data In Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947, 957 (2016); L. Thorne McCarty, *Reflections on TAXMAN: An Experiment in Artificial Intelligence and Legal Reasoning*, 90 HARV. L. REV. 837, 837 (1977); STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 17 (2d ed. 2003).

networks, have used huge datasets to model intelligent behavior.[3] Not only has this meant a revolution in the accuracy and autonomy of AI software, it has also created systems that behave in ways that are clearly intelligent, but not in a human way.

These systems promise to transform all areas of law, but the field where data-driven AI will change the law most obviously, and most quickly, is in the criminal justice sector. This is true due to a range of economic, technical, and social factors examined below, but the core insight is this: the criminal justice system is largely grounded in making predictions of human behavior. As Ric Simmons notes:

> The criminal justice system has always been concerned with predictions. Police officers on patrol predict which suspects are engaged in criminal activity in order to determine where to focus their investigative efforts. Magistrates deciding whether to grant a search warrant predict the odds that contraband will be found based on the facts presented in a warrant application. Judges conducting bail hearings predict the chances that a defendant will return to court for trial, and sentencing judges try to determine whether a convicted defendant is likely to reoffend if he is given a nonincarceration sentence.[4]

Making predictions based on data about prior decisions is precisely what modern AI systems are best at, so criminal law is a particularly ripe area for the application of AI systems.

Historically, and to this day, predictions of future human behavior have been based on crude, generalized, and non-tested assumptions:

> Since the inception of our criminal justice system, law enforcement officers and judges have relied primarily on experience, training, intuition, and common sense in making their predictions. In response, courts have crafted broad standards to accommodate these subjective judgments and allow for flexibility in application. For example, police officers may briefly detain an individual if they reasonably believe that "criminal activity may be afoot," while magistrates should issue a warrant if "a man of prudence and caution [believes] that the offense has been committed."[5]

The courts have deliberately left these standards flexible due the enormous range of considerations and variables that impact criminal matters and because "police and courts have historically lacked the necessary tools to evaluate the accuracy of their predictions with

---

3. *See, e.g.*, Yann LeCun et al., *Deep Learning*, 521 NATURE 436, 436 (2015). Hereinafter, we will use the terms "artificial intelligence," "AI," and "machine learning" synonymously with deep layered neural networks of various types. This is formally wrong in a range of ways, but for the purposes of this Article the differences are unimportant.

4. Simmons, *supra* note 2, at 948-49.

5. *Id.* at 949. (alteration in original).

any precision."[6] Accordingly, "state actors have been forced to rely on their own subjective beliefs and anecdotal evidence in making their predictions."[7]

This is true no longer. Data-driven machine learning systems will be applied to every aspect of the criminal justice system. In the pre-trial phase, data-driven techniques and AI systems are being applied to predict when and where crime will occur, and will be used to make decisions about whether to monitor, arrest, and search a suspect, and whether to charge or indict them.[8] Many of the early approaches in so-called "predictive policing" systems relied on uncleaned prior data that enshrined discriminatory treatment based on race and class.[9] This has been the basis of outraged and concerted commentary about the limits and dangers of predictive policing, and has been a seminal driver in the development of ethical AI and the movement to ensure fairness, accountability and transparency in machine learning.[10]

However, the application of machine learning is not confined to predictive policing. During the parole and sentencing phases of criminal matters, data-driven systems are currently being used to assess recidivism likelihood and will increasingly be used to provide guidance to judges in their sentencing process.[11] The likelihood of offending is also a key consideration at the bail stage of the criminal justice process. While bail, sentencing, and parole decisions occur at different stages of the criminal justice system and have different objectives, there is one key integer which plays a defining role at all of these stages in terms of determining whether a defendant will be imprisoned: community safety. In crude terms, this requires an assessment of whether there is a meaningful risk that the defendant will commit a serious offense in the foreseeable future. If there is a significant risk of this occurring, the defendant will likely be refused bail or parole; in the sentencing context, they will likely receive a lengthy prison term. Risk assessment tools, systems based on the reoffending patterns of other offenders and particular traits of the defendant,[12] are already used extensively in many states to inform parole decisions, and they are now increasingly being used in sentencing cases.[13] It is

---

6. *Id.* at 950.
7. *Id.*
8. *See infra* Part III.
9. Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. L. REV. (2016); Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. Online 15 (2019); Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237 (2015).
10. *See infra* Part III. *See generally* RAFAEL A. CALVO ET AL., SUPPORTING HUMAN AUTONOMY IN AI SYSTEMS: A FRAMEWORK FOR ETHICAL ENQUIRY (2019).
11. *See infra* Part IV.
12. *See infra* Part IV.
13. *See infra* Part IV.

in relation to these decisions that AI is likely to have the greatest role in the near future. In the context of bail, not only is AI likely to assess the risk that a defendant will commit an offense, but it will also predict the likelihood that he or she will abscond. Thus, all of the key variables that determine bail and parole outcomes will soon be determined by computers. In addition to this, there have already been calls for sentencing to be done automatically by AI systems.[14]

The vast potential for AI to be imbedded into the criminal justice system promises to provide enormous benefits to society, but also generates a range of troubling questions. As already noted, critics of predictive policing have raised concerns about the automatic encoding of systematic bias, the absence of transparency of the algorithms, and how we should ascribe liability for biased decisions. More generally, there is concern about the ethics of allowing machines to make automated decisions over people's lives. Of course, sentencing, parole, and bail decisions are political hot buttons.[15] Likely advances in the availability of data and access to AI systems will mean that political action groups, politicians, executives, and legislatures will be able to use recidivism and sentencing prediction systems to advance political agendas against judicial officers whom they see as too tough or—more likely—too soft on crime. This has serious implications for the judiciary and is likely to increase pressure on judges.

These are potentially difficult and worrying movements and there are numerous points of concern. However, there are a range of interventions that can be made to ensure a just system in a world dominated by AI and data. The thoughtful and considered application of this technology might make it possible to ensure fairness and parity of decision-making, making good on the constitutional guarantee of equal treatment under the Equal Protection Clause. But it will require a deep understanding of both the data and the algorithms to safeguard this. Indeed, AI can be used to control for some of the more troubling aspects of decision-making by law enforcement, prosecutors, and judges. There is a broad literature from cognitive science, social psychology, sociology, and criminology which shows that limitations in human decision-making can lead to numerous forms of injustice.[16] AI techniques have the potential to safeguard against this if properly deployed. Understanding the interactions

14. *See, e.g.*, Mirko Bagaric & Gabrielle Wolf, *Sentencing by Computer: Enhancing Sentencing Transparency and Predictability and (Possibly) Bridging the Gap Between Sentencing Knowledge and Practice*, 25 GEO. MASON L. REV. 653, 654 (2018).

15. *See, e.g.*, Mirko Bagaric et al., *Bringing Sentencing into the 21st Century: Closing the Gap Between Practice and Knowledge by Introducing Expertise into Sentencing Law*, 45 HOFSTRA L. REV. 785, 786–87 (2017).

16. *See, e.g.*, 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, 196 (2015) (discussing the application of this body of research to the limitations and risks inherent in an unchecked and opaque judicial discretion in determining sentence); Michele Benedetto Neitz, *Socioeconomic Bias in the Judiciary,* 61 CLEV. ST. L. REV. 137, 158–60 (2013).

between technology, decision-making, the justice system, and the wider systems of control are necessary to control the future that we face in an AI-driven world. The AI-dominated world which we are entering promises great benefits, if we can only understand the strengths and weaknesses of the systems and apply them accordingly.

In this next Part of this Article, we provide a technical overview of the workings of AI systems and discuss whether they are compatible with the workings and operation of the criminal law. We discuss some of the reason why the application of AI is so troubling—humans have an inherent aversion to automated decision-making. We examine whether this aversion is warranted, and how this might be addressed in the criminal justice context. After this introduction, we examine the implications of AI in each part of the criminal justice system. In Part III we examine how police can use AI to deter crime and detect and apprehend offenders. This Part examines the rise of predictive policing and discusses the benefits and detriments of these sorts of systems. Since these AI systems will inevitably be used in policing, we provide some recommendations for their use.

Part IV then considers the application of AI to bail, sentencing, and parole hearings. These obviously occur at different stages of the criminal justice system but they share one important commonality: the key consideration informing the outcome of these matters is an assessment of the defendant's likelihood of reoffending. Another reason for considering these stages of the criminal justice system jointly is that the main criticisms that have been levelled against the use of AI in the criminal law apply in all of these areas. Algorithms which predict the likelihood that a defendant will commit an offense have been heavily criticized on the basis that they are biased against certain minority groups and are opaque in their operation.[17] In Part IV we discuss whether these criticisms are justified and how they might be addressed. To the extent that there are important distinctive considerations at these stages of the criminal justice system, for example the likelihood that a defendant will abscond is an important consideration only in bail matters, we also assess the role that AI has in relation to these considerations. In Part V we conclude by laying out a framework for how AI should be incorporated into the workings of the criminal justice system in a manner where it facilitates more efficient and effective responses to crime, while ensuring that the system operates in a normatively sound manner. We summarise our recommendations in the concluding remarks.

17. *See infra* Part III.

II. ARTIFICIAL INTELLIGENCE AND CRIMINAL LAW

A. Defining Artificial Intelligence

Artificial intelligence is a notoriously slippery concept. A recent House bill used an inclusive definition, relying on a portmanteau of features:

> The term "artificial intelligence" includes the following:
>
> (A) Any artificial systems that perform tasks under varying and unpredictable circumstances, without significant human oversight, or that can learn from their experience and improve their performance. Such systems may be developed in computer software, physical hardware, or other contexts not yet contemplated. They may solve tasks requiring human-like perception, cognition, planning, learning, communication, or physical action. In general, the more human-like the system within the context of its tasks, the more it can be said to use artificial intelligence.
>
> (B) Systems that think like humans, such as cognitive architectures and neural networks.
>
> (C) Systems that act like humans, such as systems that can pass the Turing test or other comparable test via natural language processing, knowledge representation, automated reasoning, and learning.
>
> (D) A set of techniques, including machine learning, that seek to approximate some cognitive task.
>
> (E) Systems that act rationally, such as intelligent software agents and embodied robots that achieve goals via perception, planning, reasoning, learning, communicating, decision-making, and acting.[18]

Back in the 1980s, when one of us was first studying AI, the sardonic definition was that "AI is anything that computers can't do yet." However, the best definition is probably one that combines elements of the House definition above: it is a set of techniques within computer science, aimed at creating computer systems which can demonstrate behavior that is generally thought of as intelligent.

Artificial intelligence is a venerable discipline within computer science, born in 1956 at a conference at Dartmouth College.[19] The subdiscipline of artificial intelligence and law is nearly as old,

18. FUTURE of Artificial Intelligence Act of 2017, H.R. 4625, 115th Cong. § 3 (2017).

19. RUSSEL & NORVIG, *supra* note 2, at 17 (calling the conference the "birth of artificial intelligence").

starting at least as early as 1971[20] and operating continuously as a field since then, albeit with alternating periods of excitement and disillusionment.[21] The first highpoint for AI and law was during the eighties and nineties, a period of enormous apparent promise where researchers worked on legal expert systems that they hoped might provide legal advice that was better, cheaper, faster, and less prone to error than that of human lawyers.[22] The technology of the day involved what are called "symbolic systems,"[23] ones that rely on the symbolic representation of legal rules and cases that can be manipulated by various types of reasoning algorithms.

This early excitement waned, as these symbolic systems failed to live up to the hype. In part this was caused by some difficult jurisprudential problems, and by some path-dependent difficulties caused by the adoption of law as a domain by logic programmers who were interested in applying their techniques without really understanding legal reasoning.[24] But the AI winter[25] that lasted

20. *See* L. Thorne McCarty, *Reflections on TAXMAN: An Experiment in Artificial Intelligence and Legal Reasoning*, 90 HARV. L. REV. 837, 837 (1977) ("[t]he work on this project was begun while the author was a Law and Computer Fellow at the Stanford Law School, 1971–1973 . . . "). Layman E. Allen at Yale Law School (and later Michigan) had demonstrated the application of formal logic systems to the drafting of legal language, as early as 1957, although he did not use automated reasoning systems. *See, e.g.*, Layman E. Allen, *Symbolic Logic: A Razor-Edged Tool for Drafting and Interpreting Legal Documents*, 66 YALE L.J. 833 (1957); Layman E. Allen & Gabriel Orechkoff, *Toward a More Systematic Drafting and Interpreting of the Internal Revenue Code: Expenses, Losses and Bad Debts*, 25 U. CHI. L. REV. 1 (1957). There was a flowering of early interest in symbolic logic during the middle part of the 1970s. *See, e.g.*, Walter G. Popp & Bernhard Schlink, *JUDITH, A Computer Program to Advise Lawyers in Reasoning a Case*, 15 JURIMETRICS J. 303 (1975); Thomas Haines Edwards & James P. Barber, *A Computer Method for Legal Drafting Using Propositional Logic*, 53 TEX. L. REV. 965 (1975). For a comprehensive account of the history of the AI & Law movement, including the rise of symbolic logic systems, see Dan Hunter, Representation and Reasoning in Law: Legal Theory in the Artificial Intelligence and Law Movement (1995) (unpublished LLM thesis) (copy on file with author).

21. *See* Trevor Bench-Capon et al., *A History of AI and Law in 50 papers: 25 Years of the International Conference on AI and Law*, 20 ARTIFICIAL INTELLIGENCE & L. 215, 218 (2012).

22. *See, e.g.*, M.J. Sergot et al., *The British Nationality Act as a Logic Program*, 29 COMM. OF THE ACM 370 (1986); Alan Tyree et al., *Legal Reasoning: The Problem of Precedent*, *in* ARTIFICIAL INTELLIGENCE DEVELOPMENTS AND APPLICATIONS 231, 239-40 (J.S. Gero & Robin Stanton eds., 1988); KNOWLEDGE-BASED SYSTEMS AND LEGAL APPLICATIONS (T.J.M Bench-Capon ed., 1991).

23. *See, e.g.*, Michael Aikenhead, *The Uses and Abuses of Neural Networks in Law,* 12 SANTA CLARA COMPUTER & HIGH TECH. L. J. 31, 33 (1996).

