How Algorithms Expose the Law

Colin Doyle

Abstract. This Article proposes a new role for legal algorithms as a tool for critiquing the law. To date, algorithms have been used as a tool for improving how legal predictions are made — typically by reducing the influence of human bias and human error. These algorithms have come under fire, particularly in the domain of criminal law, where they have been shown to produce racially biased and inequitable outcomes. But oftentimes legal algorithms are not independently *causing* harm — they are instead *revealing* the harm of the laws that they apply. In fact, machine learning algorithms have a hidden talent for exposing how legal predictions work and what outcomes they can produce. This Article coins the phrase "algorithmic exposure" to describe how basic machine learning techniques can be used to assess a law's efficacy and fairness. Through this process, legal algorithms can provide empirical support for critiques of law by revealing how a law falls short of its goals or systematically burdens certain groups.

Algorithmic exposure should be adopted as a routine, best practice for assessing laws that involve prediction. This Article uses pretrial incarceration doctrine as lens for surveying how algorithmic exposure could be incorporated within political advocacy, legislation, and agency policy. Advocacy groups could refine their policy positions and galvanize the public by using algorithms to reveal how contemporary pretrial incarceration laws are ineffective and racially inequitable — no matter how they are applied. Likewise, legislatures could use algorithms as a diagnostic tool for calibrating potential legislative reforms. Independent of statutory changes, prosecutors could use algorithmic analysis of pretrial doctrine to justify — both in court and to the public — internal policy reforms that reduce pretrial incarceration. Taken together, these examples reveal how legal algorithms could become an unexpected ally in efforts to transform our legal systems.

Author. Climenko Fellow and Lecturer on Law, Harvard Law School; Affiliate, Berkman Klein Center for Internet & Society, Harvard University. I am thankful for the helpful input from Susannah Barton Tobin, Rebecca Crootof, Kiel Brennan-Marquez, Andrew Crespo, Nikolas Guggenberger, Christopher Havasy, Thomas Kadri, Benjamin Levin, Amanda Lewendowski, Leah Litman, Martha Minow, Ngozi Okidegbe, Peter Salib, Alicia Solow-Niederman, Carol Steiker, and Sarah Winsberg; for thoughtful comments from participants in the Junior Law & Tech Workshop, Yale Information Society Project, and Crimfest 2021; and for the support of my fellow Climenko Fellows.

Introduction

Legal algorithms seem up to no good. Public assistance algorithms have incorrectly accused and fined thousands of innocent people for alleged unemployment fraud.[1] Child welfare algorithms have instructed agencies to investigate families for child abuse based, in part, upon the family's poverty.[2] And criminal risk assessments have told judges to incarcerate people to protect the public, even when it's almost certain that those people will not commit a violent crime if released.[3] Predictive algorithms were once heralded as an objective way to tackle mass incarceration and its racial disparities, but wherever algorithms have taken over, a familiar pattern emerges: the algorithms make predictions based upon impermissible factors,[4] produce inequitable outcomes,[5] and take unjustifiable state action.[6] Algorithms intended to challenge mass incarceration seem instead to be perpetuating it.

A new field of study of algorithmic fairness, accountability, and transparency has appeared on scene to contain the damage, and a common pattern has emerged. After a legal system replaces humans' predictions with a predictive algorithm, researchers acquire data that reveals that the algorithm is either following unfair rules, producing unfair consequences, or both.[7] Conscientious data scientists explore ways to recalibrate the algorithm to produce fair results within legal constraints.[8] When this effort falls

---

[1] Robert N. Charette, *Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold*, IEEE Spectrum: Technology, Engineering, and Science News (January 24, 2018), https://spectrum.ieee.org/riskfactor/computing/software/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold.

[2] Virginia Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, Wired (January 15, 2018), https://www.wired.com/story/excerpt-from-automating-inequality/.

[3] Chelsea Barabas, Karthik Dinakar & Colin Doyle, *The Problems With Risk Assessment Tools*, N.Y. Times (July 17, 2019) https://www.nytimes.com/2019/07/17/opinion/pretrial-ai.html.

[4] *See* Sonja B Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 Stanf. Rev. 803, 805 (2014).

[5] *See* Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[6] *See* Charette, *supra* note 1. This is a non-exhaustive list of criticisms of legal algorithms, limited to the issues explored in this article. A separate — and important — line of criticism evaluated how legal algorithms fail to adhere to procedural safeguards such as transparency in decision-making, the opportunity for a hearing, and so on. *See generally* Danielle Keats Citron, *Technological Due Process*, 85 Wash. Univ. Law Rev. 1249, 1251 (2008); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 Stanf. Rev. 1343, 1346 (2017).

[7] *E.g.*, Angwin et al., *supra* note 5.

[8] *E.g.*, Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, ArXiv160905807 Cs Stat (2016).

short, advocates seek to remove the unfair algorithm from the field.[9] Proponents of the algorithm push back: Removing the algorithm would be counterproductive because humans tend to perform worse at the same predictive tasks.[10] A mountain of psychological research has shown how difficult it is for humans to reason statistically and how personal and cognitive biases distort human prediction.[11] Algorithms have been introduced to mitigate these human errors. Here, the algorithmic fairness literature presents a dilemma: We can either accept algorithms, despite their documented shortcomings, or we can defer to humans, who will produce worse results.[12] The discussion of fair legal algorithms often stalls at this peculiar conundrum. Somehow, algorithms are both unjustifiable and the best option available.

Something doesn't seem right. This apparent dilemma is the result of an error in causal reasoning. When an algorithm produces data that reveals unfairness, that unfairness is often attributed to the algorithm. But in many cases, predictive algorithms are not independently *causing* harm — they are instead *revealing* the harms of the laws that they apply.

We've been looking at the problem the wrong way. Replacing humans with algorithms is the way to fix the law *only if* the problem with the law is that humans are making poor predictions. But what if the problem is the law itself? What if the most accurate, least biased applications of our criminal laws would still result in racial inequities and mass incarceration? In these circumstances, replacing humans with algorithms won't ever do the trick. Optimizing unjust laws will simply produce optimized injustice — and worse, it will hide that injustice behind a mask of scientific objectivity.

---

[9] *See*, *e.g.*, American Civil Liberties Union, *How to Fight an Algorithm (ep. 7)*, https://www.aclu.org/podcast/how-fight-algorithm-ep-7 (last visited May 19, 2021); The Leadership Conference on Civil and Human Rights, More than 100 Civil Rights, Digital Justice, and Community-Based Organizations Raise Concerns About Pretrial Risk Assessment (2018), https://civilrights.org/2018/07/30/more-than-100-civil-rights-digital-justice-and-community-based-organizations-raise-concerns-about-pretrial-risk-assessment/ (last visited Aug 14, 2020); Amnesty International, *Ban facial recognition technology*, https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/ (last visited May 19, 2021); Matthew Guariglia, *Technology Can't Predict Crime, It Can Only Weaponize Proximity to Policing*, Electronic Frontier Foundation (2020), https://www.eff.org/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing (last visited Sep 21, 2020).

[10] Alex P. Miller, *Want Less-Biased Decisions? Use Algorithms.*, Harvard Business Review, 2018, https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms (last visited Apr 26, 2021); James Austin, Sarah L Desmarais & John Monahan, *Open Letter to the Pretrial Justice Institute.pdf*, http://www.jfa-associates.com/publications/Open%20Letter%20to%20the%20Pretrial%20Justice%20Institute.pdf (last visited May 19, 2021).

[11] Miller, *supra* note 10. (collecting research).

[12] Sharad Goel et al., *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment*, SSRN Electron. J. (2018), https://www.ssrn.com/abstract=3306723 (last visited Sep 25, 2019).

Perhaps surprisingly, these same algorithms could be recruited as an ally in efforts to transform our legal systems. Rather than accidentally revealing a law's negative outcomes when applying that law, machine learning algorithms could be repurposed to study laws and project their effects. This article coins a phrase — "algorithmic exposure" — for a process wherein standard machine learning techniques can be used to assess a law's fairness and efficacy. Constructing a legal algorithm requires testing a world of possible predictions within a given legal framework. Machine learning is a process of trial and error that involves exploring, primarily through automation, countless ways that a particular prediction can be made.[13] Machine learning could be used to generate a range of predictive models that represent the scope of what predictions are permissible under a law. These predictive models are a sandbox for discovery. By charting the many ways that a prediction can be made under a law, algorithms can reveal the scope of what a law can achieve. Three insights about legal prediction can be systematically extracted: accuracy, distribution of outcomes, and strength of predictive factors. Algorithms' much-ballyhooed strength is that they exploit huge datasets and an array of statistical methods to outperform humans at making predictions.[14] But because algorithms are the best available option for prediction, their shortcomings are telling. Limits on algorithms' predictive accuracy reveal the limits of predictive accuracy possible under a given law. Algorithms can also reveal the possibilities for how a law's outcomes and errors can be distributed across the population. And the process of constructing these models reveals the comparative strength and relevance of different predictive factors.[15] This knowledge can reinforce legal critiques by revealing how a law falls short of its intended purpose, when a law systematically burdens certain groups, or how a law works in arbitrary, inefficient, or redundant ways.[16]

Repurposing predictive algorithms as tool for critiquing law dovetails with a recent turn in algorithmic-fairness scholarship that encourages the field to locate unfairness at structural levels and engage with questions of political economy.[17] As part of a new wave of algorithmic scholarship, this Article seeks to center questions of power, mar-

[13] *Id.* at 21.

[14] Goel et al., *supra* note 12 at 2.

[15] This disparate treatment can happen even when a law is written in neutral language, has not been designed to be discriminatory, and is being administered by court actors who intend to treat people equally.

[16] Conventional algorithmic fairness approaches have had too narrow a view of both law and algorithms. Algorithms' potential has been limited to the role of optimizing existing legal practices. And the scope of algorithmic fairness has been confined to scrubbing a decision-making process clean of overt human bias. Interventions have looked to tweak the algorithm in its current task but have not looked to problems with the task itself. This approach can also synthesize seemingly isolated studies of algorithmic fairness by folding one-off critiques of a particular algorithm's predictions or outcomes into a broader analysis of the background legal process.

[17] *E.g.*, Ben Green & Salomé Viljoen, *Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought* 13 (2020); Chelsea Barabas, *Beyond Bias: Re-Imagining the Terms of 'Ethi-*

ginalization, and structural inequality within algorithmic discourse.[18] This Article follows up on a call that I and co-authors previously made to encourage data scientists and other researchers to shift their research focus away from using algorithms to study only marginalized populations and toward using algorithms to study people and institutions of power and authority.[19]

The payoff is not just academic. This article uses pretrial incarceration doctrine as a lens for examining algorithmic exposure's potential as a tool for political advocacy, legislation, and agency policymaking. Across these domains, predictive algorithms should be repurposed as a routine, best practice for diagnosing the efficacy and fairness of predictive laws.[20] Advocacy groups could use algorithms to clarify their policy positions, educate the public, and galvanize support for their cause. Within the bail and pretrial reform movements, algorithmic exposure can help orient advocacy away from purely procedural reforms and toward substantive doctrinal change. In a political environment in which legal reforms must be evidence-based — and evidence-based reforms are typically minor tweaks to current practices — algorithms can be repurposed to provide empirical support for a more fundamental reshaping of law and policy.[21]

Legislatures and other rulemaking bodies could use algorithms as a diagnostic tool for assessing and calibrating potential pretrial reforms. Algorithmic exposure can enrich the public and legislative debate over predictive laws by breaking prediction down to its constitutive elements.[22] Legislatures and administrative agencies could subject predictive laws and regulations to algorithmic audits to ensure that the laws are able to achieve their goals and are producing fair outcomes. The process of writing predictive laws could include the use of predictive algorithms from start to finish.[23] Algorithms

---

cal AI' in Criminal Law, Geo J Mod Crit Race Persp (2019), https://www.ssrn.com/abstract=3377921 (last visited Oct 27, 2021).

[18] *See* Frank Pasquale, *The Second Wave of Algorithmic Accountability*, Law and Political Economy (2019), https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/ (last visited Jul 31, 2020); Ben Green, *Algorithmic Imaginaries: The Political Limits of Legal and Computational Reasoning*, LPE Project (2021), https://lpeproject.org/blog/algorithmic-imaginaries-the-political-limits-of-legal-and-computational-reasoning/ (last visited May 19, 2021).

[19] Chelsea Barabas et al., *Studying Up: Reorienting the study of algorithmic fairness around issues of power* 9 (2020).

[21] *See* Erin Collins, Shifting the Evidence-Based Paradigm (forthcoming) (on file with the author).

[22] Bernard E. Harcourt, *The Systems Fallacy: A Genealogy and Critique of Public Policy and Cost-Benefit Analysis*, 47 J. Leg. Stud. 419–447 (2018).

[23] Insights gained through the algorithmic exposure of law do not then require that the predictive law be administered by algorithms. There's no iron rule that software must be integrated into the legal process. A law could be made simpler for the sake of honest and transparency while maintaining predictive accuracy.

could be used to test and compare draft legislation to see which versions of a law best adhere to legislative goals and priorities.

Independent of statutory changes, prosecutors could use algorithmic analysis of pretrial doctrine to justify less carceral pretrial policies — both in court and to the public. In recent years, progressive prosecutors have been elected in major cities across the country.[24] These district attorneys have turned away from tough-on-crime rhetoric, denouncing the racism pervasive to the criminal legal system and rejecting incarceration as the only means to protect the public.[25] But the policies that these prosecutors have adopted for money bail and pretrial incarceration may be at odds with their broader message. Although progressive prosecutors talk of a clean break with their predecessors' track record on pretrial incarceration, their approach has been a gentler application of pretrial incarceration doctrine, rather than an alternate paradigm.[26] Algorithms could help progressive prosecutors reshape their pretrial policies by using data from the prosecutor's own district to show how the local community would fare under current, past, and potential policies.[27] If a prosecutor's office chose to turn away from a purely predictive model of pretrial incarceration and adopt pretrial policies more aligned with their broader vision of reform, this data could justify that change in court and in the bully pulpit. And if a prosecutor's office fell short of that commitment, advocacy groups could use these same tools to hold the office accountable.

The Article has four parts. Part I examines a tension at the heart of contemporary algorithmic fairness debates. As algorithms proliferate through our legal systems, they

---

[24] Allison Young, *The Facts on Progressive Prosecutors*, Center for American Progress , https://www.americanprogress.org/issues/criminal-justice/reports/2020/03/19/481939/progressive-prosecutors-reforming-criminal-justice/ (last visited May 14, 2021).

[25] *Id.*

[26] Rachael Rollins On High Bail: "That Is Not How We Operate," , News (2020), https://www.wgbh.org/news/local-news/2020/09/24/rachael-rollins-on-high-bail-that-is-not-how-we-operate (last visited May 19, 2021); Chris Palmer, *Tensions are boiling over between Philly DA Larry Krasner and bail reform advocates*, https://www.inquirer.com , https://www.inquirer.com/news/philadelphia/philadelphia-da-larry-krasner-cash-bail-reform-advocates-20200729.html (last visited May 19, 2021); Colin Doyle, *Chesa Boudin's New Bail Policy is Nation's Most Progressive. It Also Reveals Persistence of Tough-on-Crime Norms.*, The Appeal Political Report , https://theappeal.org/politicalreport/chesa-boudin-cash-bail-predictions/ (last visited May 19, 2021). This approach to pretrial incarceration emerged in the tough-on-crime Nixon era, was codified into law in the 1980s, and has long been opposed by scholars and activists as violating the presumption of innocence and harming communities in a racially discriminatory manner — all without improving public safety.

[27] At its most accurate, preventive pretrial incarceration flips the Blackstone principle on its head: to prevent one person from committing a violent crime, ten legally innocent people must be incarcerated. Colin Doyle, *All Models Are Wrong, But Are Risk Assessments Useful?*, *in* American Society of Criminology DCS Handbook on Corrections and Sentencing: Pretrial Justice (2021).

somehow seem to be both responsible for untold harm and the best option for any predictive-legal problem. This dilemma stems from an error in causal reasoning. Harm that is assumed to stem from an algorithm is often harm that an algorithm has revealed but not caused.

Part II demonstrates how machine learning can be used to expose how a predictive law works and what outcomes it produces. In part because algorithms are the most accurate method of prediction available, they can reveal the limits of what a predictive law can achieve. Algorithms can reveal how accurately an outcome can be predicted and the possible distributions of outcomes and errors within the population. They can also reveal what factors are useful for this prediction and the predictive strength of those factors. These insights could be used to better understand and evaluate laws by showing when a law falls short of its intended purpose, when a law systematically burdens certain groups, and how a law works in arbitrary, inefficient, or redundant ways.

Part III argues that algorithmic exposure should be adopted as a routine, best practice for diagnosing the efficacy and fairness of predictive laws. Pretrial incarceration doctrine serves as a lens for exploring algorithmic exposure's potential within political advocacy, legislation, and agency policymaking. Advocacy groups, legislatures, and government agencies could all repurpose legal algorithms to inform their work of galvanizing political support, crafting legislation, setting policy, and litigating cases.

Part IV takes a step back to address both concerns and opportunities with this approach. Algorithmic exposure may be impeded by biased data, poor data quality, and policies against data disclosure. The process itself may be coopted. And longstanding critiques of empirical approaches to legal analysis apply here as well.