24. Ending up, as machine learning folks would say, in a sub-optimal local minimum. A neat history is given in Philip Leith, *The Rise and Fall of the Legal Expert System*, 1 EUR. J.L. & TECH. 1 (2010).

25. The first AI winter came after the initial flush of success during the 1960s waned. The start of this first winter is often ascribed to the stinging conclusions of the UK's Lighthill Report, delivered in 1973. *See* James Lighthill, *Artificial Intelligence: A General Survey*, ARTIFICIAL INTELLIGENCE: A PAPER SYMPOSIUM (1973).

from the late 1990s until about 2010[26] was not confined to legal applications of AI, and came about largely as a response to the brittleness of symbolic systems, and the public perception that artificial intelligence was not creating anything that could really be called "intelligent."

Of course, these days there is an enormous amount of excitement and hype around AI. This is almost entirely due to the remarkable advances that have been made in one technology: deep neural networks, or "deep learning," as it is often called.[27] Although artificial neural networks have been around almost since the beginning of artificial intelligence,[28] the field exploded in 2012 when Krizhevsky, Sutskever, and Hinton demonstrated remarkable results in image classification and object recognition. Krizhevsky, Sutskever, and Hinton used large scale multi-layer, deep networks[29] based on Yann LeCun's earlier seminal work on convolution.[30] At that point, the combination of huge computational power and large datasets made machine learning practical, accurate, fast, and relatively inexpensive. Deep learning was suddenly front-page news,[31] and the hype has not diminished since then.[32]

---

26. *See, e.g.*, James Hendler, *Avoiding another AI winter*, 23 IEEE INTELLIGENT SYSS. 2, 2 (2008); *see also* ANDREAS HOLZINGER, ET AL., CURRENT ADVANCES, TRENDS AND CHALLENGES OF MACHINE LEARNING AND KNOWLEDGE EXTRACTION: FROM MACHINE LEARNING TO EXPLAINABLE AI 4 fig. 2 (2018); Kathleen Walch, *Are We Heading For Another AI Winter Soon?*, FORBES (2019), https://www.forbes.com/sites/cognitiveworld/2019/10/20/are-we-heading-for-another-ai-winter-soon/#2ae59a356d69 [https://perma.cc/3UXT-WB8J].

27. *See, e.g.*, Yoshua Bengio, *Learning Deep Architectures for AI*, 2 FOUNDATIONS AND TRENDS IN MACHINE LEARNING 1, 9 n.1 (2009); Gideon Lewis-Krause, *The Great A.I. Awakening*, N.Y. TIMES MAG. (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html [https://perma.cc/QQA5-9ZPL].

28. *See, e.g.*, FRANK ROSENBLATT, CORNELL AERONAUTICAL LABORATORY, THE PERCEPTRON A PERCEIVING AND RECOGNIZING AUTOMATON (1957), https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf [https://perma.cc/C6JX-2QFY]; Frank Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, 65 PSYCHOL. REV. 386 (1958); *Perceptron*, WIKIPEDIA (Jan. 14, 2020), https://en.wikipedia.org/wiki/Perceptron [https://perma.cc/3LRB-4JV8].

29. Alex Krizhevsky et al., *ImageNet Classification with Deep Convolutional Neural Networks*, 25 ADVANCES IN NIPS' OF THE CONF. ON NEURAL INFO. PROCESSING SYS. 2012 1097, 1097–1105 (2012). Similar work was being undertaken elsewhere. *See, e.g.*, Dan Cireşan et al., *Multi-Column Deep Neural Network for Traffic Sign Classification*, 32 NEURAL NETWORKS 333 (2012). The seminal review by the leaders in the field is Yann LeCun et al., *Deep Learning*, 521 NATURE 436 (2015).

30. YANN LECUN, GENERALIZATION AND NETWORK DESIGN STRATEGIES, TECH. REP. CRG-TR-89-4 (1989). The third genius behind the development of deep learning was Yoshua Bengio. Recently Hinton and LeCun were given the ACM's Turing Award, the "Nobel Prize of Computing." *See* ASSOCIATION FOR COMPUTING MACHINERY, https://amturing.acm.org [https://perma.cc/7QUQ-2KED] (last visited July, 19 2020).

31. *See* John Markoff, *Scientists See Promise in Deep-Learning Programs*, N.Y. TIMES (Nov. 23, 2012), https://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html [https://perma.cc/4FYM-JNUA].

32. *See e.g.*, Gideon Lewis-Krause, *The Great A.I. Awakening*, N.Y. TIMES MAG. (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html [https://perma.cc/2Z93-WRMG].

In order to understand the significance of deep learning to criminal justice, it is important to have a basic idea of how these types of approaches work.[33] At its core, deep learning is a statistical method for classifying patterns based on large amounts of sample data while using neural networks that have multiple layers. The networks are constructed with input nodes connected to output nodes via a series of "hidden" nodes, arranged in a series of layers. The input nodes can represent any data—in the examples of image recognition and speech recognition they involve pixels or words—and the outputs involve the decision or coding that the researcher is looking for, such as the picture classification 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 come 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.[34]

A simple criminal justice example may help. Imagine that we have a dataset that provides historical data on every sentencing decision for all criminal defendants in a given jurisdiction. This dataset contains all of the salient factors as inputs to the sentencing decision–the presence of mitigating factors like contrition or juvenile status and the presence of aggravating factors like recidivism or violence. Other factors would include the name of the judge, the nature of the crime, etc.—along with some presumably irrelevant considerations—for example, the time of day of the decision, the color of the defendant's clothes, and so on—along with the eventual sentence given for each case. The sentencing factors are the inputs on the network, and the sentencing determinations are the outputs. The network is initially coded with random activations and weightings so it cannot predict accurately the outcome of any case. But if we train it with hundreds of cases—or better, hundreds of thousands of cases—where we know both the factors and the sentences, then we will eventually have a fully

33. To be sure, there are a number of other connectionist approaches that differ somewhat from the supervised network described here (notably unsupervised and reinforcement algorithms). Yet, all of them are dependent on large datasets which generally present a set of inputs and outputs, and they all operate in ways that are similar enough within the legal domain that the differences need not detain us. For a detailed analysis of some of the general problems with deep learning and machine learning approaches, see GARY MARCUS, DEEP LEARNING: A CRITICAL APPRAISAL (R. Pfeifer et al. eds., 2018), https://arxiv.org/abs/1801.00631 [https://perma.cc/W4ZP-PBAY].

34. This is a process called "gradient descent." For a technical description of the process, see generally SEBASTIAN RUDER, INSIGHT CENTRE FOR DATA ANALYTICS, AN OVERVIEW OF GRADIENT DESCENT OPTIMIZATION ALGORITHMS (2017), https://arxiv.org/pdf/1609.04747.pdf [https://perma.cc/YF5Q-7KJ9].

trained network where the outcome of an undecided case can be predicted accurately based on the presence or absence of various inputs.[35]

Deep neural networks have made good on the promise that, one day, machines could actually learn. The areas where we see this most obviously are in machine vision and speech, and the headline applications of this are self-driving cars, voice recognition systems, speech production, and game playing. Other advances in semantic representation and analysis have tied neural networks to data systems like the web or music databases and have given us the miracle of Google's Pixelbuds earphones translating language on the fly, or Amazon's remarkable little cylinders queuing up your favorite song when you say, "Alexa, play some music that I like."

B. Algorithmic Aversion

AI has been applied to criminal justice for almost as long as legal researchers have had access to computers. An early example of this from the 1980s was a sentencing expert system called "Sentencing Advisor."[36] In describing the advantages of expert systems for criminal justice and sentencing, Gruner noted that:

Expertise contained in an expert legal system can be easily transferred, often through means as simple as copying a computer program or database. Further, where analyses are dependent upon numerical calculations or repetitious reasoning, the tireless operation of an expert legal system may produce significantly better results than human experts in a shorter amount of time. Once freed from these tedious tasks, human workers can perform more interesting and detailed analyses in more difficult areas. Finally, expert legal systems can produce especially well-documented results, since their printing capacities are not limited by human impatience with paperwork. [37]

More recent techniques share many of these useful features, they are totally rational and deterministic, they never have "off days," and they do not tire or express a desire to go to the beach. Machine learning techniques however, unlike symbolic systems, involve algorithms and statistical models that can make decisions or perform functions without explicit instructions, relying instead on patterns and inference derived from large scale data analysis. As a result, they are harder to understand as purely deterministic, and they

35. In theory, deep learning systems are powerful enough to represent any finite deterministic classification between any set of inputs and corresponding outputs. However, there are a range of real-world issues that place practical limitations on deep learning techniques including: finite and indeterminate datasets, datasets that present local minima that defeat gradient descent-based algorithms, outcomes that require extrapolation from data not interpolation within the data, and knowledge that is hierarchically structured. For a serious analysis of these and other issues, see MARCUS, *supra* note 33.

36. Richard S. Gruner, *Sentencing Advisor: An Expert Computer System for Federal Sentencing Analyses*, 5 SANTA CLARA COMPUTER & HIGH TECH. L. J. 51 (1989).

37. *Id.* at 53.

lack explanatory coherence. For example, confronted with inhumanly brilliant behavior from recent game playing AIs, like AlphaGo's winning move thirty-seven in game two against Go master Lee Sedol, or AlphaZero's play in game ten against the symbolicly-based algorithm Stockfish, we are left wondering about the new type of intelligence displayed and ask "how on earth did it come up with that move?"[38]

This leads to an initial problem that we must confront: will humans be able to accept decisions by AIs? As we will examine in the Parts that follow, there is the clear ability for artificial intelligence to inform decisions in the criminal justice system. However, there are several obstacles that need to be overcome before the current forms of AI can have a useful and defining role in the criminal justice domain. One of the key difficulties is the innate human preference for decisions to be made by people instead of computers. People are accepting and tolerant of errors and mistakes made by humans and extremely intolerant of those made by computers. This phenomenon is termed "algorithmic aversion."

Research shows that evidence-based algorithms more accurately predict the future than do human forecasters. Yet when forecasters are deciding whether to use a human forecaster or a statistical algorithm, they often choose the human forecaster. . . . [A]lgorithm aversion, is costly, and it is important to understand its causes. . . . [P]eople are especially averse to algorithmic forecasters after seeing them perform, even when they see them outperform a human forecaster. This is because people more quickly lose confidence in algorithmic than human forecasters after seeing them make the same mistake. In 5 studies, participants either saw an algorithm make forecasts, a human make forecasts, both, or neither. They then decided whether to tie their incentives to the future predictions of the algorithm or the human. Participants who saw the algorithm perform were less confident in it, and less likely to choose it over an inferior human forecaster. This was true even among those who saw the algorithm outperform the human.[39]

Algorithmic aversion is irrational, but it is real. Thus, any proposal that suggests that AI should be incorporated into a criminal justice system at the outset needs to be aware of this phenomenon and

---

38. *See* Steven Strogatz, *One Giant Step for a Chess-Playing Machine*, N.Y. TIMES (Dec. 26, 2018), https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html [https://perma.cc/7BL2-K8D3]; Cade Metz, *How Google's AI Viewed the Move No Human Could Understand*, WIRED (Mar. 14, 2016), https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand [https://perma.cc/KL7D-AVBA]. For a formal discussion of the decision-making processes of AlphaGo, see generally Xiangrui Chao et. al., *Jie Ke versus AlphaGo: A Ranking Approach Using Decision Making Method for Large-Scale Data With Incomplete Information*, 265 EUR. J. OPERATIONAL RES. 239 (2018).

39. Berkeley J. Dietvorst et. al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCHOL.: GEN. 114, 114 (2015).

propose how to circumvent the bias against AI decision-making. This aversion is likely to be especially strong in the criminal law given the important interests at stake. In proposing how to deal with algorithmic aversion, a study by Dietvorst, Simmons, and Massey is particularly illuminating. They note that if the user is given a degree of input into the outcome of algorithm that the user will be far more likely to utilize the algorithm. They note:

> Although evidence-based algorithms consistently outperform human forecasters, people often fail to use them after learning that they are imperfect, a phenomenon known as algorithm aversion. In this paper, we present three studies investigating how to reduce algorithm aversion. In incentivized forecasting tasks, participants chose between using their own forecasts or those of an algorithm that was built by experts. Participants were considerably more likely to choose to use an imperfect algorithm when they could modify its forecasts, and they performed better as a result. This research suggests that one can reduce algorithm aversion by giving people some control—even a slight amount—over an imperfect algorithm's forecast.[40]

In light of these studies and the innate reluctance of humans to subjugate their decision-making to machines, the reform proposals in this Article will generally be advanced in a recommendatory, as opposed to prescriptive, manner. In general, judges and law enforcement officers should have the results of the algorithm available to them, but should not be required or expected to implement the conclusions uncritically. Our reforms are also premised on the basis that the workings of the algorithms and the data upon which they are based will generally be transparent and publicly available. The main exception to this relates to algorithms which predict future criminal events and those which detect criminal acts in the process of being committed. It is not feasible to disclose these algorithms given that it would provide criminals with the knowledge necessary to undermine the utility of these systems. In such cases, we suggest other methods to validate the integrity and fairness of these processes.

In addition to this, it is important to educate the legal profession and the wider community that the responses and answers provided by AI programs are not random, unpredictable, or uncontrollable. Rather, they simply consist of the processing of algorithms and data that can be validated by people. As noted above, AI consists of the extremely rapid processing of often large amounts of information in accordance with a predetermined formula to provide a response many times quicker than a person could provide. A person performing the same

---

40. Berkeley J. Dietvorst et. al., *Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them*, 64 MGMT. SCIENCE 1155, 1156 (2016).

task in accordance with the same formula would always (assuming he or she is free from error) reach the same response, but generally it would take much longer. AI is no different in principle to other automated systems—pocket calculators, cash registers, Excel spreadsheets—that have been used by the community for decades. When people need quick mathematical answers, nowadays they typically simply type the numbers into their calculator as opposed to engaging in multiplication and long division and the like. In doing so, they realize that the answer given by the calculator is simply a preprogrammed response that was in-built by a human programmer. AI involves the same underlying processes, except that the variables and the data are greater in number. But the integrity of the process is no different. And receptivity of AI by the community in the criminal justice sphere should be no less than people have towards using calculators or their mobile phone.