In closing, the article gestures toward a broader vision, of which this article plays one small part. We are at the cusp of dramatic empirical and technological change for our legal system. The future of criminal law and the prospects for racial justice are tied up with the future of machine learning and artificial intelligence. When used to optimize extant practices, legal algorithms consolidate and preserve entrenched power, inequities, and ideology. But this is not the only way that algorithms can be used. There is opportunity yet to harness algorithms' potential to expose inequities that the law creates and sustains.

Background: Legal Predictions

Prediction guides many of the decisions that judges, police, bureaucrats, and other legal actors must make every day. Before issuing a preliminary injunction, a judge must predict whether the plaintiffs will win their case on the merits.[28] Police must have probable cause for many arrests, searches, and seizures to be constitutionally permissi-

[28] *Preliminary Injunction*, LII / Legal Information Institute, https://www.law.cornell.edu/wex/preliminary_injunction (last visited Aug 8, 2021).

ble.[29] State unemployment agencies use predictions of fraud to grant or deny people unemployment benefits.[30] Child welfare agencies triage investigations of suspected neglect based on predictions of which claims will be substantiated,[31] while public housing authorities manage waitlists for housing based on predictions of who will use public housing for the shortest length of time before living independently.[32]

Across legal systems nationwide, algorithmic predictions are replacing or informing predictions traditionally made by humans. Today, algorithms can deny a person government food benefits,[33] send a social worker to investigate a home, or ban a person from flying on commercial airlines.[34] In many places, criminal procedure is now algorithmic from start to finish. Based on predictions of wrongdoing, algorithms encourage police to investigate,[35] judges to incarcerate,[36] probation to surveil,[37] and parole boards to deny release.[38]

Predictive algorithms arrived in law with much fanfare from policymakers and academics. Algorithms were seen as a way to boost predictive accuracy and efficiency, while reducing bias and error.[39] Law, particularly criminal law, is riddled with race and class

---

29 U.S. Const. amend. IV.

30 Charette, *supra* note 1.

31 Eubanks, *supra* note 2.

32 Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018).

33 Lydia X. Z. Brown et al., Challenging the Use of Algorithm-Driven Decision-Making in Benefits Determinations Affecting People with Disabilities 31 (2020).

34 Spencer Ackerman, *No-fly list uses "predictive assessments" instead of hard evidence, US admits*, The Guardian, August 10, 2015, https://www.theguardian.com/us-news/2015/aug/10/us-no-fly-list-predictive-assessments (last visited Sep 21, 2021).

35 Andrew G. Ferguson, The rise of big data policing: surveillance, race, and the future of law enforcement (2017); Rashida Richardson, Jason Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N. Y. Univ. Law Rev. 15, 19 (2019).

36 Doyle, *supra* note 25.

37 Cade Metz & Adam Satariano, *An Algorithm That Grants Freedom, or Takes It Away*, The New York Times, February 6, 2020, https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html (last visited Aug 8, 2021).

38 Angwin et al., *supra* note 5.

39 *E.g.*, Kamala Harris & Rand Paul, *To Shrink Jails, Let's Reform Bail*, N.Y. Times, July 20, 2017, https://www.nytimes.com/2017/07/20/opinion/kamala-harris-and-rand-paul-lets-reform-bail.html (last visited Aug 9, 2021); Jon Kleinberg, Jens Ludwig & Sendhil Mullainathan, *A Guide to Solving Social Problems with Machine Learning*, Harvard Business Review (2016), https://hbr.org/2016/12/a-guide-to-solving-social-problems-with-machine-learning (last visited Sep 4, 2018); Miller, *supra* note 10.

inequities.[40] To the extent that these inequities are the result of imperfect human decision-making within a legal process, algorithms may provide a helpful course correction. Algorithms tend to be both less biased and more accurate than humans making the same predictions.[41] When predicting a person's likelihood of committing a crime if released on bail, a judge may consider the person's apparent race, gender, class, or cultural background. Through implicit bias, this can happen even when a judge is trying *not* to consider that information. But an algorithm will only consider the factors it has been programmed to consider.[42] Judges' predictions are likewise distorted by cognitive biases. Decades of empirical research demonstrate how humans do not reason statistically and tend to systematically overestimate or underestimate the likelihood of events.[43] To the extent that legal decision-making is undermined by humans' imperfect thinking, algorithms should be able to provide a helpful course correction.

But in recent years, algorithms have come under scrutiny. Data-driven tools can bypass some human shortcomings, but they can still follow unfair rules and produce unfair outcomes. Legal algorithms have made predictions using biased data, have produced inequitable outcomes, and have recommended that state actors take unjustifiable actions.[44] And when developers refuse to disclose their data and methods — as they often do — the entire process can be hidden from scrutiny.[45] A burgeoning field of study of algorithmic fairness, accountability, and transparency seeks to address these concerns.[46] In the field of algorithmic fairness, a common pattern is to identify an "unfair" algorithm in some real-world domain, document its harms, and explore ways to produce fair results.[47]

Although there are a variety of ways that algorithmic harm might be redressed, the field has settled on three techniques: recalibrate the algorithm, restrict the algorithm, or remove the algorithm. Recalibrating an algorithm involves changing how the tool works internally: training the tool on different data, generating predictions based on different features, or following different statistical processes for making predictions.[48] Restricting the algorithm means using the algorithm for a different decision-making

---

[40] Becky Pettit & Bruce Western, *Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration*, 69 Am. Sociol. Rev. 151, 151 (2004).

[41] Goel et al., *supra* note 12 at 2.

[42] Ethem Alpaydin, Machine Learning (2016).

[43] See generally Daniel Kahneman, Thinking, fast and slow (2012). Unlike humans, algorithms make predictions without cognitive biases. Goel et al., supra note 12 at 2.

[44] Sandra G. Mayson, *Bias In, Bias Out*, 128 Yale Law J. 2218, 2221–22 (2019).

[45] Wexler, *supra* note 6 at 1349.

[46] Home :: FAT ML, , https://www.fatml.org/ (last visited Aug 8, 2021).

[48] *E.g.*, Richard Berk et al., *Fairness in Criminal Justice Risk Assessments: The State of the Art*, ArXiv170309207 Stat, 3 (2017).

process.[49] And if the algorithm can't be recalibrated and can't be repurposed, removing the algorithm from the field is always an option.[50]

In some circumstances, this menu of options may be sufficient. But in other cases, these options cannot fully address the unfairness they target. Pretrial risk assessments are one such domain. In recent years, actuarial risk assessments have become the hallmark of both bail reform and algorithmic fairness discourse.[51] The tools encourage judges to incarcerate or release a defendant pretrial based on predictions of whether the person will be arrested or miss a court date.[52] Critics of the tools contend that risk assessments are unjustified because they rely on racially biased data, produce racially inequitable outcomes, and have limited accuracy.[53] Conventional approaches to fixing these algorithms have proven insufficient.[54] Neither recalibrating, repurposing, nor removing the algorithms works. The algorithms cannot be recalibrated using unbiased policing and court data, because no such data exists.[55] Repurposing or removing the algorithms only does so much, because pretrial incarceration decisions still need to be made. Only now, judges would have to make these decisions without algorithmic assistance. As supporters of risk assessment tools are quick to remind us, algorithms ought to make less biased and more accurate predictions than judges.[56] So long as these predictions are being made, algorithmic predictions should be preferable to human predictions.

Can walk reader through this more

What to do? The academic literature and public policy debates are replete with permutations of this critique and defense.[57] The discussion always stalls at the same impasse: ~~Risk assessments~~ are somehow both unjustifiable and the best option available. The usual approaches to algorithmic fairness cannot resolve the deadlock.[58]

Algos

[49] Sandra G. Mayson, *supra* note 42 at 2286.

[50] Shira Ovide, *A Case for Banning Facial Recognition*, The New York Times, June 9, 2020, https://www.nytimes.com/2020/06/09/technology/facial-recognition-software.html (last visited Aug 8, 2021).

[51] Colin Doyle, Chiraag Bains & Brook Hopkins, Bail Reform: A Guide for State and Local Policymakers 14 (2019).

[52] Doyle, Bains, and Hopkins, *supra* note 48.

[53] *Id.* at 14–16.

[54] Chelsea Barabas, Karthik Dinakar & Colin Doyle, Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns (2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf (last visited Oct 3, 2019).

[55] Ngozi Okidegbe, *Discredited Data*.

[56] Goel et al., *supra* note 12 at 2.

[57] Sandra G. Mayson, *supra* note 42 at 2227–33.

[58] *Id.* at 2248–49. Both the critique and the defense of risk assessments are incomplete in their own ways. The critics have held back from following their argument to its necessary

Consider the controversy over racial disparities with pretrial risk assessment algorithms. In 2016, a ProPublica investigation led with the headline “There’s software used across the country to predict future criminals. And it’s biased against Blacks.”[59] The report found that a pretrial risk assessment algorithm had different error rates for Black people than white people.[60] The tool incorrectly predicted Black people as being at high risk of future crime much more frequently than it incorrectly predicted white people as being at high risk. And it incorrectly predicted white people as being at low risk of future crime much more frequently than it incorrectly predicted Black people as being at low risk. The report galvanized progressive groups against pretrial risk assessment algorithms,[61] which they still oppose today.[62] The report — and the accompanying dataset — also inspired a wealth of empirical research.[63] Using this data, researchers found that disparate error rates would occur in any predictive system in which different racial groups had different baseline rates of being arrested.[64] In other words, any legal system of purely preventive pretrial incarceration would produce these racial disparities, whether the legal system relied upon algorithms or relied upon humans to make these predictions. It’s now understood that the problem of racial disparities in error rates stems from the legal process of preventive pretrial incarceration —

Break up and include to develop whole picture in above section?

conclusion. If algorithms are both unjustifiable and better than any available alternatives, then the critics’ concerns would seem to precede the introduction of algorithms. Likewise, by only comparing legal algorithms’ performance to humans, the defenders have offered only a partial justification, ignoring the possibility that neither humans nor algorithms can produce predictions capable of justifying important legal decisions.

59 Angwin et al., *supra* note 5.

60 *Id.*

61 The Leadership Conference on Civil and Human Rights, *supra* note 9 at 100.

62 Coalition letter on the use of the PATTERN risk assessment in prioritizing release in response to the COVID-19 pandemic, American Civil Liberties Union , https://www.aclu.org/letter/coalition-letter-use-pattern-risk-assessment-prioritizing-release-response-covid-19-pandemic.

63 *E.g.*, Eugenie Jackson & Christina Mendoza, *Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not*, 2 Harv. Data Sci. Rev. (2020), https://hdsr.mitpress.mit.edu/pub/hzwo7ax4 (last visited Aug 14, 2020); Anthony W Flores, Kristin Bechtel & Christopher T Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks.”*, 80 9 (2016); Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Alexandra Chouldechova, *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, ArXiv161007524 Cs Stat (2016), http://arxiv.org/abs/1610.07524 (last visited Aug 14, 2020).

64 Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Berk et al., *supra* note 45. Absent an intervention that explicitly adjusted predictions based on race. Kleinberg, Mullainathan, and Raghavan, *supra* note 8.

not from risk assessment algorithms.[65] What has not been fully appreciated is that a risk assessment algorithm produced the information that led to this insight.[66]

There's a pattern here. Because an algorithm had brought unfairness to light, unfairness was attributed to the algorithm. But oftentimes the algorithm is not the source of the unfairness. Rather, the algorithm is reflecting unfairness within culture, law, and society.[67] Accordingly, the unfairness brought to light by legal algorithms may often be symptomatic of unfairness in underlying law.[68] As the ProPublica example illustrates, legal algorithms have inadvertently been revealing important information about underlying laws. The next Part of this Article examines how we might exploit this hidden talent and repurpose algorithms as a diagnostic tool.

Concept: Algorithmic Exposure

By revealing information about predictions that the law requires, algorithms can teach us about those laws.[69] Consider the following syllogism: Some laws justify legal judgments based upon predictions. Algorithms can reveal information about these predic-

---

[65] Sandra G. Mayson, *supra* note 42 at 2224–25.

[66] Some critics doubt the value of this insight, preferring alternative definitions of fairness. Flores, Bechtel, and Lowenkamp, *supra* note 60.

[67] Looking "beyond the algorithm" has its own pitfalls. Problems that are structural or systemic are — by definition — pervasive and hard to change. This conception of unfairness risks becoming too abstract and responsibility for the unfairness risks becoming too diffuse.

[68] A new wave of scholarship approaches algorithmic fairness differently by locating unfairness outside the algorithms themselves. In this approach, algorithmic unfairness is often a reflection of deeper structural and cultural problems. Second wave research into pretrial risk assessments has questioned whether these tools can justifiably be relied on to promote criminal justice reform or make incarceration decisions, often concluding that these algorithms entrench and obscure harmful penal ideologies. Chelsea Barabas, *Beyond Bias: Re-imagining the Terms of "Ethical AI" in Criminal Law* 40; Barabas et al., *supra* note 17; Ben Green, *``Fair" Risk Assessments: A Precarious Approach for Criminal Justice Reform* 5 (2018); 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. Univ. Law Rev. 42 (2019). Rather than try to optimize how the algorithms work, this scholarship asks whether an algorithm is needed and what interests it serves. This Article dovetails with second-wave scholarship by reorienting the algorithmic gaze away from marginalized populations and toward institutions of power, asking what machine learning algorithms can reveal about the law itself. Barabas et al., *supra* note 17.

[69] As these tools have proliferated and been subject to research and critique, the assumption has *not* been that they can offer much insight about the law itself. The assumption has been that analysis of how an algorithm applies the law explains only how the algorithm works — not how the law itself works. There have been some exceptions. *E.g.*, Sandra G. Mayson, *supra* note 42; Andrew Ferguson, *The Exclusionary Rule in the Age of Blue Data*, 72 Vanderbilt Law Rev. 561, 594 (2019).

tions. Therefore, algorithms can reveal information about these laws.[70] Like all syllogisms, the conclusion is true only if the premises are true. The first premise — that some laws justify legal judgments based upon predictions — hardly needs proving, as many laws explicitly require this.[71] The second premise — that algorithms can reveal information about these predictions — is the conceptual claim of this article.[72]

Whether done by humans or machines, prediction is fundamentally the same. Patterns from the past are used to anticipate the future. When making predictions, both humans and algorithms attempt to find an underlying pattern that connects information about the present with outcomes in the future.

But machines and humans make predictions in different ways. Humans rely upon intuition, informed by their background experiences and knowledge. With humans, it's hard to know how predictions are being made and whether ~~the results~~ are accurate or equitable. Compared to human prediction, machine prediction is much less mysterious. Machines require more precise programming and information. Predictive models must obey defined rules to produce defined outcomes. With statistical models, it's often possible to observe how a prediction has been made and what outcomes it has produced. In these circumstances, the rules that a law follows and the range of possible legal outcomes can become subject to richer analysis.

those predict~~ions~~ed outcomes

To date, machine learning has been used to optimize legal predictions. But machine learning could also be used to explore and find the limits of legal predictions. Machine learning is a process of testing, through brute-force automation, many ways that a particular prediction can be made. Rather than test many predictive models and discard all but the most optimal model, machine learning could be used to uncover the range of ways that a prediction can be made under the rules of a particular law. This exploratory process could regularly and systematically reveal three insights about a legal prediction:

1. How accurately an outcome can be predicted
2. The potential distributions of outcomes and errors
3. The relevancy and strength of factors used to predict a particular outcome

These algorithmic insights could be used to better understand and evaluate laws by showing how a law meets or falls short of its intended purpose; how a law systematically affects different groups; how a law works in arbitrary, inefficient, or redundant ways; and how seemingly neutral legal predictions include value judgments that may conflict with other legal principles and goals.

This Part starts by comparing human and machine predictions. It then explains how machine learning can be used to expose how a predictive law works and what out-

[70] To put it another way:

[71] *See supra* Part I, cataloguing many of these laws.

[72] This Part demonstrates the validity of that premise. Parts III and IV examine potential applications and implications.

comes it produces. The Part concludes by examining the insights that can be routinely derived from this process and identifying how these insights can be used to assess laws.

Understanding Prediction

Whether prediction is made by humans or algorithms — some fundamental elements of prediction are the same. Patterns from the past are used to anticipate the future. When making predictions, both humans and algorithms attempt to find an underlying pattern that connects information about the present with outcomes in the future. Prediction is only possible for situations in which the world follows regular patterns. Dark clouds foretell rain; obesity portends high blood pressure, and SAT scores indicate ~~college~~ law school prospects. To the extent that factors "X1, X2, X3 … X*n*" reliably preceded outcome "Y" in the past, both humans and machines will predict "Y" when factors "X" arises again.

An important way in which algorithmic and human processes differ is what the computer science community has termed "general intelligence." General intelligence is a type of intelligence that can "possess a reasonable degree of self-understanding and autonomous self-control, and ha[s] the ability to solve a variety of complex problems in a variety of contexts, and [can] learn to solve new problems that [it] didn't know about at the time of their creation."[73] Although machine learning and artificial intelligence have progressed by leaps and bounds in recent years, we are still far from creating machines that have an artificial general intelligence. In contrast, human-oriented systems — like law — depend upon people's capacity for general intelligence. When given any new task, people bring a wealth of background knowledge, habits, common sense, and intuition to bear — for better or worse.

Laws tend to be written with humans' general intelligence in mind. Because humans can rely on general intelligence, the rules for prediction can be vague and under-specified. In law, the rules that must be followed in making predictions are often open-ended or include more factors than one person can juggle.[74] When multiple factors are considered, it's hard to know the proper weight to be given to each factor.[75] Even when the rules for prediction are clear, there will always be uncertainty about a person's fidelity to those rules. Bias — including subconscious bias — and human error can slip in and distort a person's predictions. As a result, we usually don't know how legal predictions are being made. We don't know what a person considers when making a prediction — oftentimes the person making the prediction can't be sure. In the moment, it's hard to identify how accurate or fair those predictions wind up being. Even

---

[73] Artificial Brains VI.

[74] *E.g.*, Cal. Const. Art. I, § 12.

[75] As is shown later, oftentimes the factors themselves bear little relationship, statistically speaking, to the outcome being predicted.

in circumstances in which outcomes of predictions are observed, it's difficult to discern how to improve those predictions.

Consider an administrative agency like a state's child welfare services. The agency is tasked with investigating alleged child neglect and abuse. But the number of potential cases might overwhelm the agency's staffing capacity. Not every case can be investigated. Some set of rules must be adopted to decide which cases deserve investigation. In a purely human-driven system, the rule might be as simple as "prioritize and investigate the highest risk cases." This might not be an optimal rule — perhaps one with great specificity would produce better results — but it is a functional rule. Bureaucrats working for the agency can, upon reading intake forms and case files, sort cases into what seem to them to be high and low priority. Each state actor in each case must make a subjective determination of whether the case before them is "high risk" or not. In such a system, an outside observer can't ascertain what "high risk" means, what factors the state actor has considered, or how accurate the predictions are. Even if a statute defines high risk or lists specific factors for bureaucrats to consider, people are still influenced by extraneous — sometimes impermissible — information.[76]

Because computer systems do not have general intelligence, similar shortcuts are not available when software is programmed to make predictions. If the same child welfare service agency wanted to replace their human-based triage system with an algorithmic system — as many agencies have in recent years[77] — the rules for prediction would need to be more clearly specified.[78] A statistical model cannot be told to simply "prioritize and investigate the highest risk cases." Compared to humans, the algorithmic prediction process is, by mechanical necessity, more rigorous and transparent.[79] When algorithms are introduced, an opaque system of human prediction is replaced with a predictive system that must obey defined rules to produce defined outcomes. Algorithms reduce ambiguity in the predictive process because they require specified predictive

[76] No doubt, such a system may be deeply problematic. Different bureaucrats may have different ideas of what constitutes high risk, resulting in inconsistent decision-making. Some factors that they consider important — like a parent's income or a mental health history — may not be relevant or may be unfair to be used against someone. Bias against marginalized people may result in Black households or single-parent households being unfairly singled out for investigation. But such a system can still operate, however poorly. Human beings' general intelligence can fill in the gaps in the rules, and humans can make their best guesses under the circumstances.

[77] *See e.g.*, Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. Times, January 2, 2018, https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html (last visited Aug 10, 2021); The Allegheny Family Screening Tool, , https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx (last visited Aug 10, 2021).

[78] James et al., *supra* note 13 at 21–24.

[79] At least to those building the system. See *infra* X for a discussion of transparency and data disclosure concerns.

variables and specified outcomes of interest.[80] Algorithms reduce uncertainty about outcomes because training the algorithms requires measuring their accuracy against test data.[81] Because machines make predictions in a methodical, testable way, they can provide a window into the ways that a given prediction can be constructed and the results that can be produced. With machine prediction, both the process and the outcomes of predictions can become knowable.[82]

Building predictive models require some setup. Although it's widely understood that machine learning depends upon large-scale datasets, building models is not yet quite as simple as pulling up an Excel sheet of court data and pressing "go." Decisions need to be made before the predictive model can be trained on the dataset and run in the field. In general, the data must be cleaned and split, the outcome of interest must be specified, predictive variables must be defined and labeled, and statistical methods must be chosen and tested.[83] The first step is to split the data into a training dataset and a testing dataset.[84] The model will learn how to predict accurately using the training data. After the model has been built, it must be tested using the testing data to see how well the model can make predictions with data that it has not been trained on.[85]

An output variable must be specified.[86] The output variable — interchangeably referred to as the "outcome of interest" or the "dependent variable" — is the outcome that the model will try to predict. This outcome needs to be defined in the dataset. Each case in the training dataset ought to specify whether the outcome of interest been met. Without an outcome specified with the training examples, the predictive model will not be able to predict the outcomes based on the other variables.

---

[80] James et al., *supra* note 13 at 21–24.

[81] *Id.* at 29–31.

[82] Some learning methods are less interpretable than others, particularly deep learning methods. The complexity and uninterpretability of a model does not determine its success at prediction. People may become enamored with complicated methods like convolutional neural networks, but for many problems, neural networks can be inferior to even simple statistical methods. The effectiveness of using algorithms to expose law depends, in part, on the interpretability of the learning method chosen because less interpretable methods yield less information about how a prediction is made.

[83] John D. Kelleher & Brendan Tierney, Data Science 97–98 (2018).

[84] Giorgos Myrianthous, *How to Split a Dataset Into Training and Testing Sets with Python*, Medium (2021), https://towardsdatascience.com/how-to-split-a-dataset-into-training-and-testing-sets-b146b1649830 (last visited Aug 10, 2021).

[85] The aim is not to build a model that just predicts the training data well. The aim is to build a model that can be effective at predictions in the field. A good model is effective at making predictions using new data that the model has not been trained on.

[86] James et al., *supra* note 13 at 15.

Input variables must also be specified.[87] In machine learning, input variables are often called "predictive variables" or "features," and in statistics, they are often called "independent variables." They generally refer to the same thing: the information in a dataset that may help predict the outcome the outcome of interest. At an early stage in building a model, many features might be considered as it's not yet known what features will correlate with the outcome of interest. The process of constructing and building a machine learning model requires discovering what variables are predictive of the outcome of interest.[88] Before a model can be trained, these potential predictive variables need to be collected and identified within the dataset.[89]

The next step is choosing a statistical learning method.[90] Within legal domains, the methods that tend to be used are often regression and decision tree models,[91] although this may change in the future, particularly since machine learning is a rapidly changing field. For a given project, researchers often explore multiple statistical learning methods, as different methods present different tradeoffs. Even after choosing a particular statistical method, researchers may explore different variations on the same method — often with the help of automated tools.[92] There is no one-size-fits-all perfect method.[93] One method may work very well on a particular dataset, while a different method may work better on a very similar dataset.[94] Ultimately, the "right" choice

---

[87] *Id.* at 15.

[88] Randy Au, *Data Science foundations: Know your data. Really, really, know it*, Towards Data Science 105–10 (2019), https://towardsdatascience.com/data-science-foundations-know-your-data-really-really-know-it-a6bb97eb991c (last visited Feb 21, 2019).

[89] With machine learning, most of the human-level work is the cleaning and labeling of datasets so that researchers can construct a model that produces meaningful results.

[90] James et al., *supra* note 13 at 21.

[91] Doaa Abu Elyounes, *Bail Or Jail? Judicial Versus Algorithmic Decision- Making in The Pretrial System*, 21 Colum Sci Tech Rev 376, 381 (2020); Jon Kleinberg et al., *Human Decisions and Machine Predictions**, Q. J. Econ., 239 (2017), http://academic.oup.com/qje/article/doi/10.1093/qje/qjx032/4095198/Human-Decisions-and-Machine-Predictions (last visited Sep 4, 2018).

[92] Jeremy Jordan, *Hyperparameter tuning for machine learning models.*, jeremyjordan.me (2017), https://www.jeremyjordan.me/hyperparameter-tuning/ (last visited Aug 10, 2021). The goal is the same: we assume that there is a relationship between the input data and the output data. Different statistical learning methods make different assumptions about that relationship. It's often a process of trial and error to determine which method best fits the project at hand.

[93] Different methods have different benefits and drawbacks. Some methods are more interpretable — that is, humans are better able to understand how and why the model makes predictions. And other methods are more flexible: the model is better able to adjust its shape to fit the data. But flexibility often comes at the cost of interpretability, and vice versa. James et al., *supra* note 13 at 25.

[94] *Id.* at 29.

of a statistical learning method is a domain-specific issue that depends upon the data being used and the problem being solved. Once the data is cleaned, the input and output variables are specified, and a statistical method is chosen, the machine learning can begin.

Assessing Law

These predictive models are a sandbox for discovery. It's an opportunity to explore the universe of possibilities that a legal framework allows.[95] By exploring the various ways that a prediction can be made in accordance with the predictive factors prescribed or allowed by underlying law, algorithmic exposure can reveal the limits of a given legal doctrine or theory.[96] Three insights about legal prediction can be regularly and systematically extracted from these models: accuracy, distribution of outcomes, and strength of predictive factors.

Accuracy. Algorithms can reveal how accurately a legal outcome can be predicted as measures of accuracy are intrinsic to constructing a predictive model.[97] The very process of learning a model is iteratively adjusting how the model considers predictive factors to arrive at a model that most accurately predicts outcomes.[98] Building a model that optimizes accuracy necessarily reveals the limits on our ability to accurately predict an outcome based on the factors being used. Because machine learning models outperform humans at prediction, the limits on a model's accuracy represent the accuracy limits for any form of prediction.[99] Therefore, if a machine learning model can predict who will commit employment fraud with a high degree of accuracy, humans will be unlikely to outperform the statistical model. And if a machine learning model cannot predict who will commit employment fraud with a high degree of accuracy, then humans will still be unlikely to outperform the statistical model.[100]

---

[95] The measurement possibilities extend beyond calculating the results of one specific prediction. Experimenting with different models can reveal how outcomes change when certain factors are weighted differently or removed altogether. With machine learning to inform the process, researchers can learn how changes to the legal rules guiding a prediction would affect the outcomes produced.

[96] The outcomes across many different predictive models can vary, but only by so much. In legal systems that ask prediction to do too much, it can be helpful to learn what cannot be achieved through prediction.

[97] James et al., *supra* note 13 at 29.

[98] Id.

[99] Goel et al., *supra* note 12 at 2.

[100] An assumption built into this claim is that there is an adequate dataset for the machine learning model to be trained upon. Without relevant data, statistical models cannot predict anything and therefore cannot outperform humans. One might expect that a combination

Distribution of Outcomes and Errors. The machine learning process reveals how a law's outcomes and errors will be distributed across the population.[101] If the dataset is labeled for any characteristic — such as the race, gender, or age of a person — the model can produce information about the outcomes and error rates across different groups. The value here rests in the quantity of models that can be produced. The process doesn't just reveal one legal prediction's outcomes and errors. Rather, the process can reveal the distribution of outcomes and errors of the full range of different predictions permissible under law.

Strength and Relevancy of Predictive Factors. The process of constructing a machine learning model can also reveal what predictive variables can be relevant to the prediction at hand and the strength of different predictive variables that are considered. Both machine learning and human intuition are processes of pattern recognition.[102] The question that both humans and machines ask is, how does a change in factor "X" affect output "Y?"[103] Statistical correlation reveals how the presence of "X" relates to outcome "Y." If there's no relationship between factor "X" and output "Y" — in other words, if changes in "X" do not result in changes in "Y" — then that factor is not helpful for making predictions. When humans make predictions, we can guess that that a relationship between "X" and "Y" isn't there. When we attempt the same task with statistical learning, we can learn definitively whether a relationship exists.[104] Likewise, through statistical learning we can learn the relative strength of different factors.[105] This can reveal what factors, compared to other factors, are the most helpful at predicting the outcome. It can also reveal which factors are not independent of each other but are both capturing the same signal.

---

of judges and algorithms would fare better than algorithms on their own, but that's not case. Judicial overrides have been found to produce worse outcomes than algorithms acting alone. Studies of judges, probation officers, and other criminal justice professionals all reveal that human overrides tend to decrease accuracy. *Id.* at 3–4.

[101] James et al., *supra* note 13 at 20.

[102] They both use patterns from the past to predict outcomes in the future. For unstructured prediction, humans rely upon intuition. Consider the example of a multi-factor test in law. A multi-factor test is not a form of structured judgment that walks a state actor through the process of arriving at a prediction. Instead, a multi factor test anchors a judge's perception to certain factors, and the judge makes an intuitive prediction based on a mental consideration of those factors. In contrast, predictive models rely upon the statistical relationship between inputs and outputs. Kelleher and Tierney, *supra* note 79 at 104–05. Intuition is unavailable with machine learning, and statistical correlation replaces it.

[103] With humans, it's intuition and experience. With machine learning models, it's statistical correlation.

[104] *See* Rahil Shaikh, *Feature Selection Techniques in Machine Learning with Python*, Towards Data Science (2018), https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e (last visited Aug 10, 2021).

[105] *Id.*

Insights about accuracy, distribution of outcomes, and strength of predictive factors can be used to ascertain a law's systemic potential. These insights are not exclusive of one another but can work together to reveal: how a law falls short of its intended purpose; how a law systematically affects certain groups; how a law works in arbitrary, inefficient, or redundant ways; and how seemingly neutral predictions include value judgments that may conflict with other legal principles and goals.

How a law falls short of its intended purpose. Algorithms can expose how effective a law is at achieving its own objectives. Accuracy is the essential insight here. Insights into accuracy can reveal how frequently legal decisions will be made based on incorrect predictions. Consider a preventive incarceration law that seeks to imprison people who would harm others and seeks to release people who would not harm others. The ideal application of this law would correctly identify both sets of people: those who would harm others and those would not. In this case, algorithms can reveal how frequently any application of this law makes errors in both directions: either mistakenly identifying people as dangerous or mistakenly identifying people as not dangerous.

How a law systematically affects certain groups. If the errors in prediction fall upon certain groups rather than others, the law may violate commitments to equal treatment and distributional fairness. Through algorithmic construction, one can see the distribution of outcomes for not just one particular application of a law or legal doctrine but the whole range of possibilities. If particular ranges of outcomes are better than others, an algorithm can reveal which applications of a legal doctrine should be foreclosed and which should be permitted.

How a law works in arbitrary, inefficient, or redundant ways. By revealing the relevancy and strength of different predictive factors included in the law, algorithmic exposure can reveal whether the predictors included in the law have a statistical relationship with the outcome of interest. Algorithms can reveal that a law relies upon irrelevant or misleading predictive factors. Accuracy matters here, too. If a certain type of prediction cannot be made accurately and reliably, then a law depending on that prediction may produce arbitrary results.

How seemingly neutral legal predictions include value judgments that may conflict with other legal principles and goals. Some factors that are useful for prediction may conflict with other values in the law. It may be unfair to consider these factors at all or it may be unfair to weigh them so heavily. For example, when predicting whether a parent has committed child abuse or neglected their child, a useful predictive factor is whether that parent was involved in the child welfare system when they were a child.[106] But using an immutable aspect of a person's childhood to predict that person's current actions may betray a value of the legal system to treat people as moral actors with independence and dignity. Algorithms cannot resolve these value-laden questions, but they can reveal how these normative commitments are hidden within seemingly neutral legal predictions.

---

[106] Eubanks, *supra* note 2.

Exposing Law

Machine learning is a process that is often used to optimize a predictive model.[107] With machine learning, computer software independently tests out many ways of making a prediction to arrive at an optimal way of making that prediction.[108] It's a process of almost brute-force experimentation.[109] In iteration after iteration, the model's reliance on predictive factors is slightly adjusted, and then the model's predictive performance on the training data is measured according to some criterion — often accuracy.[110] After testing many variations, the best performing model is chosen.

But optimization isn't everything. Machine learning has more to offer than just an optimized legal prediction — it can reveal the full range of predictions that the law allows.[111] Within many fields, machine learning is used not to optimize but to explore. This can also be done in law. Rather than generate a range of predictive models to arrive at one optimal version of the law, machine learning could be used to generate a

---

[107] Machine learning used for predictions is called "supervised learning." This Article is only concerned with applications of supervised learning. Other popular applications of machine learning include unsupervised learning and reinforcement learning. With unsupervised learning, there is no outcome of interest being predicted. In this context, machine learning is used to better understand the relationship of clusters of data to one another. What is Unsupervised Learning? | IBM, , https://www.ibm.com/cloud/learn/unsupervised-learning (last visited Sep 17, 2021). Reinforcement learning is a method for allowing a computer agent to interact within an environment and learn through a process of trial and error based upon rewards and punishments for its actions. Reinforcement Learning 101. Learn the essentials of Reinforcement… | by Shweta Bhatt | Towards Data Science, , https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292 (last visited Sep 17, 2021).

[108] James et al., *supra* note 13 at 15. In today's culture, machine learning can seem synonymous with magic. Machine learning is how a website can identify your face from a photograph, how streaming services predict the next show you'd like to watch, and even how cars can learn to drive themselves. Although exceptional applications of machine learning capture public attention, most uses of this technology are more mundane, particularly in law.

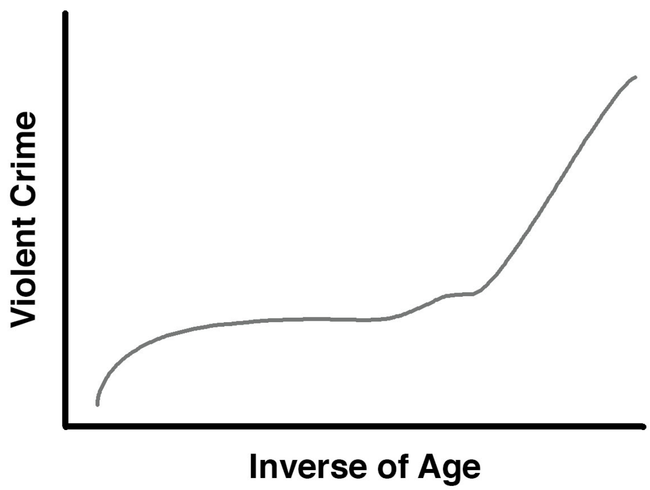
[109] *See* Andreas Stöckl, *Watching machine learning models fitting a curve!*, Medium (2021), https://towardsdatascience.com/watching-machine-learning-models-fitting-a-curve-c594fec4bbdb (last visited Aug 10, 2021).