The remainder of the Article maps the path for how AI can enhance community flourishing by massively reducing crime and the suffering that criminal acts inflict on victims while minimizing the fiscal burden that the criminal justice system has on the community. We now examine the current use of AI in the criminal justice system and make reform proposals regarding how it can be utilized to make the system more normatively sound, efficient, and effective. We focus first on the criminal detection stage.

III. DETECTION OF CRIME

A. Use of AI in Deterring and Detecting Crime

1. Predictive Policing

The detection and regulation of criminal activity has traditionally been reactive, in that it generally occurs in reaction to real-time events rather than as a proactive analysis of historical and evolving evidence and data.[41]

Typically, a crime occurs or is in the process of unfolding, and police respond to the event after being notified by a member of the public or the victim. To the extent that policing involves the proactive measures to stop crime, this sometimes involves randomized behavior, for example, routine police patrols. But most policing is directional and strategic. Police currently rely on information from a variety of sources in order to direct their activities and resources. Police departments gather and then collate crime data and use this to identify "crime hotspots." This involves using past events of criminal

41. *See* Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977, 981 (2017).

activity in an attempt to formulate patterns which can anticipate locations of future criminal acts. Pursuant to this process, police monitor and attend locations and events where there is a perceived meaningful risk of criminal behavior, such as large gatherings of people (for example, demonstrations, sporting events, and social events such as music concerts) and known criminal hotspots, such as locations where gang activity has frequently occurred in the past.[42]

This directional behavior not only relates to geographical locations but to targeting specific people. Police collect and collate data regarding individual offenders or groups of offenders in a bid to reduce the incidence of crime. This typically involves the utilization of crude and intuitive judgments. The intuitive approach taken by some police as a basis for conducting stop and frisk procedures was challenged in the class action decision of *Floyd v. City of New York,* where the court held that the searches were undertaken without reasonable suspicion and hence violated the Fourth Amendment.[43] Police sometimes based their judgment about who to target by reference to what are known as "furtive movements" which are set out in the case in following terms:

> Two officers testified to their understanding of the term "furtive movements." One explained that "furtive movement is a very broad concept," and could include a person "changing direction," "walking in a certain way," "[a]cting a little suspicious," "making a movement that is not regular," being "very fidgety," "going in and out of his pocket," "going in and out of a location," "looking back and forth constantly," "looking over their shoulder," "adjusting their hip or their belt," "moving in and out of a car too quickly," "[t]urning a part of their body away from you," "[g]rabbing at a certain pocket or something at their waist," "getting a little nervous, maybe shaking," and "*stutter[ing].*" Another officer explained that "usually" a furtive movement is someone "hanging out in front of [a] building, sitting on the benches or something like that" and then making a "quick movement," such as "bending down and quickly standing back up," "going inside the lobby . . . and then quickly coming back out," or "all of a sudden becom[ing] very nervous, very aware." If officers believe that the behavior described above constitutes furtive movement that justifies a stop, then it is no surprise that stops so rarely produce evidence of criminal activity.[44]

Considerations of this nature are inherently vague, impressionistic, not grounded in research, and hence, not surprisingly the court found

---

42. *See e.g.,* Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 Emory L.J. (2012); Ferguson, *supra* note 9.

43. 959 F. Supp. 2d 553, 667 (S.D.N.Y. 2013).

44. *Id.* at 561 (alteration in the original) (footnotes omitted).

that many police searches were not underpinned by reasonable suspicion.[45]

AI has the capacity to greatly increase the effectiveness of proactive policing. Algorithms have been designed which can predict the likelihood of crime in a certain geographical location and time with a high degree of accuracy. These are based on previous patterns of behavior. A straightforward illustration of this is the use of speed cameras to detect speeding motorists. These cameras are generally located where there has been a previously high incidence of speeding or increased risk of collision. Previous history of driver behavior is a very accurate guide to future behavior.[46]

Predictive policing algorithms are now used in a number of jurisdictions, including Los Angeles.[47] The system utilized in Los Angeles is called PredPol. The algorithm used to predict crime incorporates aspects which have been developed to describe 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.[48]

The integers which drive predictive policing algorithms are confidential. They have been criticized for their secrecy and, in particular, on the basis that they may target minority groups. It has been claimed that predictive policing instruments "could increase police presence in poor and minority communities by creating a 'ratchet effect.'"[49] Currently, there is litigation in place which aims to compel police departments in New York, Chicago, and Los Angeles

---

45. *Id.* at 559.

46. *See, e.g.*, Sonja E. Forward, *The Theory of Planned Behaviour: The Role of Descriptive Norms and Past Behaviour in the Prediction of Drivers' Intentions to Violate*, 12 TRANSP. RES. PART F: PSYCHOL. AND BEHAV. 198, 199–200 (2009) (discussing past behaviour and effects of habit).

47. Ind. Univ., *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 [https://perma.cc/4CB8-ZPL5].

48. 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 [https://perma.cc/RCC4-WE5N].

49. *Id.*

to disclose their algorithms.[50] The secrecy relating to the algorithms has been defended on the basis that "[p]olice officials . . . 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."[51]

While the use of the algorithms remains controversial, the limited data that is available suggests that systems like PredPol are statistically more likely to predict when and where crime will occur than human crime analysts.[52] Further, while some studies have shown that algorithms can target minority groups when applied in certain contexts,[53] a recent study of PredPol has shown "no statistically significant difference between arrest rates by ethnic group."[54]

In addition to using AI to determine where crime is likely to occur next, more nuanced algorithms are used by some police departments to assist police to determine whether particular individuals are likely to commit a crime or to have committed a crime. In Chicago, people who are arrested or observed by police receive a threat score from 1 to 500-plus calculated by an algorithm which is designed to measure the risk that the individual will get shot or shoot another person.[55] The score influences who police target for proactive intervention and the manner in which they deal with suspects and people who are arrested.[56] The code utilized by the algorithm is confidential, but some of the integers that are used include individualized factors such as the individual's past history of offending and their age.[57] More generic factors are also utilized, such as whether criminal activity is generally increasing or decreasing.[58] A number of police departments in other cities in the United States are also using similar algorithms.[59]

The algorithms have been supported on the basis that they have accurately predicted a high rate of shooting victims. However, critics argue that high threat scores inappropriately distort police decisions

---

50. *See* 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 [https://perma.cc/M2QR-YBPF].

51. *Id.*

52. *See* Jouvenal, *supra* note 48; *see generally* G.O. Mohler et al., *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS'N 1399 (2014); *supra* note 47.

53. *See supra* note 41.

54. *Id.*

55. *See* 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/ [https://perma.cc/G8GN-9KW7].

56. *Id.*

57. *Id.*

58. *Id.*

59. *See generally* ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT (2017).

relating to the use of force and results in disproportionate police monitoring of minorities—a risk which is exacerbated by the fact that the algorithm is confidential.[60]

The policy rationale informing the trend towards proactive policing is that traditional reactive policing strategies of detection and investigation do not work.[61] Given the catastrophic costs of crime to federal, state, and local governments,[62] let alone the social and community costs, proactive policing driven by predictive algorithm methods, which does lead to reductions in offending rates and recidivism, is clearly something that is here to stay and is set to become even more ubiquitous. The well-documented problems of entrenched bias and potential for unethical and inequitable application and enforcement means that the natural tendency of coders and technocrats to place too much value on the objective accuracy of computational models, however, will likely attract as much focus in the evolution of predictive policing algorithms. To equate a particular location or neighborhood with criminality, and then profile it with a "black box" algorithm with an in-built racial bias, in an environment in which surveillance technology has become ever present, ought not to be the goal of predictive policing in a civil society. But a properly designed system, where equal emphasis is placed on a regulatory and ethical framework at the development and deployment stages, is surely not beyond us.

The first major data-driven policing algorithm, which has now become the most widely used, is Compstat.[63] The system evolved in response to public concerns in the early 1990's of spiking crime rates in New York City and the apparent inability of the New York Police Department to address these concerns. At that time, the Department collected crime data almost solely to meet its obligation to report statistics to the FBI. Anything like real time trends in crime rates, types, or locations were basically anecdotal.[64] Compstat became highly regarded among agencies which used it due to its effectiveness in allowing them to better concentrate resources on where crime was occurring and the purported causes of crime. It was also an effective tool for information sharing between agencies and

60. *See* Ferguson, *supra* note 55.

61. *See* Lawrence W. Sherman, *The Rise of Evidence-Based Policing: Targeting, Testing, and Tracking*, 42 CRIME AND JUSTICE 377 (2013).

62. *See generally* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-17-732, COST OF CRIME: EXPERTS REPORT CHALLENGES ESTIMATING COSTS AND SUGGEST IMPROVEMENTS TO BETTER INFORM POLICY DECISIONS (2017).

63. Compstat—Computerized Statistics Managerial System. The system is used under different names by various agencies.

64. David Weisburd et al., *Changing Everything so that Everything Can Remain the Same: CompStat and American Policing*, *in* POLICE INNOVATION: CONTRASTING PERSPECTIVES 284-301 (David Weisburg & Anthony A. Braga eds., 2006).

for data matching.[65] But although Compstat is proactive in terms of matching resources to needs, its algorithms do this by identifying existing trends, rather than by engaging in any robust predictive process.

With the exponential rise in the collection and retention of information in the form of both public and private sector data, police have access to shared datasets which contain granular information about people who have never been offenders or otherwise come into contact with the criminal justice system. Along with advances in coding techniques and big data analytics, this has made it possible for policing algorithms to become truly predictive. Systems such as PredPol[66] make predictions of future offending based on a near-repeat model, which analyzes data according to three criteria: offence type, date and time of offence, and location of offence. This enables resources to be utilized pursuant to a "risk-based deployment" under which a local police jurisdiction is mapped on a grid of boxes, and each box is given a risk classification.

More sophisticated predictive policing systems are beginning to make use of machine learning to learn how a much wider range of factors correlates with crime. These more advanced systems then use that data to predict where and when crime will occur in the future. The algorithm 'learns' and improves its accuracy by correlating the results of crime predictions or forecasts against the factors used to make the prediction. 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."[67] Although the extent to which these nudges work in practice to limit over-policing can only be established by external evaluation.

2. Automated Visual Monitoring

Machine learning approaches can also assist crime reduction and detection in ways which supplement the use of existing criminal justice technological innovations. Increasingly, police are relying on technology in order to assist with proactive policing. This is best

65. In a survey of its members, the Police Executive Research Forum asked "'Why 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." BUREAU OF JUSTICE ASSISTANCE, COMPSTAT: ITS ORIGINS, EVOLUTION, AND FUTURE IN LAW ENFORCEMENT AGENCIES, 8 (2013).

66. *See* Jouvenal, *supra* note 48.

67. Aaron Shapiro, *Reform Predictive Policing*, 541 NATURE 458, 459 (2017); Hunchlab was acquired by Shotspotter in 2018, Press Release, Robert Cheetham, Why We Sold HunchLab (Jan. 23, 2019), https://www.shotspotter.com/press-releases/shotspotter-announces-acquisition-of-hunchlab-to-springboard-into-ai-driven-analysis-and-predictive-policing/ [https://perma.cc/W9PG-CE2A].

illustrated by the now high use of CCTV cameras which are located in millions of locations throughout the United States.[68] These have a two-fold role in the criminal justice sphere. First, they discourage the commission of crime in circumstances in which offenders are aware of the location of the cameras. The empirical data establishes that the best way to reduce the incidence of crime is to increase the perception in people's minds that if they offend they will be detected and apprehended,[69] and hence it is not surprising that cameras have been shown to reduce the incidence of crime.[70] The second role of cameras is to gather evidence which can be used by police and prosecutors for detecting crime, identifying offenders, and establishing their guilt in court.

A major problem associated with the use of CCTV cameras is that their effectiveness in stopping crime and apprehending criminals is limited by the fact that it is extremely labor intensive to visually monitor CCTV in live-time. This process can be made far more cost-effective by computer-based monitoring of the CCTV footage. Recent advances in machine learning visual processing has allowed for large scale automated monitoring of locations, and the flagging of problematic behavior within that space.

This works in a straightforward manner. Imagine a static camera trained on a closed door, and an image processor that checks the image once per second. The images are, of course, large datasets, and over a large number of iterations, the algorithm develops a statistical picture of the world that codes the way that the location looks when the door is closed. Any opening of the door will register as a perturbation of the model, and can be flagged for security guards to investigate. And this is not just limited to static scenes, as the same basic approach can be applied to complex patterns of behavior. We can train the algorithm on a location that has many people moving through it during the day, but no one at night. [T]he presence of a person moving through the

68. *See* Liza Lin & Newley Purnell, *A World With a Billion Cameras Watching You Is Just Around the Corner,* WALL ST. J. (Dec. 6, 2019), https://www.wsj.com/articles/a-billion-surveillance-cameras-forecast-to-be-watching-within-two-years-11575565402 [https://perma.cc/92D9-HFZB]. Global numbers to grow almost 30% as higher image quality allows better facial recognition. The authors state, "[t]he U.S. rivals China in terms of security-camera penetration, with one camera for every 4.6 people, not far from China's one camera for 4.1 people." *Id.*

69. *See* Mirko Bagaric & Theo Alexander, *(Marginal) General Deterrence Doesn't Work – and What it Means for Sentencing*, 35 CRIM L. J. 269, 280-82 (2011) [hereinafter Bagaric & Alexander, *(Marginal) General Deterrence Doesn't Work*]; Mirko Bagaric & Theo Alexander, *The Capacity of Criminal Sanctions to Shape the Behaviour of Offenders: Specific Deterrence Doesn't Work, Rehabilitation Might and the Implications for Sentencing*, 35 CRIM. L.J. 159, 163-64 (2012) [hereinafter Bagaric & Alexander, *The Capacity of Criminal Sanctions to Shape the Behaviour of Offenders*].

70. AUSTL. INST. OF CRIMINOLOGY, EFFECTIVENESS OF PUBLIC SPACE CCTV SYSTEMS (2017).

location at 2am will be flagged as suspicious, triggering an alarm alerting a law enforcement officer so that she can make an immediate judgment call regarding the appropriate response.