[110] Ethem Alpaydin, Machine Learning 38 (2016). An algorithms determines the process for exploring potential models. Some algorithms are exhaustive in their exploration of different models, while others seek to save time and computing power by exploring a more limited range of potential models.

[111] This is how machine learning models can reveal information about prediction more generally even when a prediction in law is not clearly specified. For a specific prediction, an algorithm must have a clearly defined outcome of interest and clearly defined predictive factors — even when the law is vague about both. But the process of constructing a predictive model is much more expansive than making one specific prediction. Constructing a predictive model can include considering a multitude of possible variables, and the full range of weights that each of these variables could be given.

range of predictive models that represent the scope of what is possible under the law. The advantage of machine learning is its capacity. With modern computing power, we can generate countless variations of a prediction. Any particular version of a predictive model represents just one way of making a prediction under the law. But taken in total, the many versions of the model could represent the range of possible ways of making a certain legal prediction. Researchers could use this collection of models to evaluate the results of all the ways that a law can be applied.

A graphical example can help illustrate the concept. Imagine a law that requires judges to consider at sentencing a convicted person's risk for committing a violent crime in the future. Under this law, judges must consider four factors when making this prediction: age, prior arrests, prior convictions, and prior length of imprisonment. The hypothetical true relationship of these four predictive factors with violent crime risk is depicted with the four graphs below:



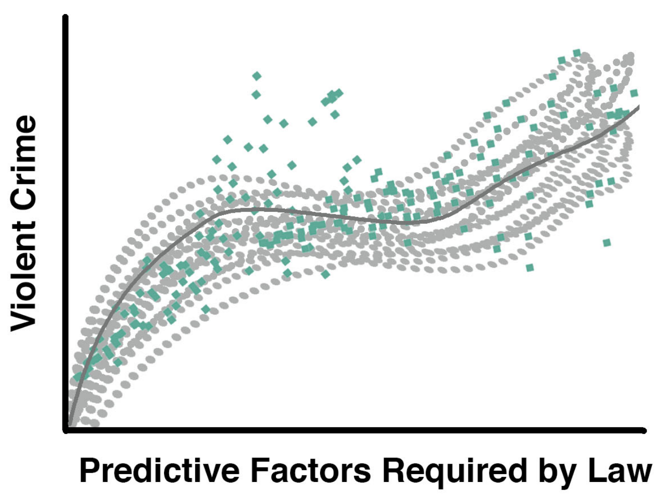
Missing number of pics in this section

These four lines represent the hypothetical true function of how each of these factors correlate with violent crime. To simplify our example, we can reduce the dimensionality of our example and collapse those four graphs into one graph.

This x-axis on this graph represents all four factors. The line represents the hypothetical true function of how these combined four factors correlate with violent crime. In the real world, we don't know what the real relationship is between predictive factors and outcomes. Prediction is a process of estimating that hypothetical true function. Humans do this with intuition. To apply this hypothetical law, judges would rely upon their background and experience to estimate how likely a person is to commit a violent crime based on the factors the law requires them to consider. We can demonstrate in our graph three different judges' understanding of how these factors relate to recidivism.

These lines depict how different judges would weigh different factors differently. None of these predictions exactly matches the hypothetical true function, but each judge in this example would be faithfully applying the law and trying to capture the same real, unknown relationship. Statistical models can be used for the same task. Constructing a predictive model would require data on the predictive factors and case outcomes, depicted in the scatter plot below.

This graph shows a scatterplot superimposed over the line that represents the hypothetical true function of how each of these factors correlate with violent crime. Just as judges rely on experience to inform their predictions, a predictive model would use the available data to estimate the relationship between the factors and the outcome of interest. Optimizing a model means adjusting the shape of the model to better fit the data.[112] The conventional way of using machine learning to optimize legal prediction generates many models. Each is evaluated based on how well it performs. The model that best approximates the true function is selected and the rest are discarded.



The left-hand graph illustrates how the machine learning process generates many models. The right-hand graph pares down the many models so that only the one optimized model remains. Rather than select one optimized model, this Article suggests that we should linger with the left-hand graph and explore a range of models that together

[112] The choice of a particular model is often based on an estimation of the shape of the real function. Linear models will draw a straight line through the data. Logistic models bend. Flexible models make few assumptions about the shape of the real function. Data scientists commonly explore a variety of models to determine what works best for a particular dataset and prediction.

represent the range of predictions that are permissible under a particular law.[113]

Application: Pretrial Incarceration Laws

Algorithmic exposure should be adopted as a routine, best practice for diagnosing the efficacy and fairness of predictive laws. To date, courts, legislatures, advocacy groups, and lawyers have flown blind with predictive laws. We don't know the error rates these laws have, the outcomes they're capable of producing, or the relevancy of the predictive factors included in the law. And so it follows that we have not been able to discern the range of fair and unfair applications of these laws and have not been able to ascertain the legitimacy of legal judgments that have been justified by these predictions. Legal actors and institutions would benefit from knowing how a predictive law works and what outcomes it produces. And there are many ways in which legal, governmental, or private actors might be able to apply that knowledge.

---

[113] There are many ways that this kind of exploration could be structured. A follow-up technical paper examines a series of options. *Statistical Methods for Exposing and Exploring Law with Algorithms* (on file). For the purposes of this Article, it's sufficient to show that this process *can* be structured. One way of setting the boundaries of this exploration would be 1) restrictting models to consider only the predictive factors allowed under law and 2) only include models that satisfy a performance criterion such as accuracy. The purpose of the performance criterion would be to limit the range of models considered to exclude models that are so inaccurate that they would represent a misapplication of the law.

In many cases, it will be difficult to decide exactly how poorly a model must perform to be a misapplication of the law. Even attempting to draw a line may appear to be forcing a quantitative response to a humanistic inquiry, demanding a level of statistical specificity that the law does not typically provide. Although the law often uses probabilistic language, such as probable cause, courts have long resisted formalizing these terms in a mathematical way. Andrew Manuel Crespo, *Probable Cause Pluralism*, SSRN Electron. J. (2019), https://www.ssrn.com/abstract=3342902 (last visited Jul 29, 2021).

Fortunately, the decision over what range of models to exclude is unlikely to influence a global understanding of the law. Although the range of models that permissibly represent the law may be contested, the insights described in the following section can be derived from models that fall squarely in the permissible range — indeed largely from models that outperform humans making the same predictions. Thus, a precise determination of what models to exclude is unnecessary. Although the process described in this Article requires exploring a range of ways to apply the law, the poorest performing models are, by nature, the least helpful. The goal is to use machine learning to explore the limits of what a predictive law can accomplish: how accurate it can be, how its outcomes are distributed, and what predictive factors are most important. Poorly performing models represent the very margins of what is allowable under the law, not how the law ought to be applied. These models don't speak to the limits of predictive accuracy and can only provide redundant information about the strength and relevancy of predictive factors. Depending on the domain, poorly performing model may provide some information about potential distributions of outcomes when predictions are made poorly.

This Part uses contemporary pretrial incarceration doctrine as a lens for exploring algorithmic exposure's potential within political advocacy, legislation, and agency policymaking. Advocacy groups, legislatures, and government agencies could repurpose legal algorithms to expose a legal doctrine's systemic potential and inform their practices of galvanizing political support, crafting legislation, setting policy, and litigating cases. Beyond pretrial doctrine and the applications examined in this part, there are more opportunities for algorithmic exposure worth exploring. This Part's goal is not to be exhaustive but to begin a discussion of how the method might be deployed in practice and what it might help achieve.

Pretrial incarceration is a prime example of a predictive legal doctrine ripe for algorithmic exposure. The decision of whether to jail someone pretrial is a weighty legal decision that is made routinely and systematically and is justified purely by a prediction about the future. Nearly every state legally allows judges to incarcerate some people pretrial based on upon predictions of dangerousness.[114] The justification for pretrial incarceration is purely predictive and forward-looking, not based on the person's guilt in the current case. Looking to first principles of criminal law, people are jailed pretrial not because of retribution or just deserts but purely for the purpose of incapacitation, to prevent those people from harming the community at large.

Because the pretrial field is flush with datasets and algorithmic tools, pretrial laws are already ripe for algorithmic exposure. In the span of just a few years, pretrial algorithmic risk assessment tools have spread across the country to over 1,000 counties in all but four states.[115] If a pretrial risk assessment tool is already in use in a jurisdiction,

[114] KY R. Crim. P. 4.06. Most jurisdictions allow judges to incarcerate only people facing certain, serious violent charges. *E.g.*, Cal. Const. Art. I, § 12. In some states, judges can incarcerate someone only to prevent grave physical harm. *Id*. Defendants awaiting trial in a criminal case may be released on personal recognizance, released on certain conditions, or detained in jail. For a broader background on general bail practices, see generally Criminal Justice Policy Program, Harvard Law Sch., Moving Beyond Money: A Primer on Bail Reform (2016). Most defendants are ordered to be released pending trial. *E.g.*, Cal. Penal Code § 1270 (West 2017) ("Any person who has been arrested for, or charged with, an offense other than a capital offense may be released on his or her own recognizance by a court or magistrate who could release a defendant from custody upon the defendant giving bail."). A person released on recognizance promises to return for future court dates. A person conditionally released must fulfill additional requirements such as posting a money bond, checking in with a pretrial services agency, maintaining employment, staying away from the victim or witnesses, or refraining from using alcohol or drugs.

[115] Some states, like New Jersey, have adopted a uniform risk assessment for every court in the state. Public Safety Assessment New Jersey Risk Factor Definitions, (2018), https://njcourts.gov/courts/assets/criminal/psariskfactor.pdf?c=99i (last visited Aug 14, 2020). But this is the exception. The decision to adopt pretrial risk assessments is more often made at the county level, which can result in a patchwork of risk assessments in use across a single state. In California alone, over a dozen different pretrial risk assessment tools are used, while some counties do not use pretrial risk assessments at all. National Landscape, , Mapping Pretrial Injustice (2020), https://pretrialrisk.com/national-landscape/ (last visited Aug 13, 2020).

then the existing models and data could jumpstart the process of algorithmic exposure.[116] Although dozens of different actuarial pretrial risk assessments exist, they are fairly uniform in design and loosely follow state law for pretrial incarceration predictions. Risk assessment tools use historical court and police data to label a particular defendant as low-to-high risk based on the rate at which people with similar characteristics were arrested or missed court dates while on pretrial release.[117] To make these predictions, the tools' developers build statistical models based on factors that correlate with these outcomes.[118]

Pretrial incarceration laws are also an apt example because this is a unique historical moment for criminal law reform and pretrial reform in particular. There is now broad political support to end mass incarceration, including mass pretrial incarceration. The pretrial incarceration levels in this country are staggering: On any given day, American jails imprison nearly half a million people who have not been convicted of a crime.[119] The astonishing truth is that there are more legally innocent people behind bars in

---

[116] But in many cases, neither the model nor the dataset are available as the third-party developers who create these tools typically assert trade secret protections to shield the underlying data and code from public view. Most risk assessment models would need to be adjusted to more closely match state law's preventive incarceration requirements. Doyle, Bains, and Hopkins, *supra* note 48.

[117] Public Safety Assessment- Risk Factors and Formula, . Consider a modern risk assessment tool like the Public Safety Assessment: Every defendant receives a "new criminal activity" risk score between 1 (lowest risk) and 6 (highest risk) and is either flagged or not flagged for "new violent criminal activity." *Id*. Most pretrial risk assessments also recommend that a judge incarcerate or release a person based on these scores. After calculating a person's risk score, that score is filtered through a "decision-making framework" or "decision-making matrix" that encourages judges to lock up people with high risk scores and free most of the rest. Pretrial Release Recommendation Decision Making Framework (DMF), (2018).

[118] These characteristics often include age, history of arrest, history of convictions, and time spent in jail or prison. Megan T Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and The Double-edged Sword of Youth*, 96 Wash. Law Rev. 1 (2018). Some tools consider only a person's age and criminal history. For example, the Laura and John Arnold Foundation's Public Safety Assessment (PSA) looks to nine risk factors that include age, a defendant's criminal history, and a defendant's history of missed court appearances. Laura and John Arnold Found., Public Safety Assessment: Risk Factors and Formula 2 (2016). The full set of factors includes age at current arrest, current violent offense, pending charges, prior misdemeanor conviction, prior felony conviction, prior violent conviction, prior failure to appear in past two years, prior failure to appear longer than two years ago, and prior sentence to incarceration. *Id*. Based on these factors, the PSA ranks the person on a six-point scale from low to high risk for two pretrial risks, "failure to appear" and "new criminal activity." *Id*. at 3. Other tools are more eclectic and include personal information such as owning a cellphone or renting, rather than owning, a home. The Colorado Pretrial Assessment Tool Revised Report, (2012).

[119] Todd D. Minton & Zhen Zeng, Bureau of Justice Statistics, Jail Inmates at Midyear 2014, at 1 (2015), https://www.bjs.gov/content/pub/pdf/jim14.pdf; Roy Walmsley, World Pre-trial/Remand Imprisonment List 1 (3d. ed. 2016).

America today than there were convicted people in jails and prisons in 1980.[120] Across the country, increases in pretrial incarceration rates are "responsible for all of the net jail growth in last twenty years."[121] With just over 4% of the world's population, the United States has almost 20% of the world's pretrial jail population.[122] Pretrial reform has attracted the support of the media, politicians of both parties, professional organizations, and the public at large.[123]

And yet this opportunity to transform our pretrial legal systems may not be realized. The typical approach for pretrial reform has been to facilitate a better application of pretrial incarceration doctrine — not to reexamine or change the doctrine itself. The bail reform movement has implicitly assumed, at its peril, that the problem with pretrial incarceration is with how the doctrine is applied, not with the doctrine itself. This can be seen with the two hallmarks of contemporary bail reform: procedural protections and algorithmic risk assessments. The purpose of procedural protections is to ensure that incarceration decisions are not made hastily but after a thorough adversarial process. And the purpose of risk assessments is to optimize incarceration decisions by helping judges make more accurate, consistent predictions of dangerousness.[124] These reforms are both designed to help judges apply the law better. To be fair, all things being equal, procedural protections and risk assessments ought to produce better outcomes than the status quo. But all other things don't need to be equal. Reducing the scope of legal change to methods for optimizing pretrial doctrine ignores the possibility that pretrial incarceration doctrine itself should change.

---

[120] Prisoners in 1980, (1981), https://www.bjs.gov/content/pub/pdf/p80.pdf (last visited Aug 13, 2020).

[121] Peter Wagner & Wendy Sawyer, Prison Policy Initiative, *Mass Incarceration: The Whole Pie 2018* (Mar. 14, 2018), https://www.prisonpolicy.org/reports/pie2018.html.

[122] Walmsley, *supra* note 115, at 13; Michelle Ye Hee Lee, *Does the United States Really Have 5 Percent of the World's Population and One Quarter of the World's Prisoners*, Wash. Post (Apr. 30, 2015), https://www.washingtonpost.com/news/fact-checker/wp/2015/04/30/does-the-united-states-really-have-five-percent-of-worlds-population-and-one-quarter-of-the-worlds-prisoners/?utm_term=.d10281e3c39c.

[123] Doyle, Bains, and Hopkins, *supra* note 48 at 11–12.

[124] These algorithmic tools should outperform judges at making predictions. Predictive models driven by millions of data points can more accurately predict recidivism than a judge making quick judgments based on limited information. Laura and John Arnold Foundation, Research Summary: Developing a National Model for Pretrial Risk Assessment 2 (2013). If a jurisdiction uniformly adopts these tools, then pretrial decisions could be more consistent and less influenced by the whims or prejudices of individual judges. And if the risk assessment tools contain or reflect bias along racial, class, gender, or other lines, then the tools have some potential to be analyzed and adjusted — unlike judges, whose biases remain hidden within the inaccessible "black box" of their minds. Pretrial risk assessments have become a divisive issue, and the critiques of these tools are varied. For a helpful summary of a variety of these critiques, see Sarah L Desmarais & Evan M Lowder, Pretrial Risk Assessment Tools 12.

The current crisis in mass pretrial incarceration and the groundswell of support for bail reform and racial justice should invite a closer critique of pretrial incarceration doctrine itself. While it may seem self-evident that people should be incarcerated pretrial based on predictions of dangerousness, the doctrine is of recent vintage and its moral and legal legitimacy has been hotly contested over time. Pretrial incarceration based on dangerousness assessments, a policy first proposed as legislation by the Nixon administration,[125] swept the country in the 1970s and 80s. It permitted courts, for the first time in American history, to legally jail people awaiting trial based on a public safety rationale.[126] Nixon's approach turned away from the conclusions of the Johnson administration's seminal 1967 report "The Challenge of Crime in a Free Society," which had considered pretrial detention as way to reform bail but had determined that it "might well create more of a problem than the imposition of money bail, in the light of the difficulty of predicting dangerousness."[127] Legal scholars, including Laurence Tribe, Caleb Foote, and Ronald Dworkin, opposed preventive pretrial incarceration as violative of due process and human dignity.[128] In a 1987 opinion, *United States v. Salerno*, the Supreme Court found the federal government's new preventive detention scheme to be constitutionally permissible.[129] In dissent, Justice Thurgood Marshall chastised the court for "disregard[ing] basic principles of justice."[130] He warned of "the coercive power of authority to imprison upon prediction" and "the dangers which the almost inevitable abuses pose to the cherished liberties of a free society."[131] In the years following *Salerno*, the debate over preventive pretrial incarceration has petered out.[132] Algorithmic exposure may be one avenue for reigniting this debate and recentering the legitimacy of preventive pretrial incarceration doctrine within broader discussions of pretrial reform.

---

125 Guy M. Blynn, *Pre-Trial Detention*, New York Times, February 9, 1969, https://www.nytimes.com/1969/02/09/archives/pretrial-detention.html (last visited Aug 10, 2021).

126 Kellen Funk, *The Present Crisis in American Bail*, Yale Law J. Forum 1098, 1104–05 (2019).

127 The Challenge of Crime in a Free Society: A Report by the President's Commission on Law Enforcement and Administration of Justice., 98 (1967).

128 Laurence H Tribe, *An Ounce of Detention: Preventive Justice in The World of John Mitchell*, 56 Va. Law Rev. 38 (1970); Caleb Foote, *Comments on Preventive Detention*, J. Leg. Educ. 9.[Insert Dworkin from book].

129 United States v. Salerno, 481 U.S. 741 (1987).

130 *Salerno* 481 U.S. at 755.

131 *Id.* at 766–67.

132 Although some scholars have kept the torch burning. *See* Robin Steinberg, *Freedom Should Be Free: A Brief History of Bail Funds in the United States* 19; Hegreness, *America's Fundamental and Vanishing Right to Bail*; Shima Baradaran, *Restoring the Presumption of Innocence*, 72 55; Jeff Thaler, *Punishing The Innocent: The Need For Due Process And The Presumption Of Innocence Prior To Trial*, Wis. Law Rev. 45; Albert W Alschuler, *Preventive Pretrial Detention and The Failure of Interest-balancing Approaches to Due Process* 61.

This Part explores a variety of practical ways that legal actors and institutions might use algorithmic exposure to challenge and transform pretrial incarceration doctrine. By using algorithms to reveal the systemic limits of contemporary pretrial laws, advocacy groups could refine their policy positions and galvanize the public. Likewise, legislatures could use algorithms as a diagnostic tool for assessing and calibrating potential pretrial reforms. Independent of statutory changes, prosecutors could use algorithmic analysis of pretrial doctrine to justify less carceral policies — both in court and to the public.

Advocacy

Algorithmic exposure can provide advocacy groups with a better understanding of how a predictive law works and what outcomes it produces. Advocacy groups can use these empirical insights to clarify their policy positions, educate the public, and galvanize support for their cause. With money bail and pretrial reform movements, algorithmic exposure can help orient advocacy away from purely procedural reforms and toward substantive doctrinal reform.

Clarifying Policy Positions. Advocacy groups can use algorithmic exposure to clarify their policy positions on pretrial reform by discerning how current pretrial incarceration laws accord with their values. Many advocacy groups tacitly accept pretrial incarceration justified by predictions of dangerousness as the legal backdrop for reform.[133] Instead of targeting the substantive doctrine of pretrial incarceration, these groups have focused their policy guidance and advocacy on promoting procedural protections for pretrial incarceration decisions.[134] These procedural protections are important. Rather than incarcerate people pretrial through imposing unaffordable money bond amounts at a perfunctory hearing, jurisdictions should have robust hearings in which the prosecution must prove its case and defense counsel can cross-examine witnesses and present evidence.[135] But the current trend in bail reform is strictly procedural.

---

[133] Some groups have already taken a strong stance to oppose or support pretrial incarceration based on generalized predictions of dangerousness. Many community bail funds take an explicitly abolitionist stance that opposes incarceration on any grounds.

[134] Doyle, Bains, and Hopkins, *supra* note 48.

[135] The standard way that people are incarcerated pretrial is still money bail. The study of Cook County, Illinois in the appendix is an example of this problem. The county has robust preventive detention procedures and requires judges to determine that a defendant has the ability "to pay the amount necessary to secure his release" before imposing bail. General Order No. 18.8A, Circuit Court of Cook County, (July 17, 2017), http://www.cookcountycourt.org/Portals/0/Orders/General%20Order%20No.%2018.8a.pdf. But judges in Cook County continue to impose unaffordable bond amounts that result in detention. The Coalition to End Money Bail, Shifting Sands: An Investigation Into The First Year of Bond Reform in Cook County 6 (2018), https://www.chicagobond.org/reports/ShiftingSands.pdf (last visited Sep 20, 2018). At an initial hearing following arrest, a judge — or

Advocacy groups ought to expand their focus to reckon with underlying pretrial incarceration doctrine. Algorithmic exposure can help, by showing how a pretrial incarceration law operates and what effects it produces — and can produce. This process would enable advocacy groups to better evaluate how a law could conform with their values — or whether their values and a legal doctrine are incompatible. There's reason to think that algorithmic exposure could help some advocacy groups reconcile a particular tension in their current policy platforms: their opposition to pretrial risk assessment tools and their acceptance of preventive incarceration on predictions of general dangerousness as a background legal doctrine. A core value for progressive advocacy groups is racial justice. A commitment to this principle has led advocacy groups to critique and oppose pretrial risk assessment algorithms because of their racial bias and poor accuracy.[136] But they have not been so outspoken about the legal doctrine that these algorithms apply. Algorithmic exposure can show how their emphasis may be misplaced. The racial inequities and inaccuracy attributed to risk assessment tools are problems inherent to the legal doctrine that these algorithms apply.[137]

By revealing the distribution of outcomes possible under pretrial incarceration laws, algorithms could reveal if a jurisdiction's pretrial incarceration laws result in the disproportionately incarceration of Black people on weak predictions of future violence.[138] There's reason to suspect that in most places the distribution of both adverse outcomes and errors would disproportionately falls upon Black people[139] This consequence is a

---

often a magistrate — can condition a person's release from jail upon the payment of a certain bond amount. Doyle, Bains, and Hopkins, *supra* note 48. Those who can afford bond, or who can at least afford a bail bond company's fee, leave jail. Those who can't pay are locked up until their case is over. In the process, people may be detained pretrial without having an adequate hearing and without having been represented by defense counsel. In recent years, a series of civil rights lawsuits have challenged bail practices in states across the country, and courts have tended to find these practices to be unconstitutional. Funk, *supra* note 122 at 1111. These cases have resulted in federal courts issuing preliminary injunctions against local governments and have prompted many states and counties to consider changing their laws and policies. Civil Rights Corps, *Challenging the Money Bail System* (2020), https://www.civilrightscorps.org/work/wealth-based-detention (last visited Aug 13, 2020).

[136] More than 100 civil rights and community groups, including the ACLU and the NAACP, have signed a statement opposing pretrial risk assessment tools. The Leadership Conference on Civil and Human Rights, *supra* note 9. Scholars have likewise written open letters detailing the harms that these tools perpetuate and encouraging jurisdictions to consider alternative pretrial reforms. Barabas, Dinakar, and Doyle, *supra* note 51.

[137] In fact, risk assessments may be the least biased and most accurate application of the doctrine.

[138] In practice there should be some local variation. Pretrial incarceration laws, judicial practices, and criminal laws all differ — not to mention differences in the population at large, including rates of criminal offending, socioeconomic circumstances, racial composition, and more.

[139] In the wake of ProPublica's *Machine Bias* report, Angwin et al., *supra* note 5. the distribu-

mathematical necessity given how preventive incarceration laws function and given what demographic groups tend to have criminal records. Using data from pretrial risk assessment algorithms, researchers have already found that disparate error rates will occur in any predictive system in which different racial groups had different baseline rates of being arrested.[140] It will occur without any personal bias or discriminatory intent on the part of people or machines that apply these laws.[141] Higher rates of prior arrest within a subset of the population will lead to higher rates of pretrial incarceration.[142] Predictions of dangerousness will have more false positives for the group that has more positive outcomes and will have more false negatives for the group that has more negative outcomes.[143] In practice, this means that the distribution of errors for Black defendants will be different than the distribution of errors for white defendants. When mistakes are made with dangerousness predictions, Black peoples' dangerousness would tend to be overestimated and white peoples' dangerousness would tend to be underestimated.[144]

When applying pretrial incarceration laws, these mistaken predictions are common because predictions of future violence are highly speculative. Largely because pretrial violence is so rare, it is virtually impossible for judges or any statistical model to identify people who are more likely than not to commit a violent crime. Predicting the future is always difficult but is nearly impossible when predicting very rare, interpersonal events, like violence, within a short timeframe, like the pretrial period. Current pretrial risk assessments data can again be a useful demonstration.[145] With the leading tool on the market, the Public Safety Assessment, every defendant is flagged or not flagged for "new violent criminal activity." Within the tool's training data, of those who were flagged for "new violent criminal activity," only 7% went on to commit a violent crime on pretrial release.[146] In other words, the "new violent criminal activity" flag can be

tion of outcomes for pretrial incarceration decisions was thoroughly examined in the literature. See Sandra G. Mayson, *supra* note 42 at 2333–38 (collecting sources and explaining the phenomenon).

140 Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Berk et al., *supra* note 45. Absent an intervention that explicitly adjusted predictions based on race. Kleinberg, Mullainathan, and Raghavan, *supra* note 8.

141 *Id.*

142 To put it another way, groups of people that have more members with criminal histories are more likely to be preventively incarcerated. *Id.*

143 In essence, to create a predictive algorithm with equal error rates for different groups, unless the algorithm may require explicitly considers and adjusts for group differences.

144 Angwin et al., *supra* note 5.

145 As a reminder, because these tools outperform humans at predicting, their limits represent the limits of our ability to make violence predictions. *See supra* Part II. B–C.

146 14B Measuring and Managing Pretrial Risk with the Public Safety Assessment: Assessor Training.

expected to correctly identify who will commit violence about 7% of the time and can be expected to incorrectly identify who will commit violence about 93% of the time.[147] Risk assessments make errors when predicting who will not commit violence, too. Risk assessments do not flag the overwhelming majority of people who go on to commit violence on pretrial release.[148]

Algorithmic exposure of the range of possible outcomes of pretrial laws could prompt advocacy groups to reconcile their policy stances on risk assessment tools and pretrial incarceration doctrine. If values of racial justice have led these organizations to critique

---

[147] This pattern has proven to be fairly consistent in places where the PSA has been adopted, despite regional variations in pretrial incarceration rates, policing practices, crime rates, and more. Over different time periods, the percent of people flagged for NVCA who go on to be arrested for a violent crime on pretrial release has been 14% in New Jersey, 3% in Kentucky, and only 1% in Cook County, Illinois, which includes Chicago. Ethan Corey, *How a Tool to Help Judges May Be Leading Them Astray*, The Appeal (2019), https://theappeal.org/how-a-tool-to-help-judges-may-be-leading-them-astray/ (last visited Jan 8, 2021). Other studies of risk assessments have found similarly low rates. Megan T. Stevenson & Sandra G. Mayson, *Pretrial Detention and the Value of Liberty*, 39 (collecting studies).

[148] There is little reason to expect the quality of these predictions to improve. Recent decades have seen only slight improvement in our ability to predict violence, among any populations. Although pretrial risk assessments have only recently come to prominence, these tools are an offshoot of a broader field of risk prediction in the social sciences, particularly psychology, which has developed a rich literature on the subject. A leading meta-analysis of risk assessment tools concludes that "the ceiling of predictive efficacy may have been reached with the available technology." Min Yang, Stephen C. P. Wong & Jeremy Coid, *The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools.*, 136 Psychol. Bull. 740–767 (2010). Rare, interpersonal events are among the hardest to predict, and historical background information about people can only tell us so much about what they will do in the future. Given these limits, the authors conclude that risk assessments, "should not be used as the sole or primary means for clinical or criminal justice decision making that is contingent on a high level of predictive accuracy, such as preventive detention." Statistical models can modestly outperform unaided human prediction for predicting arrest and arrest for a violent offense. A meta-analysis of over half a century of psychology research concludes that "statistical prediction methods are, in general, more accurate than clinical prediction methods." (Ægisdóttir et al., 2006). But statistical methods are no crystal ball. This same meta-analysis acknowledges that these tools are, in general, only slightly more accurate than humans. For pretrial risk assessments, a recent empirical paper arrives at the opposite conclusion. Julia Dressel & Hany Farid, *The accuracy, fairness, and limits of predicting recidivism*, 4 SCI. ADV. eaao5580 (2018). Within a study framework that gave people specific factors to consider and provided feedback on the accuracy of their predictions, human subjects recruited on the internet were able to slightly outperform the COMPAS pretrial risk assessment tool. Some have criticized the study's framework for being too divorced from real-world circumstances, given that judges have more information in front of them and rarely know whether their predictions were correct or not. Goel et al., supra note 12. Perhaps the problem is not so much with the study design but with a legal system that doesn't provide judges feedback on their predictions. But whether this criticism is valid or not, the point remains that statistical tools are, at best, a modest improvement over human prediction.

and oppose risk assessments, these same values ought to the lead them to criticize pretrial incarceration laws that these algorithms apply.[149] Algorithmic exposure may also help advocacy groups evaluate pretrial incarceration laws' impact on other marginalized groups. To date, Black and white populations have been the primary subjects of empirical research on this question. Algorithmic exposure could reveal more details about how other population subgroups are treated under the law. So long as the dataset being used is labeled for characteristics of the population — age, race, gender identity, ethnicity, and so on — this analysis is possible.

Reconciling this policy position should also help to resolve internecine conflicts among advocates and researchers with shared goals. Too often, progressive advocates and sympathetic researchers are pitted on opposite sides of the algorithmic debate. Pretrial risk assessments are a case in point. Advocates oppose algorithmic risk assessments for the harm they reveal. Researchers support algorithmic risk assessments because they're superior to human prediction. Neither side gets it quite right.[150] The choice between harmful algorithms or worse humans is a false dilemma. If an algorithm seems to be causing harm but recalibrating the algorithm cannot fix that harm, then it is likely that the algorithm is not causing this harm but is revealing harm inherent to the law it applies. When a legal doctrine is the source of problems, the concern should be transforming the law — not choosing whether algorithms or humans should apply that law.[151]

In this way, algorithmic exposure can help to expand the scope of evidence-based or data-driven law reform. Technology and data are potent tools for political advocacy, particularly in a time of data-driven public policy. In the domain of criminal law, evidence-based criminal law reforms reign supreme. But scholars have noticed a troubling contradiction.[152] The shift to evidence-based policy standards now requires that any changes to the criminal law be justified through an empirical or data-driven lens. But

[149] As a policy matter, heightened procedural protections still matter, but they can't bring pretrial incarceration doctrine in line with these groups' broader values of racial justice.

[150] As explored in detail in Part I, when humans are making biased and error-prone predictions, algorithms are an improvement — all other things being equal. But that phrase, "all things being equal," does a lot more work than supporters of algorithms let on. In politics and law, things are rarely in stasis — and all things never stay equal in the zero-sum game of limited political will, public engagement, and financial resources. Algorithmic reforms to law necessarily come at the expense of other legal reforms. For those who desire fundamental changes to our criminal legal systems, it's rational to consider algorithms a distraction that siphons away attention and resources from other opportunities. The conflict within the algorithmic fairness discourse between proponents and critics of algorithmic interventions ought to be understand against this political economic backdrop.

[151] To put it another way: When prediction can justify a legal process, algorithms are preferable to unaided human judgment. But algorithms can also reveal when prediction does not justify a legal process.

[152] Erin Collins, Shifting the Evidence-Based Paradigm (forthcoming) (on file with the author).

existing criminal laws are not evidence-based. In fact, most of the laws that have driven mass incarceration in recent decades were adopted despite empirical evidence to the contrary.[153] Algorithmic exposure can provide a way to expand the scope of the "evidence-based" paradigm beyond tweaks or optimizations of existing practices to reckon with fundamental values.

Educating and Galvanizing the Public. These empirical insights aren't just useful for an organization's internal decision-making; advocacy groups can use this same information to educate and galvanize the public to pressure elected officials to change the law. Insights concerning accuracy and racialized distribution of outcomes can help justify the argument that pretrial incarceration based exclusively on predictions of violence is incompatible with either a commitment to racial justice or a commitment to due process of law.[154] Activists' similar arguments against algorithmic risk assessment tools have already proven politically effective.[155]

Insights into the accuracy and racial distribution of dangerousness predictions can be a powerful public education tool. The predictive turn in the criminal law has become deeply rooted within American culture. There is a sense that violence is predictable and that judges should be able to discern who should be incarcerated and who should be released pretrial. Particularly when machine learning algorithms are used, the public may overestimate the potential accuracy of these predictions. Although the phrase "preventive detention" conjures up ideas of rationality and science, the data collected here can expose the public to the fallout of these practices. The actual numbers are humbling, if not shocking.[156] Preventive incarceration that depends upon these predictions defies one of the central maxims in our legal tradition: William Blackstone's

---

[153] Michael Tonry, *Predictions of Dangerousness in Sentencing: Déjà Vu All Over Again*, 48 Crime Justice 439–482 (2019).

[154] These insights invite other considerations as well. How much is pretrial incarceration justified by factors beyond someone's control? Given that age is the most predictive factor for violence, how does pretrial incarceration conflict with other values [insert commitment to rehabilitation / opportunities for kids]. How much does prediction create a feedback loop of prediction and punishment rather than open up opportunities for positive change?