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.[71] The system used at this location is iCetana, one of the leading manufacturers of real-time AI-assisted video monitoring [72] The technology has been in use at the Swinburne campus for over seven years. The tool learns by monitoring the relevant area for a period of time and then flags unusual behavior. The system is constantly recalibrating movement patterns in order to classify the type of behavior which is normal. Thus, irregular behavior is used as a proxy for activity that is potentially criminal activity. It detects behavior such as running, loitering, falling, and punching. It even can recognize pre-aggression stances that occur due to differences in posture that coincide with hostility. The system is not sufficiently nuanced to pick up all forms of criminal conduct, such as drug selling. However, in addition to self-learning automated CCTV algorithms, there are also rule-based systems, where the 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, such as drug offending.

The other area where AI-based visual processing is used extensively is in facial recognition. Machine learning techniques have advanced quickly in this area, and now are remarkably reliable in ideal conditions.[73] Facial recognition technologies can be used by law enforcement to identify offenders in public settings, for example those with outstanding warrants or those wanted for questioning. 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 [74] its privacy-

---

71. *See University Improving Situational Awareness*, ICETANA (Jan. 28, 2019), https://icetana.com/university-enhancing-situational-awareness/ [https://perma.cc/SV4L-EWTL].

72. *About*, ICETANA, https://icetana.com/company/#about [https://perma.cc/333C-274W].

73. *See* Kate Kaye, *This Little-Known Facial-Recognition Accuracy Test Has Big Influence*, INT'L ASS'N PRIVACY PROFS. (Jan. 7, 2019), https://iapp.org/news/a/this-little-known-facial-recognition-accuracy-test-has-big-influence/ [https://perma.cc/TT35-3NBC] (reporting on NIST tests, reporting facial recognition accuracy rates as high as 99.8%).

74. *See* Jieshu Wang, What's in Your Face? Discrimination in Facial Recognition Technology (Apr. 13, 2018) (unpublished M.A. thesis, Georgetown University), available at https://repository.library.georgetown.edu/bitstream/handle/10822/1050752/Wang_georgetown_0076M_14043.pdf?sequence=1&isAllowed=y [https://perma.cc/8YDK-RG6L].

invading nature[75] some limitations in datasets that have led to misidentification of people in certain groups,[76] and the way that the technology can be used by authoritarian governments to control dissidents or ethnic groups.[77] Although these concerns are appropriate, they are not particularly problematic where the facial recognition technology is used by the police, whose use is proscribed by regulation and constitutional protections including the Fourth Amendment's proscription on unreasonable search and seizure.

Thus, there are manifest benefits that can emerge from AI in terms of discouraging crime and apprehending criminals. But as alluded to above, there are numerous problems with the technology that need to be overcome before its full potential can be reached. We now address these challenges.

B. Criticisms of AI in Policing

The use of AI to assist in policing has been criticized on several grounds. One is that it involves racial bias and hence discriminates against already socially and economically disadvantaged groups. This is considered at length in the next part of the Article given that it is a criticism that relates to the use of algorithms at all stages of the criminal justice system, including sentencing. Other criticisms relate to the rectitude of the systems and the claim that predictive policing and enhanced AI monitoring results in the violation of numerous rights, including privacy, and those that are normally incidental to arrest, including the right to liberty. We now consider these criticisms.

1. Establishing the Validity and Improving the Efficiency of Predictive Policing

The principal benefit of predictive policing is that it improves the ability of police to stop crime and apprehend criminals by deploying police resources to locations where crime is most likely to be committed. Any system that reduces the harmful effects associated with crime is clearly desirable. However, in order to consolidate the use of predictive policing and potentially increase reliance on it, it is

75. *See, e.g.*, 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 [https://perma.cc/4TG6-7J7J]; Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion,* 163 U. PA. L. REV. 327, 329-31 (2015)*;* I. Bennett Capers, *Crime, Surveillance, and Communities*, 40 FORDHAM URB. L.J. 959, 963–64 (2013); Wayne A. Logan, *Policing Identity*, 92 B.U. L. REV. 1561, 1603 (2012).

76. *See* 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 [https://perma.cc/FZV7-A6CD].

77. *See* 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 [https://perma.cc/7ZYT-26EG].

necessary to first establish the validity of the system and, if possible, to improve the accuracy of the system. In addition to this, it is important that a cost-benefit assessment of the system is undertaken to demonstrate that the financial resources devoted to predictive policing do not exceed the additional cost of the extra police that it would take to achieve similar reductions in crime.

As alluded to above, there is evidence that predictive policing is more effective at reducing crime than traditional policing approaches. However, the number of studies that have been undertaken is not significant, and the results of these studies are not definitive. Thus, there is a need to better evaluate current predictive policing systems. This process will presumably facilitate a comparison of different predictive policing methods and thereby result in improved systems. In addition to this, a detailed cost benefit analysis needs to be undertaken of the capital cost involved in developing predictive policing algorithms and the costs associated with maintaining, updating, and implementing them on a day-to-day basis. Most well-designed automated processes are cheaper to run than systems using human labor, thus it is likely that predictive policing is cost effective, but this needs to be established, not assumed.

The same evaluative processes need to be undertaken in relation to the automated monitoring of CCTV or facial recognition. Thus, the accuracy of current algorithms used for depicting criminal acts needs to be assessed and further research should be undertaken to improve their reliability and accuracy. A cost-benefit analysis needs to be undertaken in relation to their roll-out and usage.

2. Infringement of Privacy and Breach of Rights Relating to Search, Seizure, and Arbitrary Arrest

A likely criticism of the increased use of AI in relation to policing—especially relating to facial recognition and the automated monitoring of CCTV cameras—is that it will violate the right to privacy. This is not an overwhelming obstacle. The first reason for this is that the right to privacy itself is a contentious interest. The definition and justification of the right is unclear. Robert Post has lamented "[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."[78] Perhaps the most enlightening definition of

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

privacy is simply "the right to be let alone."[79] The rationale for privacy is generally thought to stem from the broader virtues of autonomy and dignity.[80]

Despite doctrinal uncertainty regarding the nature and source of the right to privacy, the Supreme Court has acknowledged it as a legally protected interest. 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.[81] In *Roe v. Wade*, for example, Justice Blackmun stated in his majority opinion:

> The Constitution does not explicitly mention any right of privacy. In a line of decisions, however, . . . the Court has recognized that a right of personal privacy, or a guarantee of certain areas or zones of privacy, does exist under the Constitution. In varying contexts, the Court or individual Justices have, indeed, found at least the roots of that right in the First Amendment; in the Fourth and Fifth Amendments; in the penumbras of the Bill of Rights; in the Ninth Amendment; or in the concept of liberty guaranteed by the first section of the Fourteenth Amendment.[82]

The right to privacy, however, is virtually negated in the context of some aspects of the criminal justice system, including where criminal sanctions are imposed. In *Hudson v. Palmer*, the Court noted that it would not be possible to achieve many of the security objectives of prisons, which involve prohibiting the introduction of drugs and weapons into prisons, if prisoners retained the right to privacy.[83]

Thus, while the right to privacy does receive some legal recognition, it is a weak right, which is often impinged upon, often 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. CCTV monitoring exists in many parts of America. A person who walks the streets of Manhattan or most large American cities will have their image taken hundreds of times. The increasing monitoring of people that potentially stems from the use of AI to monitor CCTV cameras or facial recognition systems is little different in nature to that which currently occurs. 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

---

79. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193 (1890).

80. VICTORIAN LAW REFORM COMM'N, WORKPLACE PRIVACY: ISSUES PAPER (2002), http://www.lawreform.vic.gov.au/sites/default/files/IssuesPaperfinal.pdf [https://perma.cc/Z6RV-ALES].

81. *See, e.g.*, Lawrence v. Texas, 539 U.S. 558, 571–72 (2003); Roe v. Wade, 410 U.S. 113, 129 (1973); Griswold v. Connecticut, 381 U.S. 479, 485 (1965).

82. *Roe*, 410 U.S. at 152 (citations omitted).

83. 468 U.S. 517, 517-18 (1984); *see also* Williams v. Kyler, 680 F. Supp. 172, 173 (M.D. Pa. 1986).

is being monitored in live time and a crime is occurring, the operator will typically do all that is reasonably possible to prevent the crime, including notifying police or, where the technology is available—for example, where a speaker system is attached to the CCTV device—notifying the offender that the event is being viewed and recorded with the purpose of discouraging the offender from continuing with the conduct. When an offence is recorded by CCTV, this will be used to assist in the detection and prosecution of the offender. It is clear that observation of this nature does not breach acceptable privacy limits. There are countless instances of crimes that have been solved by police viewing CCTV footage of the event, generally when the offender was unaware that the location was being filmed, and the offender being identified after his or her image was screened in the mainstream media.[84] The important point being that 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 to existing limitations of this right. To the extent that the incursions are more frequent and targeted, this could be readily justified by the common good that is achieved by reducing crime and the increased rate of detecting and prosecuting offenders.

Moreover, to some extent, both facial recognition and automated CCTV observance is less intrusive than live-time viewing by a human being. In the automated context, law enforcement officers will only observe the CCTV or view the facial image when a computer detects that the footage suggests that a crime is being committed or that an offender has been recognized. Thus, for most of the time, individuals will be *potentially* observable, instead of being constantly observed or monitored by law enforcement.

The rights to liberty, property, and bodily integrity are, however, more powerful and have far stronger legal protections than the right to privacy. AI directed policing will result in certain cohorts of people being more frequently arrested, searched, and stripped of their property (as a result of searches following arrests) than is currently the situation. This has already resulted in claims of unfairness, discrimination, and persecution levelled at this form of policing and suggestions that it potentially violates the Equal Protection clause and Fourth Amendment.[85] These are potentially strong objections, but again not decisive if the algorithm is developed appropriately.

---

84. *See generally* Kate Dailey, *The Rise of CCTV Surveillance in the US*, BBC NEWS MAGAZINE (Apr. 29, 2013), https://www.bbc.com/news/magazine-22274770 [https://perma.cc/9BQ2-ZDCW] (discussing the identification of the perpetrators of the Boston marathon bombings via CCTV). A specific example from the UK is that of the London nail bomber, David Copeland, who was identified by an acquaintance from CCTV footage published in mainstream media. *See* V. Bruce et al., *Matching Identities of Familiar and Unfamiliar Faces Caught on CCTV Images*, 7 J. EXPERIMENTAL PSYCHOL. APPLIED 217 (2001).

85. Simmons, *supra* note 2, at 972.

Before examining the design of any such algorithm—which we undertake in Part IV below—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 socioeconomic and racial minority groups.[86] This is a criticism that has been forcefully levelled at many police departments and one of the reasons that has been suggested for explaining the grossly disproportionate rate of arrest and imprisonment of Hispanic and African-American offenders.[87] Thus, if crime prevention and detection algorithms do result in police more frequently policing lower socioeconomic 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 which informs the algorithm is consciously and deliberately prescribed, and hence there is the opportunity to evaluate the algorithms for group profiling and ensure that this is not a design feature. This of course assumes that the workings of the algorithm are made transparent or can be independently tested to demonstrate that they are not biased in their selection suspects. This is a matter addressed further below, but in short, our view is that for AI to gain acceptance and legitimacy in the criminal justice sector it is necessary to establish that it does not result in the discriminatory targeting of certain groups in the community.

IV. BAIL, SENTENCING, AND PAROLE

A. The Key Unifying Integer in Bail, Sentencing, and Parole: The Likelihood that the Defendant Committed a Serious Offense

We now discuss the use of AI at the post-arrest stage of the criminal justice system.

---

86. *Id.* at 974.

87. *See* Angela J. Davis, *Prosecution and Race: The Power and Privilege of Discretion*, 67 FORDHAM L. REV. 13, 29-30 (1998); 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, 296–99 (2014); Tracey L. McCain, *The Interplay of Editorial and Prosecutorial Discretion in the Perpetuation of Racism in the Criminal Justice System*, 25 COLUM. J.L. & SOC. PROBS. 601, 602 n.5 (1992); Kim Farbota, *Black Crime Rates: What Happens When Numbers Aren't Neutral*, HUFFINGTON POST (Sep. 2, 2016), http://www.huffingtonpost.com/kim-farbota/black-crime-rates-your-st_b_8078586.html [https://perma.cc/NT5U-PPMM]; Task Force on Race and the Criminal Justice Sys., *Preliminary Report on Race and Washington's Criminal Justice System*, 35 SEATTLE U. L. REV. 623, 636, 642–44 (2012); Kochel et al., *Effect of Suspect Race on Officers' Arrest Decisions*, 49 CRIMINOLOGY 473, 490 (2011); Paul Butler, *Starr Is to Clinton as Regular Prosecutors Are to Blacks*, 40 B.C. L. REV. 705, 708–09 (1999) (citing to MARC MAUER & TRACY HULING, THE SENT'G PROJECT, YOUNG BLACK AMERICANS AND THE CRIMINAL JUSTICE SYSTEM: FIVE YEARS LATER 9–10 (1995)); *Decades of Disparity: Drug Arrests and Race in the United States*, HUMAN RIGHTS WATCH (Mar. 2, 2009), https://www.hrw.org/report/2009/03/02/decades-disparity/drug-arrests-and-race-united-states [https://perma.cc/6AHW-DQRN].

Bail, sentencing, and parole occur at different stages of the post-arrest phase of the criminal justice process. Bail decisions are made following the charging of a suspect and prior to the determination of guilt or innocence. At this stage of the process, the suspect has not been convicted of an offence, and a decision is made whether the offender should be released into the community until the suspect's criminal liability is determined. Sentencing occurs only once the offender has been found guilty (either following a trial or pleading guilty). Parole is the back-end of the criminal justice system. Most offenders who are sentenced to prison are eligible for release into the community prior to the expiration of their prison term.[88] If they are successful in securing this release, they are placed on parole. While these phases of the criminal justice system have different objectives and criteria that inform decision-making, they share one very important commonality. The key consideration that informs decision-making in all of these stages is community protection.

Thus, in relation to bail the main determinant is the risk that the suspect will reoffend if he or she is released into the community. The same applies in relation to parole. Sentencing has a number of objectives, including deterrence and rehabilitation, but the aim that has been paramount in the United States for the past few decades is community protection[89] Accordingly, the key consideration that informs the in/out (of prison) sentencing decisions and the length of a prison term that might be imposed is an assessment of the likelihood that the defendant will commit a serious offense.