[155] Peter Krouse, *Ohio Supreme Court proposes bail reforms that don't include risk assessments*, cleveland.com (2020), https://www.cleveland.com/news/2020/01/ohio-supreme-court-proposes-bail-reforms-that-dont-include-risk-assessments.html (last visited Aug 14, 2020); Final Report of the Special Commission to Evaluate Policies and Procedures Related to the Current Bail System, (2019), https://d279m997dpfwgl.cloudfront.net/wp/2020/01/0102_bail-reform-report.pdf (last visited Jan 29, 2020).

[156] This information may not be so new or shocking to impacted communities. In this way, the data can reinforce or corroborate community knowledge. There is a risk in using empirical data in this way: the risk that a sufficient condition could become a necessary one. Many groups, including bail funds, have argued that we shouldn't need empirical data to prove what we already know through the lived experiences of communities that are policed and punished.

“[B]etter that ten guilty persons escape, than that one innocent suffer.”[157] Pretrial incarceration based on these predictions generates — in some cases precisely — the opposite ratio: ten people must be incarcerated to prevent one from committing a violent crime on pretrial release. Recent empirical research indicates that the public may not approve of pretrial incarceration at this poor a level of accuracy.[158] The question of the legitimacy of pretrial incarceration has sometimes been framed as, “What level of accuracy for predictions of future violence is needed to justifiably incarcerate someone who has not been found guilty of a crime?” One potential answer is that whatever that level may be, it’s not in the ballpark of what’s achievable.

Accuracy and racial inequity are not separate issues but are deeply interconnected. Incarcerating people upon weak justifications may only be possible when the people being targeted are already marginalized. As Caleb Foote remarked decades ago, preventive detention’s function is to be “an acceptable public-relations rationale” for “conceal[ing] and perpetuat[ing] a discriminatory system of justice in the face of growing unwillingness from those difficult lower classes to continue to be treated more as objects than as humans.”[159] Prediction has been a way to sweep discriminatory treatment and replication of systemic inequality under the rug. As descriptive machines that reproduce what they’ve been fed, algorithms can lay bare the structural inequities have been reproduced within seemingly objective predictions.[160]

Legislation

Algorithmic exposure of law can allow for a more informed, nuanced approach to legislation and rulemaking. Society does not need to blindly construct predictive laws and discover the consequences later. Legislatures and administrative agencies could subject predictive laws and regulations to algorithmic analysis to ensure that the laws are able to achieve their goals and are producing fair outcomes. The process of writing these laws could aided by predictive algorithms from start to finish.[161]

The legislative processes explored in this section illustrates how algorithmic exposure can enrich the public and legislative debate over predictive laws by demystifying pre-

---

157 Alexander Volokh, *n Guilty Men*, 174 Univ. Pa. Law Rev. (1997).

158 Stevenson and Mayson, *supra* note 139.

159 Foote, *supra* note 124.

160 For a discussion of the tension between empirical knowledge gained through algorithms and community knowledge gained through lived experience, see infra Part IV B. Knowledge and Values.

161 Insights gained through the algorithmic exposure of law do not then require that the predictive law be administered by algorithms. There’s no iron rule that software must be integrated into the legal process. A law could be made simpler for the sake of honest and transparency while maintaining predictive accuracy.

diction and prevention, particularly within the domain of criminal law. By breaking a prediction down to its constitutive elements, algorithms can reveal the contingent, political nature of a predictive law. Predictive laws have often been characterized as an objective forecast coupled to a neutral cost-benefit analysis,[162] But a predictive-legal process cannot be objective because defining the system to be analyzed and the variables to be considered is always a normative, political choice — not a purely quantitative one.[163] As Ruha Benjamin reminds us, "even just deciding *what problem* needs solving requires a host of judgments."[164] Questions of scope, factors worth considering, and even objectives worth prioritizing are not self-evident.[165] Algorithmic exposure lays bare these considerations by showing what factors and outcomes of interest were or could have been considered in the predictive process.

Legislatures can incorporate algorithmic exposure into the legislative process as a kind of impact assessment for predictive laws. Impact assessments are nothing new. The legislative processes for state and federal governments often require impact assessments, including those that measure the environmental,[166] economic,[167] or racial[168] impact of proposed regulations and laws. A predictive impact assessment could report to the legislature a predictive law's accuracy, its distribution of outcomes and errors across different population groups, and the relevancy and strength of the predictive factors included in the law. Assessments could be performed on competing versions of proposed legislation so that legislators and the public could compare their respective performance.

To show how predictive impact statements could be useful — and not just another bureaucratic report among many produced in a legislative session — let's consider what an impact report might return about pretrial incarceration law's predictive factors and how a legislature might respond. Current pretrial incarceration statutes include many factors, few of which have been vetted to see if they correlate with pretrial violence or recidivism.[169] The report may find that some factors required by law do not have pre-

---

[162] Harcourt, *supra* note 20.

[163] *See Id.* Bernard Harcourt identifies this objectivity trap as the systems fallacy and traces its origins to the development of operations research during World War II and the rise of systems analysis in the mid-twentieth century.

[164] Ruha Benjamin, Race after technology: abolitionist tools for the new Jim code 11 (2019). Insert reference to point brought up in Ben Green's LPE blog post.

[165] Harcourt, *supra* note 20.

[166] Biosafety Unit, *What is Impact Assessment?* (2010), https://www.cbd.int/impact/whatis.shtml (last visited Aug 11, 2021).

[167] Economic Impact Statement Act - American Legislative Exchange Council, , https://www.alec.org/model-policy/economic-impact-statement-act/ (last visited Oct 21, 2021).

[168] Racial Impact Statements, , The Sentencing Project , https://www.sentencingproject.org/publications/racial-impact-statements/ (last visited Oct 21, 2021).

[169] A slightly more complicated issue is that there can be some factors that have only slight

dictive value.[170] An impact report might also find that some factors have an outsized influence. Even if a pretrial incarceration statute lists four factors, it's unlikely that each of the four factors is responsible for 25% of a prediction of future violence. Some factors will be stronger indicators of violence, and some will be weaker. Knowing which factors are most important may shape public and policymaker attitudes toward prediction, because these factors may compete with other values.

For example, policymakers may reconsider how to structure preventive incarceration legislation if a predictive impact report reveals that the law instructs judges to lock people up mostly because they are young. From a purely predictive standpoint, age is a helpful factor to include in a preventive incarceration law, as age is often the strongest factor when predicting recidivism risk.[171] In other words, compared to all other information, age is the best indication of whether someone will be rearrested for a violent crime on pretrial release. Although age statistically correlates with crime, other values might override the legislature's goal of maximizing the accuracy of preventive incarceration. Perhaps legislators want the law to reflect the importance of giving young people second chances or the overall benefit of prioritizing rehabilitation over incarceration given the plasticity of juvenile brain development.[172] If that were the case, a legislature may choose to modify a pretrial incarceration law that only considers recidivism risk to make age-related exceptions or to include alternative procedures for young people. The benefit of the predictive impact assessment is that it can alert the legislature to how a seemingly neutral law conflicts with other values and goals.

---

predictive value. These factors do not dominate any statistical model, but they are not completely irrelevant on their own. They bear some statistical relationship to arrest for a violent crime. At the level of building a statistical model, the problem with these factors is that they tend to drag in more noise than signal when included in the model. That is to say: by including these factors in the model, the model becomes less accurate, even though those factors are predictive. The problem is that their relationship with the outcome isn't clean enough. We could make better predictions by considering fewer factors that have a stronger relationship with the outcome.

[170] This concern is not hypothetical. The factors used within many predictive laws were not derived from a rigorous process. A common way of drafting these laws is BOGSAT: a bunch of guys sitting around a table.

One example of this problem is bail statutes. Following the first set of bail setting formulas pioneered by the Vera Institute in the 1960's, many states adopted laws that listed factors that judges ought to consider when setting bail, under the impression that these factors were predictive of future court appearance. The process of building pretrial risk assessment algorithms has already shown that many of these factors have a weak or non-existent relationship with future court appearance.

[171] *See* Stevenson and Slobogin, *supra* note 114 at 1–2.

[172] Megan Stevenson and Christopher Slobogin have more deeply explored the question of risk assessment policy on this issue. Megan T Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 26.

Algorithmic exposure could also be incorporated iteratively into the legislative drafting process. Algorithms could be used to experiment with draft legislation and test a variety of models to see which best adhere to legislative goals and priorities.[173] In this way, algorithmic exposure could be used not just to condone or condemn a law but to explore how it might be reshaped. Formulas for prediction could be tested and modified to capture a subset of models that perform more efficiently or better reflect moral or legal values.

Among other possible uses, this iterative process can be a way for legislatures to work through difficult decisions about predictive laws' dependence upon biased data.[174] The public and public officials today are concerned with biased data distorting predictions of risk. Whether a prediction of pretrial dangerousness is made by a judge or an algorithm, these predictions use data that reflect patterns of policing and prosecution. Decades of research have shown that, for the same conduct, Black and Hispanic people are more likely to be arrested, prosecuted, convicted, and sentenced to harsher punishments than their white counterparts.[175] As the classic computer science aphorism goes: Garbage in, garbage out.[176] Predictions made using biased data will produce biased results.

So what is a legislature to do about biased predictions? One proposal is to get rid of predictive incarceration laws. Another proposal is to explicitly consider race when mak-

---

[173] In some circumstances, changes to how predictions are made could mitigate many distributional harms. If the errors in prediction fall upon certain groups rather than others, the law may violate principles of distributional fairness. Algorithmic exposure can reveal not just a particular instantiation of a legal doctrine or theory but the whole range of possibilities. If a particular range of outcomes are impermissible, an algorithm can reveal which applications of a legal doctrine should be foreclosed to prevent these outcomes.

[174] Gone are the days when advocates and public officials proclaimed legal algorithms' objectivity and neutrality. Concerns over biased data have gone mainstream. Biased data is the algorithmic fairness concern that has found its way from academia to newspapers, documentary films, and presidential campaigns.

[175] See generally The Sentencing Project, Report of the Sentencing Project to the United Nations Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia, and Related Intolerance Regarding Racial Disparities in the United States Criminal Justice System (2018); Lynn Langton & Matthew Durose, U.S. Dep't of Justice, Police Behavior During Traffic and Street Stops, 2011 (2013); Stephen Demuth & Darrell Steffensmeier, The Impact of Gender and Race-Ethnicity in the Pretrial Release Process, 51 Soc. Probs. 222 (2004); Jessica Eaglin & Danyelle Solomon, Brennan Center for Justice, Reducing Racial and Ethnic Disparities in Jails: Recommendations for Local Practice (2015); Sonja B. Starr & M. Marit Rehavi, Racial Disparity in Federal Criminal Sentences, J. Pol. Econ. 1320 (2014); Marc Mauer, Justice for All? Challenging Racial Disparities in the Criminal Justice System (2010).

[176] Or as Sandra Mayson recasts it, Bias In, Bias Out. Sandra G. Mayson, *supra* note 42.

ing recidivism predictions, as a kind of affirmative action for crime prediction.[177] Neither option seems politically palatable.

Algorithmic exposure can provide an opportunity to test other ways of reducing the effect of biased data over predictive decision-making.[178] One idea worth testing is that biased data plays an outsized role at the margins of pretrial decision-making. The logic is this: Judges and algorithms are much more likely to be affected by bias and make the wrong predictions for borderline cases. A Black person with a few arrests on their record may only have those arrests because of policing practices in their neighborhood. A similarly risky white person would have a clean record, because they don't live in a heavily policed area. There would be a difference in the data that does not reflect a difference in reality. Relying on this data, judges and algorithms would overestimate the risk of the Black person and underestimate the risk of the white person. In contrast, when a person has a lengthy history of violent crimes, biased data should have less of an impact. That data can meaningfully indicate predilection toward violence, despite any noise from biased policing or prosecution practices.[179]

Under this conception of biased data's impact, the impact of racial bias could be reduced if the law required greater certainty of future violence before an order of pretrial incarceration can be imposed. Absent a method for testing the theory, a legislature — and the public — would have to accept or reject it at face value as a credible explanation of biased data's impact. If accepted, the legislature would also have to guess what level of predictive certainty would best reduce these disparities.

In contrast, algorithmic exposure provides an opportunity to prospectively explore a range of potential variations on a legal rule and evaluate the results. The following graph shows the results of a simple experiment that explore how racial disparities in error rates can change based on how certain a prediction of recidivism must be before it can justify a decision to label someone as "high risk." For this test, 100 different pretrial risk assessment models were constructed and tested on pretrial data.[180] The only variation between the models is the threshold at which they labeled someone as high risk. On the left side of the graph, everyone was labeled high risk. Moving from left to right, the models become more stringent on who deserves to be labeled high risk. On

---

[177] Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Deborah Hellman, *Measuring Algorithmic Fairness*, 106 Va. Law Rev. 811 (2020).

[178] This approach would (at least implicitly) adopt a counterargument to the claim that predictive laws should be abolished because of biased data is that the data coming in and the predictions coming out are not all garbage. This is a normative claim more than a technical one. The normative question that is left open is, What level of disparate treatment of marginalized groups is acceptable within predictive laws?

[179] In other words, bias in the data can distort some predictions but it does not scrub all predictions of all value.

[180] The data included both case outcomes and a risk assessment tool's predictive performance and was used in ProPublica's report on risk assessment bias. Angwin et al., *supra* note 5. The risk assessment model was built to mirror the COMPAS risk assessment tool.

the right side, no one was labeled high risk. The y-axis measures the false positive rate — the frequency with which the model mistakenly labeled someone as high risk.[181] The unbroken green line tracks the false positive rate as a function of threshold for Black defendants, and the dotted blue line tracks the false positive rate as a function of threshold for white defendants.

The difference between the two lines is the difference in the rate at which Black people and white people were mistakenly labeled high risk. The lines converge at the points where everyone is labeled "high risk" and where everyone is labeled "not high risk," but they diverge in varying amounts along the path between these two points. At all points, Black people were more frequently mistakenly labeled "high risk" than white people. But the difference narrows considerably across certain ranges.

What this graph reveals is that — at least for the sample data — racial disparities in error rates for predicting recidivism can be reduced by changing how certain a prediction of recidivism must be before someone is incarcerated. Choosing the appropriate level of certainty is a value-laden question dependent upon potentially competing ideas of racial equity and acceptable error rates.[182] Algorithmic exposure can't answer that normative question, but it can make the stakes clearer. Any legislature contemplating changes to its predictive incarceration laws ought to have this kind of information available rather than draft pretrial incarceration laws without knowing their possible effects.

Policy

Professional researchers within state or federal administrative agencies are already capable of using algorithms as a diagnostic tool for law. These are institutions of expert authority that have access to government data and already use statistical methods for different analyses.[183] In the past decade, a growing scholarly literature in criminal law has explored the opportunity for these agencies to usher in and oversee a new era of empirical, expert-driven criminal law reform.[184] State administrative agencies, administrative offices of the court, and executive offices are primed to deploy algorithmic exposure as a method.

[181] For purposes here, the definition of a mistaken label of high risk was accepted as-is from the dataset, and a person was not considered high risk if they were not arrested within a two-year window following the assessment.

[182] Reductions in racial disparities in this example will come at the cost of overall accuracy.

[183] See generally David Freeman Engstrom et al., Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies, SSRN Electron. J. (2020), https://www.ssrn.com/abstract=3551505 (last visited Aug 10, 2021).

[184] Rachel E. Barkow, Prisoners of politics: breaking the cycle of mass incarceration 15 (2019).

In the context of pretrial incarceration doctrine, prosecutors are also capable of using algorithmic exposure to reshape their offices' pretrial policies. As a group, the new wave of "progressive prosecutors" seem more eager than traditional prosecutors to hire data scientists and other technical staff to collect data on prosecutorial behavior and conduct internal data audits.[185] Larry Krasner's office is collaborating with the University of Pennsylvania to create a public-facing data dashboard to share and analyze prosecutor decisions.[186] Smaller offices that don't command a national spotlight are also taking a data-driven approach to revising internal policy. In 2021, the District Attorney's office for Yolo County, California (located just outside Sacramento) created a transparency data portal for the public to track how the office handles cases.[187] This effort has already led to policy change. Data revealed that the District Attorney's office was disproportionately denying people of color the opportunity for their cases to be diverted out of the criminal legal system before trial, largely because having a criminal history could automatically disqualify someone from diversion. The office has eliminated automatic disqualification, and the data portal will continue to track the results.[188] In offices large and small, algorithmic exposure could slip in alongside the data work already being conducted.[189]

Although prosecutor offices are capable of using algorithmic exposure to reevaluate pretrial policy, many offices might be resistant to scrutinizing pretrial incarceration practices in this way. Even among progressive prosecutors who have promised a dramatic change in the government's approach to pretrial incarceration, their approach has been a gentler version of extant pretrial incarceration doctrine, rather than an alternative paradigm. Nonetheless, these offices have been flexible in reevaluating longstanding prosecutorial practice and present the best opportunity for algorithmic exposure to reshape prosecutors' pretrial decision-making. These offices have been

---

185 Rachael Rollins's office in Boston and Larry Krasner's office in Philadelphia have fulltime staff studying the office's internal data.

186 Philadelphia DAO, *District Attorney Krasner Announces DAO Transparency & Accountability Research Collaboration*, The Justice Wire (2020), https://medium.com/philadelphia-justice/district-attorney-krasner-announces-dao-transparency-accountability-research-collaboration-27b12ad833db (last visited Aug 9, 2021).