Three different techniques have been used to determine a defendant's level of risk of offending.[90] The first involves unstructured clinical assessments, where an individual assessor determines the offender's risk of reoffending according to impressionistic criteria without empirical validation.[91] This approach has been shown to be the least reliable and, because of the subjectivity associated with this approach, there is no way that it can be built into a system based

88. Jorge Renaud, *Grading the Parole Release Systems of All 50 States*, PRISON POLICY INITIATIVE (Feb. 26, 2019), https://www.prisonpolicy.org/reports/grading_parole.html [https://perma.cc/6P85-SK72].

89. *See, e.g.*, NATIONAL RESEARCH COUNCIL, THE GROWTH OF INCARCERATION IN THE UNITED STATES: EXPLORING THE CAUSES AND CONSEQUENCES 6 (Jeremey Tavis et al. eds., 2014).

90. As discussed further in this section, the main three methodologies are unstructured clinical assessments, actuarial methodologies, and structured professional judgment assessments. *See* Michael R. Davis & James R. P. Ogloff, *Key Considerations and Problems in Assessing Risk for Violence*, *in* PSYCHOLOGY AND LAW: BRIDGING THE GAP 191, 195–96 (David Canter & Rita Žukauskienė eds., 2008); Christopher Slobogin, *Risk Assessment*, *in* THE OXFORD HANDBOOK OF SENTENCING AND CORRECTIONS 196, 198 (Joan Petersilia & Kevin R. Reitz eds., 2012).

91. Slobogin, *supra* note 900, at 198; *see also* Jordan M. Hyatt & Steven L. Chanenson, *The Use of Risk Assessment at Sentencing: Implications for Research and Policy* (Vill. U. Sch. of L., Working Paper Series, 2016), http://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps [https://perma.cc/Q7JJ-HV99].

on an algorithm.[92] However, there are more accurate risk assessment methods which can be readily computerized. The second mechanism for predicting offenders' risk of reoffending involve actuarial-based assessments.[93] These approaches are often termed "risk assessment" tools,[94] and they measure an individual's chances of endangering public safety generally by using actuarial methodologies that identify variables that contributed to their occurrence.[95] This information is extrapolated via an algorithm to create rules regarding the likelihood of future events occurring. 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."[96] These sorts of tools are relatively new and so they are sometimes treated with caution. However, both the concept and approach underpinning them 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.[97]

A large number of risk assessment tools have been developed. The main differences between them are the integers that they use and the relative weights that they apply to relevant considerations that have been ascertained as being relevant to the risk of future offending. Generally, we find that an offender's criminal history is a constant, base determinant,[98] and other key variables include an

---

92. *See* Christopher Slobogin, *Principles of risk assessment: Sentencing and policing*, 15 OHIO ST. J. CRIM. L. 583 (2018).

93. *See* Davis & Ogloff, *supra* note 90, at 195. *See also* Paisly Bender, *Exposing the Hidden Penalties of Pleading Guilty: A Revision of the Collateral Consequences Rule*, 19 GEO. MASON L. REV. 291, 313 (2011); Melissa Hamilton, *Back to the Future: The Influence of Criminal History on Risk Assessments*, 20 BERKELEY J. CRIM. L. 75, 76 (2015); Michael Tonry, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 FED. SENT'G REP. 167, 168 (2014). Such tools are in fact now used in the majority of states in the United States. *See* Shawn Bushway & Jeffrey Smith, *Sentencing Using Statistical Treatment Rules: What We Don't Know Can Hurt Us*, 23 J. QUANTITATIVE CRIMINOLOGY 377, 378 (2007).

94. Davis & Ogloff, *supra* note 90, at 195; Pari McGarraugh, Note, *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).

95. McGarraugh, *supra* note 94, at 1091–92. In addition, actuarial methodologies and other risk assessment approaches include unstructured clinical assessments and structured professional judgment assessments. *See* Davis & Ogloff, *supra* note 90, at 195.

96. Hamilton, *supra* note 93, at 92.

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

98. Hamilton, *supra* note 93, at 89.

offender's criminal associates, pro-criminal attitudes, and antisocial personality.[99] 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.[100] It is described as one of the latest (fourth) generation predictive tools,[101] and is more nuanced than many earlier predictive models. It scores not only static factors, such as prior criminal history, but also looks to dynamic variables, such as employment status, employment history, education, and family relationships.[102]

Some courts already use risk assessment tools in reaching sentencing decisions. However, most do so in a non-systematic way that does not have a significant impact on the sentencing calculus.[103] The Brennan Center summarized the use of risk assessment tools in sentencing determinations, highlighting the differences between states:

> Driven by advances in social science, states are increasingly turning toward risk assessment tools to help decide how much time people should spend behind bars. These tools use data to predict whether an individual has a sufficiently low likelihood of committing an additional crime to justify a shorter sentence or an alternative to incarceration. . . . Some courts have implemented risk assessments to determine whether defendants should be held in jail or released while waiting for trial; similarly, some parole boards use them to decide which prisoners to release. States such as Kentucky and Virginia have implemented the former, while Arkansas and Nevada have implemented the latter. More recently, states are applying risk assess-

---

99. *Id.* at 90.

100. Admin. Off. of the U.S. Courts Probation and Pretrial Servs. Offs., *An Overview of the Federal Post Conviction Risk Assessment* (2018), https://www.uscourts.gov/sites/default/files/overview_of_the_post_conviction_risk_assessment_0.pdf [https://perma.cc/5LPE-ZSCB]. 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; Wisconsin State Risk Assessment Instrument, and most of these are used for assessing post-sentencing correctional populations. Hyatt & Chanenson, *supra* note 91, at 4.

101. *Id.* at 3.

102. Hamilton, *supra* note 93, at 94. Another common similar tool is the Level of Service Inventory, which incorporates fifty-four considerations. *See* Slobogin, *supra* note 90, at 199. In terms of predicting future violence, it has been noted that dynamic measures are slightly more accurate than static measures for short- to medium-term predictions of violence. *See* Chi Meng Chu et al., *The Short- to Medium-term Predictive Accuracy of Static and Dynamic Risk Assessment Measures in a Secure Forensic Hospital*, 20 ASSESSMENT 230, 237 (2013). Given that these tools go beyond the use of static factors and incorporate dynamic factors, they are sometimes referred to as structured professional judgment tools.

103. They are most commonly used in Virginia, Missouri, and Oregon. Slobogin, *supra* note 90, at 202–03.

> ments to guide sentencing decisions. The first state to incorporate such an instrument in sentencing was Virginia in 1994. By 2004, the state implemented risk assessments statewide, requesting judges to consider the results in individual sentencing decisions. Courts in at least 20 states have begun to experiment with using risk assessments in some way during sentencing decisions. . . . Because these instruments do not change existing sentencing laws, which the authors believe are a root cause of overly long sentences, this report does not delve further into the use of risk assessment in sentencing.[104]

The third mechanism that has been developed to predict offenders' recidivism involves "risk and needs assessments." This type of approach assesses the risk of offenders reoffending and identifies needs of those offenders that, if met, would lower their probability of recidivism.[105] These sorts of instruments are often referred to interchangeably with risk assessment tools; however, there are a range of significant functional differences between them. Risk assessments measure a defendant's chances of reoffending and thereby endangering the public.[106] On the other hand, risk and needs assessments seek to reduce offenders' risk of recidivism by determining which programs and other interventions would stop them re-offending.[107] Risk and needs assessment tools rely on a technique called "structured professional judgment."[108] It differs from a strictly actuarial approach, because the main aim of this type of instrument is to generate the information required to create a needs assessment and a risk management plan, whereas the actuarial approach predicts antisocial behavior.[109] The score that results from a risk and needs assessment is not designed to predict the offender's risk of reoffending, and considerations other than those in the instrument can be taken into account to reduce the individual's risk of recidivism.

Research suggests that, while risk and needs assessment tools are far from perfect, the best instruments, administered by well-trained staff, can predict re-offending with 70% accuracy.[110] Risk and needs

---

104. JAMES AUSTIN & LAUREN-BROOKE EISEN WITH JAMES CULLEN & JONATHAN FRANK, HOW MANY AMERICANS ARE UNNECESSARILY INCARCERATED? 18–19 (2012)(footnotes ommitted). Judges often pay little regard to the results of risk assessment tools. As noted by Slobogin, in Virginia, fifty-nine percent of defendants who were considered to be at low risk of reoffending by a risk assessment tool were still sentenced to a prison. Slobogin, *supra* note 90, at 202; *see also* Simmons, *supra* note 2, at 966.

105. *See* NATHAN JAMES, CONG. RESEARCH SERV., RISK AND NEEDS ASSESSMENT IN THE CRIMINAL JUSTICE SYSTEM 1–2 (2015).

106. McGarraugh, *supra* note 94, at 1091.

107. *Id.*

108. Slobogin, *supra* note 90, at 199.

109. *Id.*

110. 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

assessment tools are far more accurate than unstructured judgments, and, 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.[111]

Given the accuracy of risk and needs assessment tools, it is not surprising that they are used extensively in determining conditions for probation[112] and the appropriateness of parole.[113] However, they are used far less frequently in the sentencing process,[114] and not widely used in relation to bail determinations.[115] Given their efficacy, there is obvious potential for this to change, since a key consideration at bail is whether the defendant is likely to commit an offense if he or she is released into the community.

Shortly, we examine these and other criticisms of risk and needs assessment tools, but before doing so, we more fully outline the key advantages associated with incorporating AI into post-arrest aspects of the criminal justice system. We commence with the bail system and then proceed to sentencing decisions and parole determinations.

B. AI and Bail – Will the Defendant Abscond?

Apart from an offender's likelihood of offending, the other main consideration that informs bail decisions is whether the defendant is a flight risk. At present, this is a matter that is determined by the

---

generally more accurate than predictions based solely on clinical judgment. *See* D.A. Andrews et al., *The Recent Past and Near Future of Risk and/or Need Assessment*, 52 CRIME & DELINQ. 7, 12–13 (2006); William M. Grove et al., *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, 62 (2018); but cf. Christopher Slobogin, *A Defense of Modern Risk-Based Sentencing Risk and Retribution: the Ethics and Consequences of Predictive Sentencing* (forthcoming) (manuscript 4–5) (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3242257 [https://perma.cc/BMY8-9SGJ]).

111. James, *supra* note 105, at 5-8. For earlier research findings regarding the accuracy of such tools, see CARLEEN THOMPSON & ANNA STEWART, REVIEW OF EMPIRICALLY BASED RISK/NEEDS ASSESSMENT TOOLS FOR YOUTH JUSTICE 33-34 (2006); FRANK MORGAN ET AL., RISK ASSESSMENT IN SENTENCING AND CORRECTIONS, CRIMINOLOGY RESEARCH COUNCIL, 99-101 (1996), http://crg.aic.gov.au/reports/22-95-6.pdf [https://perma.cc/RF8B-FABD]; MAX MALLER & RICHARD LANE, A RISK ASSESSMENT MODEL FOR OFFENDER MANAGEMENT, AUSTL. INST. CRIMINOLOGY (2002) http://www.aic.gov.au/media_library/conferences/probation/maller.pdf [https://perma.cc/UM69-3UXW]; Brooke Rae Winters & Hennessey Hayes, *Assessing the Queensland Community Corrections RNI (Risk Needs Inventory)*, 12 CURRENT ISSUES CRIM. JUST. 288, 289 (2001). *See also* Slobogin, *supra* note 90, at 200.

112. Edward Latessa & Brian Lovins, *The Role of Offender Risk Assessment: A Policy Maker Guide*, 5 VICTIMS & OFFENDERS 203, 205 (2010).

113. *Id.*

114. PAMELA M. CASEY ET AL., NAT'L CTR FOR ST. CTS., USING OFFENDER RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: GUIDANCE FOR COURTS FROM A NATIONAL WORKING GROUP (2011).

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

intuitive views of judges. However, it is also an issue about which there is a large amount of data available to identify the characteristics which are most indicative of a risk of absconding. This data could be readily collated and used to develop an algorithm to determine the traits of defendants that are at highest risk of absconding.

Since the decision of the Supreme Court in *United States v. Salerno*,116 a person may be detained in custody pending trial only where there are no conditions under which release could reasonably assure public safety. Forms of release which are permitted include: payment of a full cash bond, grant of an unsecured bond or conditional release, or by bail which is guaranteed by way of surety, that is, by a third party (in some states a commercial bail bondsman).

Cash bail is generally not available in New Jersey or Alaska. Accused criminals are risk assessed according to the Public Safety Assessment (PSA) tool. [117] It weighs nine factors to predict the likelihood of three pretrial outcomes for a given offender, one of which is Failure to Appear (absconding). The factors are:

(1) the defendant's age at the time of arrest;

(2) whether the current charge is a violent offense;

(2a) whether the current charge is a violent offense and the defendant is 20 years old or younger; (3) whether the defendant has a pending charge at the time of the offense;

(4) whether the defendant has a prior disorderly persons conviction;

(5) whether the defendant has a prior indictable conviction;

(5a) whether the defendant has a prior disorderly persons or indictable conviction

(6) whether the defendant has a prior violent conviction;

(7) whether the defendant has a prior failure to appear pretrial in the past two years;

(8) whether the defendant has a prior failure to appear pretrial older than two years; and

(9) whether the defendant has a prior sentence to incarceration.[118]

The factors were found to be the closest correlates to whether a defendant would commit another offence, commit another offence involving violence, or abscond (fail to appear). The data used to determine the correlates was contained in approximately 1.5 million bail decisions from 300 U.S. jurisdictions.

---

116. 481 U.S. 739, 750–51 (1987).

117. LAURA AND JOHN ARNOLD FOUND., PUBLIC SAFETY ASSESSMENT: RISK FACTORS AND FORMULAS 2–3 (2016), https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/PSA-Risk-Factors-and-Formula.pdf [https://perma.cc/U7T2-B66N].

118. *Id.* at 3.

It is important to note that other automated decision-making procedures are also beginning to be combined with the predictive risk systems in ways which may not have been foreseen or intended by the vendors of the algorithms. In New Jersey, for example, the weighted scores from the PSA are then entered into the Decision Making Framework (DMF), which assigns the defendant a "risk level." This level is calculated on the basis of the PSA score and the offence, or offences, with which the defendant has been charged.[119] If the defendant has a history of absconding or has been charged with murder, rape, or robbery, has an elevated risk toward violence, where the defendant was arrested while on pretrial release for two or more pending offenses, then the DMF will issue a no release recommendation regardless of the defendant's PSA risk score.[120] Some offences, such those involving the possession or use of weapons, are flagged as requiring a "heightened response" without generating an automatic no release recommendation.