187 Press Release, Commons Policy Changes, , https://yoloda.org/commons-policy-changes/ (last visited Jun 21, 2021).

188 *Id.*

189 Outside non-profits or other organizations are already positioned to assist prosecutors in these efforts. Collaborations between DA offices sharing data and non-profits analyzing the data can often blur the "inside-outside" line. The Vera Institute's Reshaping Prosecution program is a prime example of an inside-outside venture. Reshaping Prosecution, , Vera Institute of Justice , https://www.vera.org/projects/reshaping-prosecution-program (last visited Jun 21, 2021). Their self-defined goal is to "Help prosecutors implement data-driven policies and practices that reduce incarceration and promote racial equity and equal justice." *Id.* Organizations like these already have the data and staff to conduct this kind of analysis.

amenable to critiques of existing law and practice, and they are experimenting with approaches that diverge from those of their tough-on-crime predecessors.

To date, it has been easy for prosecutors, judges, and the public to overestimate the accuracy of pretrial dangerousness determinations. Prediction seems to point to objective facts in the world: Without some intervention, people will commit crimes, abuse children, defraud the government, and attempt terrorist acts. To prevent this future from happening, government action may be needed. But the predictions that justify this action are hard to parse. The human predictive process is opaque and therefore resistant to analysis. And the outcomes always justify themselves, particularly carceral outcomes.[190] A judge is always proven right for locking someone up. Because the person was incarcerated, they couldn't harm the public. And if a judge lets someone go and that person harms others, that's reason for more government intervention next time.[191]

Algorithmic exposure can lay bare how inaccurate predictions of pretrial violence are. Preventively incarcerating people for general predictions of violence requires primarily incapacitating people who do not need to be incapacitated. An algorithmic pretrial tool optimized to produce the most accurate results would not flag for anyone for violence. For any person in a dataset, the most likely outcome is that they will not commit violence on pretrial release. Therefore, to produce predictions of violence and encourage incarceration, accuracy must be sacrificed. This results in substantially more false positives — people who are flagged for violence but do not go on to commit a violent crime — than true positives — people who are flagged for violence and do go on to be arrested for a violent crime. The necessary consequence is jailing the many people labeled "high risk" to prevent the violence of a few of their members.

Accordingly, the effect of algorithmic exposure on prosecutors' pretrial policies ought to be a measure of modesty in dangerousness determinations and pretrial incarceration decisions.[192] Prosecutor offices could narrow the circumstances in which their attorneys are allowed to argue for pretrial incarceration. Permissible incarceration decisions

---

[190] This point was recognized at the dawn of the predictive-carceral era. *See* Tribe, *supra* note 124.

[191] This phenomenon has been particularly prevalent within criminal law but has metastasized through American legal systems to domains including civil commitment, child protective services, government benefits fraud, and counter-terrorism policies, among others. *See* Bernard E. Harcourt, Against prediction: profiling, policing, and punishing in an actuarial age (2007).

[192] Algorithmic exposure is a mechanistic process, open to various normative applications. Although this Article explores how algorithms could be used to critique criminal law from a progressive, decarceral perspective, the insights that algorithms reveal about law are not inherently normative, and the reconception of algorithms as a tool for reevaluating law is not bound to a particular political perspective. To put it another way, algorithmic exposure is amenable to varying definition of fairness. Fairness is a capacious, contested idea. Hellman, *supra* note 169 at 814. Because the central claim of this article is a mechanistic one, it can accommodate many different conceptions of predictive fairness.

would have to be justified by reasons beyond pure prediction, such as strength of the evidence, the gravity of the present offense, and concerns of the alleged victim(s).

With algorithmic insights in hand, a prosecutor's office would be uniquely positioned to challenge and change pretrial decision-making.[193] In some places, prosecutors have enough power over the pretrial process that a prosecutor office's policy limiting pretrial incarceration would become the *de facto* practice of the court system. Prosecutors' authority over pretrial incarceration can vary by jurisdiction. In some places, prosecutors are the gatekeeper for pretrial incarceration: A judge can order a person to be incarcerated pretrial only upon a prosecutor's motion.[194] But this is the exception. In most places, the judges are in charge. Whether by imposing an unaffordable money bond amount or directly ordering someone to be incarcerated pretrial, judges can either respond to a prosecutor's motion for pretrial incarceration or choose to jail people on their own initiative. Even in places where prosecutors cannot singlehandedly limit pretrial incarceration, arguments from the prosecution can influence judges — both with individual decisions and with systemic approaches to pretrial decision-making.[195]

Insights from an algorithmic diagnosis of law could muscle their way into frontline attorneys' arguments in court. State laws afford prosecutors broad discretion over the pretrial incarceration arguments they can make in court. Pretrial incarceration laws tend to be permissive, not mandatory.[196] They don't require the state to incarcerate anyone. They instead permit the incarceration of people who pose a unique danger to the community. Under these laws, prosecutors could argue that the conditions the law requires are rarely met. The law may allow the incarceration of people who pose a serious risk of causing "serious bodily harm" to others, but this a high threshold.[197] Prosecutors could argue that the law requires more than a weak generalized prediction of someone's future behavior to justify pretrial incarceration.

---

[193] A prosecutor who concludes that pretrial incarceration on general dangerousness cannot coexist with a commitment to racial justice will need to adopt a new paradigm for pretrial policy. The good news is that this is exactly what they were elected to do. But the academy, civil society and policy think tanks need to seize this moment and think beyond technocratic adjustments to our pretrial laws and policy. At present, the policy options tend to be pretrial incarceration on general dangerousness or bust. There are no off-the-shelf policy alternatives to pretrial incarceration on general dangerousness other than abolition or issuing a moratorium on pretrial incarceration. Both of these are political nonstarters, even in progressive cities like San Francisco or Boston.

[194] *E.g.*, N.J. Stat. Ann. § 2a:162-16, 18 (West 2017).

[195] As I've argued with co-authors before, court culture is understudied, underappreciated element of criminal law reform, particularly in circumstances in which judges have broad discretion, like pretrial hearings. *See generally* Mitali Nagrecha, Sharon Brett & Colin Doyle, *Court Culture and Criminal Law Reform*, 69 Duke Law J. Online 84 (2020).

[196] *E.g.*, Mass. Gen. Laws Ann. ch. 276, § 58A (West 2021).

[197] Cal. Const. art. I, § 12.

In an analogous pattern, the Philadelphia District Attorney's office now requires assistant district attorneys to calculate the projected costs to the state for incarcerating someone and to share that information with the court at sentencing.[198] If the assistant district attorney believes the costs of incarceration are justified, the attorney must make a cost-benefit argument at the sentencing hearing. A similar policy could be adopted in pretrial hearings. For every person that the court is considering incarcerating pretrial, frontline attorneys would have to present to the court, first, the statistical likelihood of that person committing a violent crime if released and, second, statistics of the disparate burdens that pretrial incarceration places upon local communities of color. Prosecutor recommendations for pretrial incarceration or judicial orders of pretrial incarceration would have to be made in light or in spite of this data.

Outside of court, prosecutors could use insights from algorithmic analysis of pretrial doctrine to reframe the public debate over crime and preventive incarceration. The prosecutor plays an important role as spokesperson for the government on all matters of criminal law. Prosecutors tend to be in the limelight when high-profile cases attract media attention. But prosecutors are regularly asked to comment on other aspects of our criminal legal systems. In public disagreements with judges or police, prosecutors could use algorithmic analysis of their own policies and compare them to how courts deviate from their recommendations to show what would have resulted if a prosecutorial understanding of fairness had been adopted. Data from algorithms critiquing law can be a countermeasure to the newspaper headlines that inevitably appear when someone on pretrial release commits a notorious crime. Predicting violence will always seem easier than it is, particularly in hindsight. Empirical insights, particularly insights drawn from data about the jurisdiction itself, can counter these tendencies and bring a clear-eyed realism to emotionally charged — but statistically exceptional — circumstances.

Concerns & Opportunities

So far, this Article has developed the concept of algorithmic exposure of law and has demonstrated the concept's practical utility. This Part looks ahead, noting the conditions that might prevent algorithmic exposure from being adopted or might limit its effectiveness. Algorithms can be a powerful tool for revealing a law's potential, but the method cannot answer all questions and can be susceptible to misuse. This part examines some of the prospects and pitfalls that lay ahead.[199]

---

198 Chris Palmer, *In latest edict, Philly DA Larry Krasner tells prosecutors to seek lighter sentences, estimate costs of incarceration*, The Philadelphia Inquirer, https://www.inquirer.com/philly/news/crime/philadelphia-district-attorney-larry-krasner-plea-deals-shorter-sentences-cost-of-mass-incarceration-20180315.html (last visited Aug 10, 2021).

Biased Data

Criminal law data is biased. And algorithms that are trained on this data will replicate those biases. Garbage in, garbage out: Any machine learning model is only as good as the data fed into it. If it is fed biased data, it will produce biased results. Algorithmic tools currently used in criminal law rely on historical records of arrests, charges, convictions, and sentences to generate predictions about a person's future behavior. These tools implicitly assume that criminal history data are a reliable and neutral measure of underlying criminal activity. But these records reflect not just people's activity but also the activity of courts and police.

Court and police data is dirty data. People of color are treated more harshly than similarly situated white people at each stage of the legal system, which results in serious distortions in the data used to develop legal algorithms.[200] Take arrest records as an example. Arrest records are both under- and over-inclusive of the true crime rate. Arrest records are under-inclusive because they only chart law enforcement activity, and many crimes do not result in arrest.[201] Less than half of all reported violent crimes result in an arrest, and less than a quarter of reported property crimes result in an arrest. Arrest records are also over-inclusive because people are wrongly arrested and arrested for minor violations, including those that cannot result in jail time. For decades, communities of color have been arrested at higher rates than their white counterparts, even for crimes that these racial groups engage in at comparable rates.[202]

---

[200] See generally The Sentencing Project, Report of the Sentencing Project to the United Nations Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia, and Related Intolerance Regarding Racial Disparities in the United States Criminal Justice System (2018); Lynn Langton & Matthew Durose, U.S. Dep't of Justice, Police Behavior During Traffic and Street Stops, 2011 (2013); Stephen Demuth & Darrell Steffensmeier, The Impact of Gender and Race-Ethnicity in the Pretrial Release Process, 51 Soc. Probs. 222 (2004); Jessica Eaglin & Danyelle Solomon, Brennan Center for Justice, Reducing Racial and Ethnic Disparities in Jails: Recommendations for Local Practice (2015); Sonja B. Starr & M. Marit Rehavi, Racial Disparity in Federal Criminal Sentences, J. Pol. Econ. 1320 (2014); Marc Mauer, Justice for All? Challenging Racial Disparities in the Criminal Justice System (2010).

[201] FBI, 2017 Crime in the United States: Clearances, https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topicpages/clearances (last visited June 28, 2019).

[202] Megan Stevenson & Sandra G. Mayson, The Scale of Misdemeanor Justice, 98 B.U. L. Rev. 731, 769-770 (2018). This comprehensive national review of misdemeanor arrest data has shown systemic and persistent racial disparities for most misdemeanor offences. The study shows that "black arrest rate is at least twice as high as the white arrest rate for disorderly conduct, drug possession, simple assault, theft, vagrancy, and vandalism." Id. at 759. This study shows that "many misdemeanor offenses criminalize activities that are not universally considered wrongful, and are often symptoms of poverty, mental illness, or addiction." Id. at 766. For example, Black people are 83% more likely to be arrested for marijuana compared to whites at age 22 and 235% more likely to be arrested at age 27, in spite of similar marijuana usage rates across racial groups. "[R]acial disparity in drug arrests between black and whites cannot be explained by race differences in the extent of drug offending, nor the nature of drug offending." Ojmarrh Mitchell & Michael S. Caudy, Examining Racial Dis-

Biased data is an intractable problem that raises ethical concerns for using court and police data to justify decisions to punish, incarcerate, or otherwise restrict people's liberty. But for exposing law with algorithms, the ethical concerns are not quite as clear. Biased data can present ethical concerns about using algorithms to analyze laws, but these concerns are dependent on what analysis is being done and toward what ends. There should be no doubt: the same fundamental problem of biased results persists. Algorithms that are used to study law that have been trained on biased data will produce biased results. These biases should be taken into account when making any claims based on information produced by these algorithms. But dirty data doesn't necessarily contaminate or discount an entire project.

Another aphorism worth remembering is that all models are wrong, but some are useful.[203] All models are imperfect representations of the world. Some useful claims can be made even when the data and results are biased. Consider this Article's proposal to use algorithms to understand the distribution of outcomes that necessarily result from a pretrial incarceration law. This algorithmic model would be trained upon biased criminal law data. And yet, the bias would not reduce the usefulness of the insight the model provides because the model would not be making a claim about the behavior of people accused of crimes. To make that claim would require interpreting biased data as though it were an unbiased report of ground truth. Instead, the algorithm reports on how a legal system would operate. The distribution of legal outcomes is surely distorted by biased data. But claims about the distribution are not.

Biased data introduces uncertainty into any model that depends on that data to represent truth. Biased data creates a rift between reality and the statistical model that is being built. Consider claims about accuracy in pretrial dangerousness predictions. Unlike the claim about distribution of outcomes, this claim is much more susceptible to being distorted by biased data. Any claims of predictive accuracy that are based on biased data will be biased, because the claims are based on accepting the data as a true representation of how people behave. Because the data that guides the model diverges from the real world, the model is likely to overestimate the potential accuracy of predictions of violence. Therefore, any claims about accuracy must be qualified whenever dirty data is an issue. The less reliable the underlying data set is, the less reliable the claims about accuracy can be. But this doesn't mean that a claim about accuracy can never be made if data is biased. All data and all models diverge from reality. Claims about accuracy must be qualified and informed by knowledge about the domain. In some circumstances biased data may so overwhelm a dataset that one can't reliably make certain claims. But this claim would not just be about the deficiencies of an algorithmic model

---

parities in Drug Arrests, Just. Q., Jan. 2013, at 22. Similarly, Black drivers are three times as likely as whites to be searched during routine traffic stops, even though police officers generally have a lower "hit rate" for contraband when they search drivers of color. Ending Racial Profiling in America: Hearing Before the Subcomm. on the Constitution, Civil Rights and Human Rights of the Comm. on the Judiciary, 112th Cong. 8 (2012) (statement of David A. Harris).

[203] Insert from chapter

that is trying to understand accuracy. This would be a claim about the ability to make these kinds of predictions at all. Under these circumstances, given that the information about the world is so unreliable, neither humans nor algorithms would have sufficient knowledge to make accurate predictions.

Knowledge and Values

Algorithmic exposure has the potential produce new empirical knowledge about that world gained through algorithmic processing of data. It's the type of knowledge that would be expected from this process: knowledge about the world acquired through observation.[204] Here, knowledge has been produced through observing the process and results of prediction. The process of constructing algorithms is a means by which one can view and understand how a predictive legal process operates. The claim of this Article is that algorithms can help to reveal information about legal prediction that was previously unknown.

At the same time, these algorithms may produce knowledge that is already known to some groups and not others based upon their social identities.[205] "Situated knowl-

---

[204] Within the article, existing empirical knowledge filled in some insights that algorithms would reveal about the law. If this works, why do we need the process of using algorithms to expose the law at all? Why not just rely upon social science research?

In most contexts, this kind of robust social science research doesn't conveniently exist to fill in the gaps. This case study was chosen, in part, because it's an area that has been subject to extensive study by social scientists, and this information overlaps with the information algorithms could reveal about pretrial law. Most areas of law will not have this kind empirical research to fall back on. Even when this research does exist, it's often less useful for local advocacy than insights into how local laws affect the local population. Empirical research tends to be about one particular jurisdiction, either at the state or county level. Advocates must make a series of inferential steps to connect research about Kentucky to a District Attorney campaign in San Francisco. These inferences require trust in the out-of-state study design and trust that research conducted in other places at other times applies to the current time and place.

Social science research and machine learning tools are complementary, not exclusive. In practice, machine learning is an engineering tool more than it is a tool of social science research. There are significant overlaps between what can be uncovered in this process and what can be uncovered by the empirical methods used in economics and other disciplines. Indeed, some academic research explicitly designs and tests machine learning models. Kleinberg et al., *supra* note 87. As it is generally deployed, machine learning lacks some of the rigor of social science empirical methods. Machine learning does not generate causal insights, and its findings are not held to the same exacting standards of empirical methods. Machine learning is a practical and flexible engineering tool deployed in business and government to decipher troves data and derive workable insights. Across many domains, the steady march of social science and the nimble dance of machine learning can work together.

[205] Rua M. Williams & Juan E. Gilbert, *Cyborg Perspectives on Computing Research Reform*, *in*

edge" is a term introduced by Prof. Donna Haraway "as a means of understanding that all knowledge comes from positional perspectives."[206] One's social position always determines what it is possible to know about a given object of study.[207] Too often what passes for "objectivity" is the view from the position of a white, male researcher. In criminal law, this position more closely aligns with people administering the criminal legal system than those affected by it. Some of the knowledge produced by using algorithms as a diagnostic tool for criminal law would be new information for those in that privileged social position. But much of that knowledge would not be new to people from communities who are heavily policed, prosecuted, and punished.

Arguments about the systemic inequities of criminal law are not new.[208] Part III's focus on pretrial incarceration is a good example. Community groups across the country, particularly community bail funds, have spent years protesting and advocating against contemporary pretrial incarceration laws and practices. A cursory look at their advocacy would reveal their knowledge that prosecutors — including the new flock of progressive prosecutors — are incarcerating primarily poor Black men on weak predictions of future wrongdoing. An algorithmic diagnosis of pretrial incarceration laws may confirm that knowledge, but it would not be revelatory to those groups in the way it might be revelatory to people from positions of privilege and remove from the criminal legal system.