Since PSA was introduced in New Jersey, two noteworthy outcomes have emerged. There has been a significant drop in crime rates,[121] especially in violent crimes. But there has also been a significant rise in the costs of managing defendants granted pretrial release, specifically the costs of issuing and monitoring GPS tracking devices and the labor costs incurred in investigating suspected breaches of bail conditions.[122]

C. AI and Sentencing

1. Rule of Law Benefits: Consistent, Predictable, and Transparent Sentencing Law and Outcomes

Sentencing is the state's most coercive area, as sentencing involves the deliberate infliction of sanctions on its citizens, including the imposition of financial penalties, the deprivation of liberty, and, in extreme cases, the death penalty. Given what is at stake, it is unjustifiable for courts to make decisions in this realm that are inconsistent, arbitrary, or opaque. Such decisions would fundametally

119. The operation of PSA in combination with DMF is explained by in the New Jersey Court's Annual Report 2017. Glenn A. Grant, *Criminal Justice Reform Report to the Governor and the Legislature for Calendar Year 2017*, N.J. JUDICIARY 1, 11 (2018), https://www.njcourts.gov/courts/assets/criminal/2017cjrannual.pdf [https://perma.cc/FJ96-BW8F].

120. *Id.* at 11–12.

121. *2018 Uniform Crime Report*, ST. N.J. DEP'T LAW & PUBLIC SAFETY 1 (Oct. 19, 2018), https://www.njsp.org/ucr/pdf/current/20181019_crimetrend_2018.pdf [https://perma.cc/GR85-US59] (Homicide -70%, Rape – 35.1%, Robbery – 50.4%, Assault - 44.6%).

122. Grant, *supra* note 119, at 9-10.

violate the rule of law.[123] Geoffrey de Q. Walker explains that 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."[124] The rule of law operates in a society where everyone—including judicial decision-makers—acknowledges an obligation to comply with the law, and where there is "an absence of arbitrary coercion."[125] While, as Walker appreciates, 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." [126] In helping to preclude arbitrary and uncertain justice, consistent, predictable, and transparent sentencing decisions constitute a crucial safeguard of the rule of law.

John Rawls observes that "[t]he rule of law ... implies the precept that similar cases be treated similarly,"[127] and Walker considers that, when implemented in practice, this principle of consistent decision-making significantly limits the discretion of judges and "forces them to justify the distinctions that they make between persons by reference to the relevant legal rules and principles." [128] As Maria Jean J. Hall and others also put it, "it is desirable that like cases be treated alike," [129] and "there is universal acceptance that consistency of approach should be an essential feature of sentencing decision-making."[130]

One of the main reasons for the move from indeterminate to prescriptive sentencing in the United States over the past forty years was the inconsistencies that previously plagued sentencing law and practice.[131] It seems, however, that even largely prescriptive sentencing models have failed to achieve a reasonable level of consistency. A number of recent studies have demonstrated wide-ranging sentencing disparity among judges applying the Federal Sentencing Guidelines.[132] A study of judges at the Boston division of

123. *See e.g.*, JOSEPH RAZ, THE AUTHORITY OF LAW 211, 214-16 (1979); JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 270-76 (1980); Jeffrey Jowell, *The Rule of Law Today*, *in* THE CHANGING CONSTITUTION 3 (Jeffrey Jowell & Dawn Oliver eds., 1985).

124. GEOFFREY DE Q. WALKER, THE RULE OF LAW 1 (1988).

125. *Id.* at 3.

126. *Id.* at 42.

127. JOHN RAWLS, A THEORY OF JUSTICE 237 (1971).

128. WALKER, *supra* note 124, at 19.

129. Maria Jean J. Hall et al., *Supporting Discretionary Decision-Making with Information Technology: A Case Study in the Criminal Sentencing Jurisdiction*, 2 UNIV. OF OTTAWA LAW & TECH. J. 1, 3 (2005).

130. *Id.* at 31.

131. MARVIN E. FRANKEL, CRIMINAL SENTENCES: LAW WITHOUT ORDER 8 (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).

132. *See* Nancy Gertner, *A Short History of American Sentencing: Too Little Law, Too Much Law, or Just Right*, 100 J. CRIM. L. & CRIMINOLOGY 691, 696-97 (2010); *see also* Joshua M. Divine, *Booker Disparity and Data-Driven Sentencing*, 69 Hastings L.J. 771, 790 (2018).

the District of Massachusetts showed that the three most lenient judges imposed sentences that were on average 25.5 months or less, while the other two judges, who sentenced at least fifty defendants, imposed sentences that were more than double this length. [133] Syracuse University's Transactional Records Access Clearinghouse program studied approximately 370,000 federal sentences imposed nationwide and similarly observed wide inter-judge disparity in numerous jurisdictions. For example, the median sentences between judges in Dallas ranged from 60 and 121.5 months, and between judges in the District of Columbia the median sentences ranged from 27 to 77 months.[134] A major reason for these inconsistencies is that implicit biases and deep-rooted values and beliefs of individual judges often affect their decision-making. Even though American judges normally make decisions within prescriptive and guideline sentencing systems that have presumptive penalties, there is considerable opportunity for their personal views of offenders (including those perceptions of which even they are unaware) to affect their decisions.[135]

One of the more obvious potential advantages of computerized sentencing is that it could make sentencing law and sentencing outcomes more consistent, predictable, and transparent (providing, of course, that the formula underpinning the algorithm is disclosed). Hutton has noted that "[o]ne 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" and ensure that sentencing decisions are consistent with one another.[136] Computerized sentencing does have the potential to achieve broad consistency between sentences that are imposed on offenders for similar crimes. Computers cannot make decisions pursuant to sentiments and agendas that are not explicitly incorporated into their programs. As Richard Susskind observes, "computer systems will not suffer from 'off-days' that so often inhibit the performance of human beings." [137] Indeed, lacking human irrationality, there is no reason for computers to deviate from a consistent approach to sentencing. Thus, a computerized sentencing system could ensure that similar sentences are produced where the facts of crimes are alike.

---

133. Gertner, *supra* note 132, at 697; *see also* Divine, *supra* note 132, at 790–91.

134. It was also noted that there were lower differences in some districts. *See* Gertner, *supra* note 132. *See also* Divine, *supra* note 132, at 792. In relation to the Federal Guidelines, see generally Alan Ellis & Mark Allenbaugh, *Unwarranted Disparity: Effectively Using Statistics in Federal Sentencing*, BLOOMBERG LAW: WHITE COLLAR CRIME REPORT (2017).

135. *See* Bagaric, *supra* note 16, at 105–07; see *also* Benedetto Neitz, *supra* note 16, at 158–61.

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

137. RICHARD SUSSKIND, TRANSFORMING THE LAW: ESSAYS ON TECHNOLOGY, JUSTICE AND THE LEGAL MARKETPLACE 173 (2000).

Sentencing has been a feature of the artificial intelligence and law movement since at least the 1980s. One early system, Sentencing Advisor, developed in 1989, was quite sophisticated for its time.[138] It was a rules-based inference engine which could operate in both forward chaining and backward chaining modes. It could prompt the user to enter more data if a rule generated an unknown quantity, and its code also contained a BECAUSE statement enabling it to produce statements of which inference rules it had applied, and in which order, to come to a decision.[139] The inference rules which drove Sentencing Advisor were based on the U.S. Sentencing Commission's Sentencing Guidelines which, although nebulous and laborious to apply, are at first blush, an ideal candidate for an expert system. This is becuase the process for applying them is mechanistic, based on set quantities and any departure from them occurs after the application of all prescribed factors. So any subjective factors relevant to the sentence could be considered by the judge once the algorithm has reported. Sentencing Advisor did not include any actuarial function, however, such as predictions of recidivism.

Recent approaches are able to model more sophisticated aspects of the sentencing process. An algorithmic sentencing program would, in Hutton's words, comprise "a set of rules describing the criteria which should be taken into account and the method through which account is to be taken," and "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."[140] In developing automated sentencing systems, it is important that a constant, unvarying suite of factors that inform penalty be used—including aggravating and mitigating considerations that increase or decrease penalty respectively—and the weight to attach to each of those factors can be determined by the machine learning techniques.[141] Underpinning those elements and their impact on penalty would

138. *See generally* Gruner, *supra* note 36. Although users complained that the system involved significant access delays because the 200 inference rules, in the form of IF-THEN statements, needed to access the full sentencing guidelines which were stored on floppy disks rather than a HDD.

139. With some more recent actuarial algorithms which predict risk (such as Compas, short for "Correctional Offender Management Profiling for Alternative Sanctions), significant controversies have arisen as the algorithms (and therefore the weightings attributed to individual factors) are the proprietary interest of the company which develops them. This opacity has led to accusations of unfairness, masked bias, and breaches of procedural fairness. But some other actuarial algorithms, such as the Public Safety Assessment-Court tool (PSA-Court tool), used by judges to assist in predicting the likelihood of a person re-offending if granted bail have avoided these controversies. They are less complex, consisting of just nine factors, all concerning criminal history and there is no questionnaire. The PSA's development was funded by a philanthropic organisation, does not use gender or race as predictive factors, and is not black-boxed. Jason Tashea, *Risk-Assessment Algorithms Challenged In Bail, Sentencing And Parole Decisions*, A.B.A. J. (Mar. 1, 2017), http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole/?/ [https://perma.cc/LNY4-96HG].

140. *See* Hutton, *supra* note 136.

141. *See supra* Part II.A.

be clearly articulated objectives that the sentences are designed to achieve. These objectives include: rehabilitation, community protection, incapacitation of serious sexual and violent offenders, and punishment that is commensurate with the seriousness of an offense.[142] Hutton emphasizes that incorporating "the principle of proportionality" into computerized sentencing programs in particular can "increase the formal, generalizable, rule-governed aspects of sentencing and thus provide a more rational basis for sentencing" and result in more consistent sentencing decisions.[143] To ensure that computerized sentencing leads to proportionate sentencing, calculations of the extent to which certain offenses set back the interests of their victims could also be incorporated into the algorithm.[144]

Hutton envisages an ideal 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."[145] A clear set of variables would be applied, and judicial bias that can at present lead to inconsistencies in sentences would be eliminated from the decision-making process.

In order to make sentencing fully transparent, it is important to produce a publicly-accessible algorithm that clarifies the variables and integers that are taken into account in sentencing and the weight that is attached to them, as well as the objectives of sentencing. At present, sentencing determinations can be influenced by judges' particular prejudices. As Eric Engle observed, "Courts generally 'duck' the question of exactly how they weight the [varying] interests," and "modeling law by computer" can eliminate judicial discretion and discrimination and articulate precisely how various interests are baanced in the decision-making process.[146] Indeed, Susskind observes that AI-based systems are, by their nature, "usually … transparent" because they "can generate explanations of the lines of reasoning that lead them to their conclusions."[147]

Another significant advantage that would ensue from introducing computerized sentencing is that sentencing decisions would be made much more quickly and efficiently. An algorithm can resolve a problem significantly faster than a human, so computerized sentencing could greatly reduce the current time between when an offender is found

142. 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).

143. Hutton, *supra* note 136, at 565.

144. *See* Bagaric & Gopalan, *supra* note 142, at 230.

145. Hutton, *supra* note 136, at 552.

146. Eric Engle, *Legal Interpretation By Computer: A Survey of Interpretive Rules*, 5 AKRON INTELL. PROP. J. 71, 92–93 (2011).

147. SUSSKIND, *supra* note 137, at 183.

guilty and when a sentence is imposed.[148] In producing sentencing determinations in a timely fashion, computerized sentencing could ameliorate the numerous adverse ramifications stemming from delayed sentencing decisions.

The consequences of long delays in making sentencing decisions include clogging of the court system and increased costs to the public purse. Perhaps even more importantly, the longer it takes for decisions to be made about sentences, the longer offenders must wait to learn of their fate and victims must postpone their sense of resolution. Often neither the offenders nor the victims can proceed with their lives while the sentencing decisions remain unresolved. This infringes the universal maxim that "justice delayed is justice denied." As Stefan Voigt observes, detaining suspects while they wait for their trial is a serious rights violation, and "[o]verly long court delay is not only likely to threaten the legitimacy of a country's judicial system, but can also lead to a loss in legitimacy of the political system at large," plus it can "have important economic consequences."[149] Indeed, swifter completion of sentencing matters is crucial to promoting the rule of law. Walker maintains that the right to a speedy trial is implicit in the rule of law,[150] as is "[a]ccessibility of courts," by which he means the circumstance where "a person's ability to vindicate legal rights is not made illusory by long delays or excessive costs."[151]

D. AI and Parole

As noted earlier, risk and needs assessment tools are already extensively used in relation to parole determinations in many states. The use of these instruments has increased rapidly over the past three decades. In 1970, only Illinois used an actuarial instrument to determine illegibility for parole.[152] This increased to 28 of the 32 states which had a parole system by 2004.[153] It has been suggested that these instruments are in fact operating to 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. While risk assessment is often understood as a predictive and probabilistic technology that embraces uncertainty . . . in the penal realm it operates in a way that makes risk a certainty. 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

148. Stefan Voigt, *Determinants of Judicial Efficiency: A Survey*, 42 EUR. J.L. ECON. 183, 183–84 (2016).

149. *Id.*

150. WALKER, *supra* note 124, at 5.

151. *Id.* at 40.

152. BERNARD HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 8 (2008).

153. *Id.*

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. . . . Within classification, evaluation, and prediction, there is no outside to risk, no possibility of an absence of risk.[154]

In addition to this, the design of some of the instruments is less than optimal, and the results are contingent on the manner in which the instrument is applied by the user.[155] These disadvantages can be overcome by the use of AI, which can be programmed to combine the formal definitions from risk assessment tools. There, necessary weightings of the relevant variables can be adjusted by the application of 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 on whether actual defendants reoffended during or after the parole period, it is possible to use the machine learning system to build an accurate model of which offenders will reoffend. This approach marries the benefit of assessment based on clear and specific factors (rather than a generalized gestalt model) with the fast, statistical modeling that machine learning promises.

E. The Elephant in the Room: Elimination of Subconscious Bias from Bail, Sentencing, and Parole Decisions

Having examined the key benefits that AI can bring to the bail, sentencing, and parole phases of the criminal justice system, we now focus on the key problem which has been flagged regarding the use of AI in all of these parts of the system. It has been argued that AI will invariably lead to the entrenchment of decisions which involve undue weight being accorded to existing judicial subconscious biases.[156] To assess the validity of this objection, it is important to understand the extent of subjectivity currently associated with sentencing.

Evidence establishes that judges, like most people, view themselves as being fair and objective. Yet they inevitably have preferences and biases, too, which inform their decision-making. Judges can have difficulty recognizing biases in the thought patterns involved in their decision-making,[157] and the most difficult biases to overcome are those of which one is unaware. In *How Judges Think,* Judge Richard Posner states, "We use introspection to acquit ourselves of

154. *Id.* at 329 (emphasis in original).

155. *See* Sarah L. Desmarais et al., *Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings*, 13 PSYCHOL. SERVS. 206, 216 (2016).

156. *See, e.g.*, Frank Fagan & Saul Levmore, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, 93 S. CAL. L. REV. 1, 1 (2019).

157. Jennifer K. Robbennolt & Matthew Taksin, *Can Judges Determine Their Own Impartiality?*, 41 MONITOR ON PSYCHOL. 24, 24 (2010).

accusations of bias, while using realistic notions of human behavior to identify bias in others."[158] People assume that "their judgments are uncontaminated"[159] with implicit bias, but the truth is otherwise. All people, including judges, are influenced by their life journey and "are more favorably disposed to the familiar, and fear or become frustrated with the unfamiliar."[160]

A large number of studies show that the impact of implicit judicial bias in sentencing is significant. Thus, it has been shown, for example, that:

- Attractive offenders receive more lenient penalties than other accused, except when they use their attractive appearance to facilitate the crime.[161]
- The socioeconomic background of parties also influences legal outcomes. An analysis of child custody cases showed that judges favor wealthy litigants to those who are impoverished.[162]
- The racial background of victims can also influence sentencing decisions. For example, multiple studies show that black offenders who harmed white victims were sentenced more heavily than black offenders who harmed black victims.[163]

The subconscious bias of sentencing judges operates especially harshly against offenders from racial minorities. Empirical studies have uncovered that offenders from minority groups, and especially African Americans, sometimes receive more severe sentences than white offenders who have committed comparable crimes.[164] Researchers have found that racial bias has contributed to this disparity, thereby undermining the rule of law. As Walker notes, a critical component of the rule of law is "[t]he rules of natural justice," which include "the requirement of an unbiased tribunal."[165]

An analysis of the sentences of more than 59,250 offenders found that the same courts will sentence black offenders to prison

158. RICHARD POSNER, HOW JUDGES THINK 121 (2008).

159. Timothy Wilson et al., *Mental Contamination and the Debiasing Problem*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 185, 190 (Thomas Gilovich et al. eds., 2002).

160. Rose Matsui Ochi, *Racial Discrimination in Criminal Sentencing*, 24 JUDGES J. 6, 53 (1985).

161. Birte Englich, *Heuristic Strategies and Persistent Biases in Sentencing Decisions*, *in* SOCIAL PSYCHOL. OF PUNISHMENT OF CRIME 295, 304 (Margit E. Oswald et al., eds., 2009). In one study, seventy-seven percent of unattractive defendants received a prison term, while only forty-six percent of attractive defendants were subjected to the same penalty. John E. Stewart II, *Defendant's Attractiveness as a Factor in the Outcome of Criminal Trials: An Observational Study*, 10 J. APPLIED SOC. PSYCHOL. 348, 354 (1980).

162. Bagaric, *supra* note 16, at 106–107; Benedetto Neitz, *supra* note 16, at 158–61.

163. Bagaric, *supra* note 16, at 107; *see also* Siegfried L. Sporer & Jane Goodman-Delahunty, *Disparities in Sentencing Decisions*, *in* SOCIAL PSYCHOLOGY OF PUNISHMENT OF CRIME 379, 390 (Margit E. Oswald et al., eds., 2009).

164. Matsui Ochi, *supra* note 160, at 7.

165. WALKER, *supra* note 124, at 37.

terms that are 22% longer than the sentences they impose on white offenders even where the offenders have committed identical crimes and have identical criminal histories.[166] Similar findings were uncovered by research, undertaken for the United States Bureau of Justice Statistics and the United States Department of Justice Working Group on Racial Disparity, into sentences imposed in the federal jurisdiction pursuant to the Federal Sentencing Guidelines.[167] Factoring in variables recognized by the Guidelines,[168] this study found that, between 2005 and 2012, black male offenders received sentences that imposed prison terms that were longer than the prison terms imposed on white offenders who had committed similar crimes. The same study speculated that the case of *Booker*, in holding that the Guidelines were advisory only, had increased judges' discretion in applying the Guidelines and led to inconsistent sentencing decisions being made for black and white offenders.[169] The report states:

> We are concerned that racial disparity has increased over time since *Booker*. Perhaps judges, who feel increasingly emancipated from their guidelines restrictions, are improving justice administration by incorporating relevant but previously ignored factors into their sentencing calculus, even if this improvement disadvantages black males as a class. But in a society that sees intentional and unintentional racial bias in many areas of social and economic activity, these trends are a warning sign. It is further distressing that judges disagree about the relative sentences for white and black males because those disagreements cannot be so easily explained by sentencing-relevant factors that vary systematically between black and white males . . . . We take the random effect as strong evidence of disparity in the imposition of sentences for white and black males.[170]

---

166. Ronald S. Everett & Roger A. Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANTITATIVE CRIMINOLOGY 189, 207 (2002); David Abrams et al., *Do Judges Vary in Their Treatment of Race?*, 41 J. LEGAL STUD. 347, 356 (2012).

167. William Rhodes et al., *Federal Sentencing Disparity: 2005–2012* 51–56 (Bureau of Just. Stat., Working Paper No. 1, 2015), https://www.bjs.gov/content/pub/pdf/fsd0512.pdf [https://perma.cc/7UZ9-V28D]. This report also systematically documents previous studies in the United States, which support the conclusion that subconscious bias causes racial disparity in sentencing.

168. *Id.* at 22-23.

169. *Id.* at 67-68.

170. *Id.* at 68. A more recent study focusing on sentencing patterns in Florida noted that African-Americans often received markedly longer prison terms than white offenders for the same offense. *See* Elizabeth Johnson et al., *Black Defendants Get Longer Sentences in Treasure Coast System*, DAYTONA BEACH NEWS-J. (Dec. 19, 2016), https://www.news-journalonline.com/news/20161218/black-defendants-get-longer-sentences-in-treasure-coast-system [https://perma.cc/Q33G-SCZZ].

There is also a range of other more subtle factors that have been found to influence the mindset of judges and the decisions they make. Thus, it has been noted that judges who think about negative matters, such as their own death, set bail at higher levels than other judges.[171] Another study observed that judges were far more likely to grant parole if the decision was made shortly after they had taken a meal break than prior to doing so.[172] The researchers speculated on the reason for this:

> [A]ll repetitive decision-making tasks drain our mental resources. We start suffering from "choice overload" and we start opting for the easiest choice . . . . And when it comes to parole hearings, the default choice is to deny the prisoner's request. The more decisions a judge has made, the more drained they are, and the more likely they are to make the default choice. Taking a break replenishes them.[173]

Judges are unlikely of their own volition to reduce the extent to which their preferences can guide their decisions. Posner correctly noted that judges, like all people, are utility-maximizers and hence gain satisfaction from the prestige of their role and the influence they can have in the discharge of their functions.[174] Judges, in making their decisions, give effect to their preferences, which are in turn influenced by their "background, temperament, training, experience, and ideology, which shape [their] preconceptions and thus [their] response to arguments and evidence."[175]

Thus, offenders' immutable characteristics—especially race—can in fact influence sentencing decisions in the current system in various ways. It has been suggested that algorithms which evaluate the risk of recidivism in the sentencing context may also discriminate against offenders with particular immutable traits and entrench

171. Bagaric, *supra* note 16, at 107; Abram Rosenblatt et al., *Evidence for Terror Management Theory: I. The Effects of Mortality Salience on Reactions to Those Who Violate or Uphold Cultural Values*, 57 J. PERSONALITY & SOC. PSYCHOL. 681, 683 (1980).

172. Shai Danzinger et al., *Extraneous Factors in Judicial Decisions*, 108 PROC. NAT'L ACAD. SCI. 6889, 6889-90 (2011).

173. Bagaric, *supra* note 16, at 108 (quoting Ed Yong, *Justice is Served, but More so After Lunch: How Food-breaks Sway the Decisions of Judges*, DISCOVER MAG. (Apr. 11, 2011, 3:00 PM), https://www.discovermagazine.com/the-sciences/justice-is-served-but-more-so-after-lunch-how-food-breaks-sway-the-decisions-of-judges [https://perma.cc/CHA4-3KMQ]).

174. POSNER, *supra* note 158, at 35–36.

175. *Id.* at 249; Bagaric, *supra* note 16, at 110.

racism in decision-making about sentences.[176] The same criticism equally applies regarding the use of AI in all parts of the criminal justice system.

However, a properly designed algorithm can exclude the unfair emphasis on offenders' immutable traits. Slobogin aptly notes that "[e]nhancing the punishment of an offender because of gender, age, or any other immutable characteristic strikes some as grossly unfair."[177] Thus, if immutable traits are to be used within the sentencing calculus, we must acknowledge how they operate and must justify why the trait may properly affect sentencing outcomes. Slobogin, again:

> The Supreme Court, however, does not believe that risk assessment is antithetical to criminal justice. It has even approved death sentences based on dangerousness determinations (*Jurek v. Texas* 1976, 275–276). 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 are merely *evidence* of future conduct, in the same way that various pieces of circumstantial evidence that are not blameworthy in themselves.[178]

In the first state appellate decision to consider the appropriateness of risk and needs assessment in sentencing, *Malenchik v. Indiana*,[179] the court concluded that it was not discriminatory for judges to use risk assessment tools that took into account offenders' immutable traits on the basis that sentencing law: 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

176. *See generally* Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/587N-V8UR]; Laurel Eckhouse, Opinion, *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 [https://perma.cc/G4SS-42RV].

177. Slobogin, *supra* note 90, at 203–05.

178. *Id.* at 205.

179. 928 N.E.2d 564 (Ind. 2010).

such personal and sociological data, are effective in predicting the risk of recidivism and the amenability to rehabilitative treatment.[180]

Apart from directly using offenders' immutable characteristics within a sentencing algorithm, we need to be careful of data that are proxies for these characteristics. Traits such as race can indirectly be incorporated into sentencing variables. Notably, the inclusion of prior criminality as a consideration in risk assessment tools and as an aggravating factor in sentencing determinations can have the effect of discriminating against African-American offenders because more African-Americans have prior convictions than white Americans.[181]

However, a sophisticated and nuanced risk assessment algorithm can be readily developed which is cognizant of not impliedly adopting discriminatory considerations and which can be used to overcome racial and other biases in sentencing, bail, and parole decisions. The algorithm would set out the relevant considerations that it takes into account, so that immutable characteristics will only be incorporated into the formula if it is definitively established that they can have an impact on the risk of reoffending, as opposed to being a proxy for other considerations such as deprived social and economic background. Further, if the algoithm is developed carefully with a focus on preventing the operation of factors that lead to indirect discrimination, it can minimize the potential for considerations such as race to influence sentencing outcomes inappropriately.

The results of significant research into the effects of race on one risk assessment tool in particular—the PCRA—which were published in 2016, illustrate this point. A study undertaken by Jennifer Skeem and Christopher Lowenkamp analyzed risk assessments that had been conducted using the PCRA in relation to 34,794 federal offenders in order to recommend conditions for their probation.[182] In addition to finding that the PCRA was accurate in more than 70% of cases,[183] the authors discovered the following:

> First, there is little evidence of test bias for the PCRA. The instrument strongly predicts rearrest for both Black and White offenders. Regardless of group membership, a PCRA

180. *Id.* at 574 (quoting Ind. Code Ann. § 35-38-1-9(b)(2) (West 2017)).

181. *See* Mirko Bagaric, *Three Things That a Baseline Study Shows Don't Cause Indigenous Over-Imprisonment; Three Things That Might But Shouldn't and Three Reforms that Will Reduce Indigenous Over-Imprisonment*, 32 HARV. J. RACIAL & ETHNIC JUST. 103, 151 (2016). *See generally* Anya Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257 (2020) (defining proxy discrimination as a pernicious subset of disparate impact, and providing strategies to avoid it).

182. Jennifer L. Skeem & Christopher T. Lowenkamp, *Risk, Race, & Recidivism: Predictive Bias and Disparate Impact,* 54 CRIMINOLOGY 680, 680 (2016). Risk assessments have no impact on sentencing decisions in the federal system, so Skeem and Lowenkamp did not examine the results of the application of the PCRA in relation to sentencing. *Id.* at 686.

183. *Id.* at 689.

> score has essentially the same meaning, that is, same probability of recidivism. So the PCRA is informative, with respect to utilitarian and crime control goals of sentencing. Second, Black offenders tend to obtain higher scores on the PCRA than White offenders (d= .34; 13.5 percent nonoverlap). So some applications of the PCRA might create disparate impact—which is defined by moral rather than empirical criteria. Third, most (66 percent) 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.[184]

Thus, offenders' immutable traits should not influence criminal justice decisions unless there is clear and persuasive evidence that they are relevant to an important objective of sentencing. It is possible to ensure that computerized decision-making follows these protocols, and that it does not lead to the imposition of more severe outcomes on offenders from particular racial and social groups than on others. Indeed, computers can achieve this outcome far more effectively than judges. It is well established, for example, that young people are more likely to commit crimes and to recidivate than aged offenders, and hence it would be appropriate to incorporate age into criminal justice algorithms.[185] The same consideration applies in relation to gender, given that men commit more crimes and reoffend at higher rates than women.[186]

---

184. *Id.* at 700.

185. *See e.g.*, THE OSBOURNE ASSOCIATION, THE HIGH COSTS OF LOW RISK: THE CRISIS OF AMERICA'S AGING PRISON POPULATION 5 (July 2014), http://www.osborneny.org/resources/resources-on-aging-in-prison/osborne-aging-in-prison-white-paper [https://perma.cc/Z43F-SFYV]; KIM KIDEUK & BRYCE PETERSON, URB. INST., AGING BEHIND BARS: TRENDS AND IMPLICATIONS OF GRAYING PRISONERS IN THE FEDERAL PRISON SYSTEM, 5 (2014), https://www.urban.org/sites/default/files/publication/33801/413222-Aging-Behind-Bars-Trends-and-Implications-of-Graying-Prisoners-in-the-Federal-Prison-System.PDF [https://perma.cc/4DC9-ST2C]; KIM STEVEN HUNT & ROBERT DUMVILLE, U.S. SENT'G COMMISSION, RECIDIVISM AMONG FEDERAL OFFENDERS: A COMPREHENSIVE OVERVIEW, 23 (2016), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2016/recidivism_overview.pdf [https://perma.cc/3F23-77QQ]; U.S. SENT'G COMMISSION, MEASURING RECIDIVISM: THE CRIMINAL HISTORY COMPUTATION OF THE FEDERAL SENTENCING GUIDELINES 12 (2004), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2004/200405_Recidivism_Criminal_History.pdf [https://perma.cc/T4VQ-RMV3].

186. MATTHEW R. DUROSE ET AL., BUREAU OF JUST. STAT., RECIDIVISM OF PRISONERS RELEASED IN 30 STATES IN 2005: PATTERNS FROM 2005 TO 2010 6 (2014), http://www.bjs.gov/content/pub/pdf/rprts05p0510.pdf [https://perma.cc/2EH2-LS3F]; FLA. DEP'T OF CORRECTIONS, 2011 FLORIDA PRISON RECIDIVISM REPORT: RELEASES FROM 2003-2010 8 (2012) http://www.dc.state.fl.us/pub/recidivism/2011/gender.html [https://perma.cc/JE2X-FFN9].

In contrast to humans, computers have no instinctive, unconscious bias, are incapable of inadvertent discrimination, and are uninfluenced by extraneous considerations, assumptions, and generalizations that are not embedded in their programs. They operate simply by applying variables that have been previously identified and data drawn from past events. Bias can infiltrate computerized decision-making only if an algorithm incorporates existing variables or data that result in disproportionately harsh sentences being imposed on offenders from certain groups. Consequently, for computerized decision-making to eliminate bias from sentencing decisions—and, indeed, ensure that racially-based decision-making is not entrenched as a consequence of it—the algorithm and data must be free of the discrimination that permeates the present sentencing regime. Systems need to be designed so that they do not include any integers that could have this effect by virtue of their implicit bias, and datasets must be assessed to ensure that they are clean, complete, and free from discriminatory proxies.

It is important to emphasize that the integers that influence sentencing outcomes must be transparent and set out clearly in a manner that is comprehensible to all people involved in the criminal justice system and the wider community. Promulgation of the algorithms and data that are used in computerized sentencing will reassure all interest groups including offenders, victims, and the community generally. Controversy recently erupted concerning a judge's sentencing of a Wisconsin offender to six years in prison on the basis of a computer program's assessment of his risk of recidivism, because the algorithm for this software had been kept hidden from the public.[187] The company that produced the software claimed that the algorithm was "a trade secret," but as Liptak observed, this unfairly prevented the offender from challenging the risk assessment.[188] Liptak aptly commented, "[t]here are good reasons to use data to ensure uniformity in sentencing. It is less clear that uniformity must come at the price of secrecy."[189] We agree with this criticism, but it can be readily surmounted by ensuring that all of the elements of the sentencing decision-making system are publicly disclosed.

V. NEXT STEPS

Data-driven AI systems are having a profound effect in all areas of human society. It is inevitable that they will also assume a greater role

187. Adam Liptak, *Sent to Prison by a Software Program's Secret Algorithms*, N.Y. TIMES (May 1, 2017), https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?_r=0 [https://perma.cc/H8KX-QD6M]. *See also* Tashea, *supra* note 139.

188. Liptak, *supra* note 187.

189. *Id.*

in the law and particularly in criminal law. In order for this uptake to occur in a methodological and systematic manner there are several broad developments that need to occur.

The first key step that needs to be undertaken in order to meaningfully enhance outcomes in the criminal justice system through the use of AI is to conduct a systematic evaluation of the extent to which AI is currently used in the criminal justice system and to assess its efficacy. As we have seen, AI is already used in a number of different parts of the criminal justice system, including predictive policing, crime detection, probation decisions, and, to a lesser degree, sentencing. The use of AI in these realms has occurred organically without an overarching assessment of the benefits, disadvantages, and possible dangers of AI. Given the *ad hoc* evolution of AI into parts of the criminal justice system, it is understandable that there has not been a considered, let alone systematic, evaluation of the impact of this technology, or a strategy put in place for the future use and application of AI in this sphere.

In light of the already considerable reliance on AI in some parts of the criminal justice system, there is now a pressing need to evaluate the key uses of the technology. This needs to be undertaken by reference to a number of different criteria. In relation to predictive policing, it is necessary to ascertain whether this leads to crime reduction and in a manner which is cost effective. Moreover, it is important to ensure that this system does not lead to the targeting of racial or social groups. In the context of probation and sentencing, greater research needs to be undertaken regarding whether algorithmic tools are better at predicting recidivism than other techniques. In addition to this, as with predictive policing tools, it is necessary to ascertain whether the current tools involve racial bias.

Once the current tools have been evaluated, greater clarity will emerge regarding their functionalities. This will provide a reference point for the future development and refinement of the technology. The criteria for the future enhancement of AI should be centered upon the more efficient attainment of the cardinal criminal law objectives, in the form of crime reduction, protection of the community, and the consistent and proportionate punishment of offenders. A key advantage of AI compared to human decision-making is that all of the variables are determined, as is the formula through which they are processed. This can make the workings of the criminal law far more transparent and predictable, thereby providing more confidence in the integrity of the system, including proof the system does not operate in a discriminatory manner.

However, transparency in terms of the public disclosure of the algorithms that underpin the AI processes is not tenable in relation to all of the settings where AI will operate in the criminal justice

space. From this perspective, the system can be divided into two broad areas. One area concerns how we deal with the fallout of crime. The key processes here are sentencing determinations and parole decisions. It is desirable that the programs which are developed to facilitate these systems, including risk assessment evaluations, are made publicly available in order to inform relevant stakeholders (namely the judiciary, prosecutors, defence lawyers, defendants, and victims) of the relevant variables and data, so that we can make an assessment of the appropriateness of the sentencing and probation decisions made by these systems. Disclosure of the decision-making methodology will also facilitate the ongoing refinement and improvement of the algorithms and generate good data hygiene standards.

However, different considerations relate to the other main part of the criminal justice system: those relating to the detection of crime and apprehension of criminals. It is not desirable to publicly disclose the factors that influence the manner in which police utilize their resources in order to detect crime. Public disclosure of this information would provide criminals and potential criminals with information that could be used to reduce their likelihood of detection and apprehension. If information was made publicly available that, for example, police directed most of their resources to one geographical area in the hours from 3:00 p.m. to 6:00 p.m. because that was a crime hot spot during these periods, criminals would then almost certainly commit crime in other areas or times to reduce their likelihood of apprehension. Similar considerations apply regarding the type of interactions and events which will raise alarms through the automated analysis of CCTV. While many alarm triggers would presumably be obvious to many offenders, for example, the throwing of a punch in the direction of another person, there are some more subtle transactions that would not be obvious and should not disclosed due to the likelihood that criminals will tailor their behaviour to fit within the normality parameters of the detection algorithm. Thus, for example, if the algorithm identified as an indication of drug trafficking that a person would approach four individuals within one hour, drug dealers would simply approach no more than three people in any one hour period. Nevertheless, in relation to these matters it remains important that the algorithms are validated to ensure that they are effective in detecting crime and are racially neutral. For this to occur an expert panel comprising former judges, prosecutors, defence lawyers, criminal justice scholars, and computer scientists should be appointed to independently review the algorithm.

Once the current use of AI has been reviewed and evaluated, careful research and planning should be taken to enhance the use of AI in the criminal justice sphere. This ostensibly involves the adaption and improvement of existing AI technology as it applies in the different parts of the criminal justice system (such as the detection of crime and

sentencing), however, it is desirable that this process is approached in a cohesive and integrated manner. This is because there are in fact many similarities that unify parts of the criminal justice system and for considerations of efficiency and consistency it is important that an integrated approach is undertaken. The ultimate purpose of the criminal justice system is to reduce the incidence of crime and hence to protect the community. Thus, for example, identifying individuals who are likely to commit crimes is an important component of predictive policing, sentencing, bail decisions, and parole decisions. Information and knowledge that is relevant to one aspect of the criminal justice system will often be relevant to another aspect of it, even though it might assume a different level of importance in the respective realms. In each part of the system it is also obviously important that the algorithms are free from bias.

The development of the algorithms is obviously largely a technical matter; however, an equally important consideration is gaining acceptance of the tools by users and the wider community. To accommodate this it is important that legal experts and leaders in the legal profession are centrally involved in the development and adoption of the AI systems. Unless this occurs, there will be at least passive resistance to AI legal tools. There is clear evidence of this currently in the context of the use of risk assessment tools in sentencing. Brandon Garrett and John Monahan conducted a recent survey of the use of risk assessment tools in sentencing by courts in Virginia and noted that:

> A sizable minority of judges had great discomfort with the goals and the use of risk assessment at sentencing. Some described risk assessment as just "another tool that aids but does not supplant judicial judgment." Others express extreme distaste for risk assessment. For example: "Frankly, I pay very little attention to the [risk assessment] worksheets. . . . I also don't go to psychics." That some judges were not fully cognizant of the availability of risk assessment in sentencing was also unsurprising, given the almost complete lack of judicial training on the subject.
>
> These studies of judicial practice and opinion concerning risk assessment produced several important insights into how to better institutionalize use of risk assessment. To change behavior, it is not enough to adopt a technical tool— attitudes towards the use decision-making need to be addressed if the tool is to be used well. A new approach is needed that takes account of interface between general quantitative risk information and the officials,

> such as judges, prosecutors, and probation officers, who take that information into account in decision-making.[190]

Judges are the key cohort of legal professionals who need to be most heavily involved in the development of AI legal tools given their cardinal role in the legal profession. The reality is that ultimately it is they who will adjudicate upon the validity of the incorporation of AI into the criminal justice sphere, including the appropriateness of AI directed search and seizure, the accuracy of algorithms assessing a defendant's flight risk at a bail hearing, and how much weight to grant algorithms predicting recidivism rates in sentencing decisions.

Judges therefore need to have a voice in the ethical and appropriate use of AI, within judicial decision-making and within the criminal justice system more broadly. Concrete steps to ensure this include:

- The establishment of a judicial taskforce/expert group to investigate and report on the use (present and proposed) of big data and AI methods in criminal investigation, bail, and sentencing;
- The creation of guidelines or heuristics for the design of "ethical algorithms" across all aspects criminal justice system—and especially the judicially-led development of explanatory AI systems for criminal law matters, mechanisms for the control of invisible bias in data systems, and the appropriate use of data-centric AI criminal justice technology;
- The generation of guidelines and standards for the assessment of datasets that are used in criminal justice settings, to ensure that they are appropriate, clean, reliable, and free of discriminatory proxies;
- The creation of a judicial training program in AI and criminal justice, to promulgate knowledge of the technology throughout the judiciary, and to give judges guidance as to the legal issues raised by the technology within criminal justice;
- The establishment of a group of experts to advise the judiciary about the political use of AI systems within the criminal justice system; and
- The development of algorithmic impact statements to assess the potential beneficial and detrimental effect of any proposed rollout of an AI system within criminal justice.

Once the algorithms and datasets are adopted we suggest that they operate in a supportive rather than prescriptive manner. Thus, for example, in the sentencing context judges should determine an

190. Brandon Garrett & John Monahan, *Judging Risk*, 108 CALIF. L. REV. 439, 445 (2019).

offender's likelihood of recidivism and the probable success of rehabilitative interventions by considering the results of risk and needs assessment tools. Nevertheless, they should then have discretion to make decisions that deviate from that information in individual cases (for example, if the offender's profile or nature of his or her offense is atypical). Even if judges do not follow the conclusions derived from risk and needs assessments, merely encouraging them to examine this data will inject greater rationality, predictability, and accuracy into decision-making in the criminal justice sphere. The same consideration applies regarding the concrete sentence that is ultimately set out by the sentencing algorithm—it should serve as a reference point for judges but not straitjacket their decision.

VI. CONCLUSION

The criminal law deals with the most harmful conduct in the community. Thus, the community has a strong need to lower the crime rate and to ensure that those who breach the criminal law are apprehended and punished. Rapid technological advances, and in particular the advent of AI, now provide the opportunity for the first time in human history for crime to be considerably curtailed and to ensure that criminals are dealt with transparently, fairly, and efficiently.

AI has the capacity to profoundly influence and improve the workings of all parts of the criminal justice system. Data-driven algorithms can predict where crime is likely to occur and this can be supplemented by live-time recording of criminal acts. This will not only result in the apprehension of far more criminals but also deter many individuals from offending in the first place. AI can also be used to determine whether offenders present a meaningful risk of reoffending. This information can be used to enable sentencing courts to tailor sentences to secure higher levels of community safety. In relation to bail determinations, it is also possible to distinguish with a high degree of accuracy offenders who will reoffend or abscond from those who will not. Thus, AI has the capacity to fundamentally reshape the manner in which we approach crime and punishment. Importantly, crime reduction will occur at two significant stages. First, through the use of predictive policing and secondly by maximizing the likelihood that offenders who are likely to reoffend will receive harsher sanctions, often in the form of longer prison terms. Thus, AI has the potential to massively reduce the incidence of crime.

However, there are several possible disadvantages associated with the greater use of AI in the criminal justice sphere. The key disadvantage is that it may systematize decision-making which is biased against already disadvantaged groups, such as African Americans. This problem can, however, be surmounted by the careful

development of the relevant algorithms to ensure that they are free of actual and subconscious integers that have a disproportionately adverse impact on disadvantaged groups.

But there is no point in arguing that we should not use AI systems, or even place a moratorium on their use in the criminal justice system. They will be used, and in the years to come machines will make more and more decisions in law enforcement and criminal law. The important thing is to understand the computer systems and how they can best be used to improve our currently flawed criminal justice system.