Although much of the knowledge that algorithmic exposure produces may not be new to the communities most affected by the criminal legal system, these algorithmic insights can still support community-led efforts both as an empirical foundation and as a lightning rod for attention. These algorithmic insights can make advocates' arguments more difficult to dispute, more broadly appealing, and more likely to garner attention.

Quantifying and confirming grassroots knowledge can support grassroots efforts. It can validate the experience of members of those communities to a broader public and galvanize support for change. Empirical knowledge can affirm information that is already known by some groups of people and broaden the audience of people who will acknowledge that information. This empirical knowledge can be helpful by revealing how the law works step-by-step and pinpointing where to direct efforts at reform. Data-driven insights can make concrete much of what can be uncertain or speculative within arguments rooted in personal experience.[209] And the opposite is true: stories of

---

Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems 1–11, 1 (2019), https://dl.acm.org/doi/10.1145/3290607.3310421 (last visited Jul 7, 2021). "Situated knowledge" is a term introduced by Prof. Donna Haraway. *Id.*

206 *Id.* at 2.

207 *Id.* at 2.

208 *See e.g.*,Steinberg, *supra* note 128; Tribe, *supra* note 124; Hegreness, *supra* note 128; Baradaran, *supra* note 128; Thaler, *supra* note 128; Foote, *supra* note 124; Alschuler, *supra* note 128.

209 *See e.g.*, Tribe, *supra* note 124; Foote, *supra* note 124; Alschuler, *supra* note 128; Barada-

people's experiences with criminal legal systems can enrich the data that algorithms glean about pretrial laws by contextualizing this information and framing arguments within a moral reasoning.

Technology can be a lightning rod for attention. As Rediet Abebe has highlighted, technological insights can act "as synecdoche when [they make] long-standing social problems newly salient in the public eye."[210] For better or worse, empirical information has political cache. There are constituencies who will discount community voices and advocacy but who will listen to empirical data that confirms that community's experiences.

This terrain is fraught. The process of using algorithms to expose problems in criminal law ought be approached with humility, grace, and accountability to the people affected by these laws. Claims of "new" information should not be overstated when this information confirms what many have been experiencing, protesting, and contesting for years. Researchers and advocates must be careful in their work to use this information as a support for grassroots efforts and not use their privileged positions to supplant or ignore that situated knowledge. We do not need algorithms to discover that our criminal legal system is inequitable, but algorithmic insights have the potential to reinvigorate, expand, or reinforce existing lines of critique.

Algorithms for All

The biggest obstacle to exposing law with algorithms is access to clean, high-quality datasets.[211] Machine learning depends upon high quality labeled datasets. The quality of government data sets can vary significantly by place, particularly with state and local governments. In some places, court records are still mostly paper records and not digitized. Other places have advanced to using an entirely digitized process. Many court systems do not have common databases or do not collect data in different courts and offices in a way that can work together. In many places, data is spread across different institutions. A court system may use a software that differs from probation that differs from local jails that differs from prisons that differs municipal record keeping. Each of these institutions may produce data that is labeled differently or can't interact well with another institution's data or does not track important people and actions. Most of the data that legal systems collect has not been collected with later statistical analysis in mind. It has been collected simply as a matter of record keeping. Accordingly, the

ran, *supra* note 128; Thaler, *supra* note 128.

[210] Rediet Abebe, Designing Algorithms for Social Good 11–12 (2019). Depending on how progressive prosecutors respond to these attempts at accountability, algorithmic insights can also support broader critiques of progressive prosecution as a means of transformative change to criminal law.

[211] Not too many years ago, computing power may have been an obstacle. That's no longer the case. Home and office computers are more than capable of data science projects well beyond what's contemplated here. No expensive hardware or software is required.

quality of data can often depend on the quality needed to maintain administrative files. In many cases, this results in very messy data.

A lack of quality, labeled datasets is not unique to law but is *the* classic data science problem.[212] This is the problem that most data science projects face in the real world. It's not easily fixable, it but is something that ought to be gradually overcome with time. Courts and governments may be slow to adapt the digital world, but progress is being made as our world becomes ever more digital. For the short term, it may be enough of an impediment that algorithms cannot be built to study certain laws.

Even where usable datasets exist, dataset disclosure may be another obstacle. In most places, court data is not publicly available and is also not available by request. Most states' Freedom of Information Act laws include barriers to acquiring this kind of data en masse. Sometimes, the cost of making the FOIA requests and paying fees to the government to acquire the data can be prohibitive. In most cases, getting datasets requires government cooperation. For projects that are critical of the government, this is a challenge.

But transparency in local criminal government is becoming more common, and there's hope that transparency will increasingly become the norm. In many places, apolitical non-profits have formed with the explicit goal of procuring and sharing criminal case data.[213] Progressive prosecutors' election campaigns have emphasized transparency and accountability. Some prosecutor offices have released unprecedented amounts of data, while others have established data research units within their offices.[214]

Algorithmic exposure is a tool for more than just the government and well-heeled non-profits.[215] The work can also be completed by smaller community advocacy

---

[212] The reason we have such robust algorithms for image detection is in no small part due to the open source release of ImageNet.

[213] *E.g.*, Texas Justice Initiative | Home Page, , https://texasjusticeinitiative.org/ (last visited Aug 7, 2021); Measures for Justice, , Measures for Justice , https://measuresforjustice.org/ (last visited Aug 9, 2021).

[214] Lucy Lang & Erica Bond, *Prosecutors and the 'Moral Imperative' for Transparency*, The Crime Report (2021), https://thecrimereport.org/2021/03/19/prosecutors-and-the-moral-imperative-for-transparency/ (last visited Aug 9, 2021). Kim Foxx, the prosecutor for Cook County, Illinois (which includes Chicago), famously released a dataset that detailed what happened in every felony case her office processed over a six-year period. Matt Daniels, *The Kim Foxx Effect: How Prosecutions Have Changed in Cook County*, The Marshall Project (2019), https://www.themarshallproject.org/2019/10/24/the-kim-foxx-effect-how-prosecutions-have-changed-in-cook-county (last visited Aug 9, 2021).

[215] Admittedly, larger non-profits are already well-positioned to conduct the data-intensive work of algorithmic exposure. Organizations like the Vera Institute and the ACLU have data research teams and projects. Two of the largest foundations funding criminal law and pretrial reform, Arnold Ventures and the MacArthur Foundation, have heavily funded data-driven research and reforms.

groups, like bail funds and court-watch programs.[216] In recent years, these community groups have quietly become more sophisticated authors of empirical and data-driven projects. Community bail funds operate across the country to challenge state and county money bail systems.[217] They are well known for posting money bond for people who would otherwise remain in jail pretrial on bond amounts that they cannot afford.[218] By posting bond without considering a person's potential dangerousness or flight risk, these groups challenge the efficacy and wisdom of the money bail system and the judges who operate it.[219] Less well known are their broad advocacy efforts, including detailed, data-filled reports on local criminal legal systems' functioning.[220]

Many grassroots organizations are undertaking increasingly sophisticated data projects.[221] They're not shying away from numbers, and they're using methods one would typically associated with larger government organizations or non-profits.[222] With comprehensive, freely available data science education and software available on

---

[216] This case study looks outside the boundaries of traditional expertise and data-driven authority for two reasons: 1) to show the breadth of opportunities available to use algorithms as a tool for critiquing law, and 2) to bring attention to community groups' underappreciated data analysis capabilities.

[217] *See* National Bail Fund Network, , Community Justice Exchange , https://www.communityjusticeexchange.org/en/nbfn-directory (last visited Aug 10, 2021).

[218] *See* Johna Engel Bromwich, *How a Minnesota Bail Fund Raised $20 Million to Help Jailed Protestors*, N.Y. Times, June 1, 2020, https://www.nytimes.com/2020/06/01/style/minnesota-freedom-fund-bail-george-floyd-protests.html (last visited Aug 10, 2021).

[219] *See* The Bail Project, https://bailproject.org/ (last visited Aug 10, 2021).

[220] *E.g.*, Monitoring Cook County's Central Bond Court: A Community Courtwatching Initiative, (2018), https://chicagobond.org/wp-content/uploads/2018/10/courtwatching-report_coalition-to-end-money-bond_final_2-25-18.pdf (last visited Aug 10, 2021).

[221] The case study also bridges a divide in the literature between scholarship that advocates for an expert, empirical path to criminal law reform and scholarship that champions a democratic, community-driven approach. *See* Erin Collins, *supra* note 19 (collecting sources); *see also* John Rappaport, *Some Doubts About "Democratizing" Criminal Justice*, 712 Univ. Chic. Law Rev. 711, 715–17 (2019) (tracing this tension in the literature). This case study opens the question of whether the empirical and the democratic paths are as diametrically opposed as the literature may seem to suggest. *See* Benjamin Levin, *Criminal Justice Expertise* ___ Fordham L. Rev. __ (forthcoming 2022) (analyzing different conceptions of expertise as reflective of underlying political and ideological values). For decades, data analysis was the exclusive purview of professional experts within large institutions. Their capabilities still dwarf those of bail funds and court-watch programs, but they no longer have a monopoly. With limited funding and labor, grassroots organizations need to be selective with the data projects they undertake, but these projects are increasingly part of their portfolio.

[222] Philadelphia Bail Fund, Rhetoric vs. Reality: The Unacceptable Use of Cash Bail by the Philadelphia District Attorney's Office During the COVID-19 Pandemic 5 (2020) (using random sampling to make inferences about the District Attorney's office's broader practices).

the internet, machine learning is an increasingly common skill.[223] Large scale data science projects are now possible outside traditional institutions.[224] Community bail funds have become increasingly sophisticated collectors and disseminators of qualitative and quantitative data, including their own alternative datasets that track the functioning of criminal legal systems. Many bail funds have started or partnered with court-watch programs in which volunteers observe court proceedings and collect qualitative and quantitative information.[225] Court-watch programs have used this information to track judges' adherence to pretrial laws,[226] to share data on pretrial incarceration decisions, and to share stories of the daily injustices that occur within our criminal courts.[227]

The Tyranny of Metrics

Data science is a sociotechnical process, and researchers play an active role in constructing meaning from data.

Using algorithms to critique law embodies the "studying up" research that co-authors and I have encouraged data scientists and the broader algorithmic fairness research community to pursue.[228] In prior work, we introduced the algorithmic fairness community to the anthropological concept of "studying up,"[229] by drawing upon Prof. Laura Nader's call for her fellow anthropologists to move beyond the study of people and cultures at the peripheries of Western society — who comprised the bulk of the anthropological canon — in favor of studying how power operates through elite institutions and positions of authority.[230] This article answers that call to "study up" by re-purposing legal algorithms to study the laws that algorithms are typically intended to

---

[223] David Venturi, *I ranked every Intro to Data Science course on the internet, based on thousands of data points*, freeCodeCamp.org (2019), https://www.freecodecamp.org/news/i-ranked-all-the-best-data-science-intro-courses-based-on-thousands-of-data-points-db5dc7e3eb8e/ (last visited Aug 10, 2021).

[224] [Insert reference to the data for good movement and volunteer opportunities for data scientists].

[225] Jia Tolentino, *Where Bail Funds Go From Here*, The New Yorker, June 23, 2020, https://www.newyorker.com/news/annals-of-activism/where-bail-funds-go-from-here (last visited Aug 10, 2021).

[226] *E.g.*, Monitoring Cook County's Central Bond Court: A Community Courtwatching Initiative, *supra* note 201 at 35–37.

[227] *E.g.*, Court Watch NYC Spring Newsletter, 3 (2018), https://static1.squarespace.com/static/5a21b2c1b1ffb67b3f4b2d16/t/5b0ee79cf950b7742628e513/1527703453246/Spring+2018+CWNYC+Newsletter.pdf (last visited Aug 10, 2021).

[228] Barabas et al., *supra* note 17 at 1.

[229] *Id.* at 1.

[230] Laura Nader, *Up the Anthropologist: Perspectives Gained from Studying Up* 29 (1972).

optimize. Instead of studying and predicting the behavior of marginalized groups, algorithms have an alternate potential of studying and exposing how the law uses the language and logic of prediction to justify legal judgments against these groups.

Demystify metrics.

Demystify prediction.

Surfacing normative values in algorithmic and legal discourse.

Conclusion

The future of criminal law and the prospects for racial justice are tied up with the future of machine learning and artificial intelligence. As machine learning and artificial intelligence reshape law and society, activists and scholars have identified how this new technology poses the threat of generating new inequalities and "cloak[ing] and amplifying existing ones."[231] This concern is legitimate and important. When limited to optimizing current practices, legal algorithms entrench and perpetuate current inequality. But the law is not neutral backdrop in this process, and algorithms can do more than optimize how a law is applied. Algorithms have an overlooked potential to expose unjust laws and provide empirical support for transformative changes. In campaigns to reshape our criminal legal systems, algorithms should be recruited as a welcome — if unexpected — ally.

Introduction

Legal algorithms seem up to no good. Public assistance algorithms have incorrectly accused and fined thousands of innocent people for alleged unemployment fraud.[232] Child welfare algorithms have instructed agencies to investigate families for child abuse based, in part, upon the family's poverty.[233] And criminal risk assessments have told judges to incarcerate people to protect the public, even when it's almost certain that those people will not commit a violent crime if released.[234] Predictive algorithms were once heralded as an objective way to tackle mass incarceration and its racial disparities,

---

[231] Mimi Onuoha, Notes on Algorithmic Violence (2018), https://github.com/MimiOnuoha/On-Algorithmic-Violence (last visited Aug 10, 2021).

[232] Robert N. Charette, *Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold*, IEEE Spectrum: Technology, Engineering, and Science News (January 24, 2018), https://spectrum.ieee.org/riskfactor/computing/software/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold.

[233] Virginia Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, Wired (January 15, 2018), https://www.wired.com/story/excerpt-from-automating-inequality/.

[234] Chelsea Barabas, Karthik Dinakar & Colin Doyle, *The Problems With Risk Assessment Tools*,

but wherever algorithms have taken over, a familiar pattern emerges: the algorithms make predictions based upon impermissible factors,[235] produce inequitable outcomes,[236] and take unjustifiable state action.[237] Algorithms intended to challenge mass incarceration seem instead to be perpetuating it.

A new field of study of algorithmic fairness, accountability, and transparency has appeared on scene to contain the damage, and a common pattern has emerged. After a legal system replaces humans' predictions with a predictive algorithm, researchers acquire data that reveals that the algorithm is either following unfair rules, producing unfair consequences, or both.[238] Conscientious data scientists explore ways to recalibrate the algorithm to produce fair results within legal constraints.[239] When this effort falls short, advocates seek to remove the unfair algorithm from the field.[240] Proponents of the algorithm push back: Removing the algorithm would be counterproductive because humans tend to perform worse at the same predictive tasks.[241] A mountain of

---

N.Y. Times (July 17, 2019) https://www.nytimes.com/2019/07/17/opinion/pretrial-ai.html.

[235] *See* Sonja B Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 Stanf. Rev. 803, 805 (2014).

[236] *See* Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[237] *See* Charette, *supra* note 1. This is a non-exhaustive list of criticisms of legal algorithms, limited to the issues explored in this article. A separate — and important — line of criticism evaluated how legal algorithms fail to adhere to procedural safeguards such as transparency in decision-making, the opportunity for a hearing, and so on. *See generally* Danielle Keats Citron, *Technological Due Process*, 85 Wash. Univ. Law Rev. 1249, 1251 (2008); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 Stanf. Rev. 1343, 1346 (2017).

[238] *E.g.*, Angwin et al., *supra* note 5.

[239] *E.g.*, Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, ArXiv160905807 Cs Stat (2016).

[240] *See*, *e.g.*, American Civil Liberties Union, *How to Fight an Algorithm (ep. 7)*, https://www.aclu.org/podcast/how-fight-algorithm-ep-7 (last visited May 19, 2021); The Leadership Conference on Civil and Human Rights, More than 100 Civil Rights, Digital Justice, and Community-Based Organizations Raise Concerns About Pretrial Risk Assessment (2018), https://civilrights.org/2018/07/30/more-than-100-civil-rights-digital-justice-and-community-based-organizations-raise-concerns-about-pretrial-risk-assessment/ (last visited Aug 14, 2020); Amnesty International, *Ban facial recognition technology*, https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/ (last visited May 19, 2021); Matthew Guariglia, *Technology Can't Predict Crime, It Can Only Weaponize Proximity to Policing*, Electronic Frontier Foundation (2020), https://www.eff.org/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing (last visited Sep 21, 2020).

[241] Alex P. Miller, *Want Less-Biased Decisions? Use Algorithms.*, Harvard Business Review, 2018, https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms (last visited

psychological research has shown how difficult it is for humans to reason statistically and how personal and cognitive biases distort human prediction.[242] Algorithms have been introduced to mitigate these human errors. Here, the algorithmic fairness literature presents a dilemma: We can either accept algorithms, despite their documented shortcomings, or we can defer to humans, who will produce worse results.[243] The discussion of fair legal algorithms often stalls at this peculiar conundrum. Somehow, algorithms are both unjustifiable and the best option available.

Something doesn't seem right. This apparent dilemma is the result of an error in causal reasoning. When an algorithm produces data that reveals unfairness, that unfairness is often attributed to the algorithm. But in many cases, predictive algorithms are not independently *causing* harm — they are instead *revealing* the harms of the laws that they apply.

We've been looking at the problem the wrong way. Replacing humans with algorithms is the way to fix the law *only if* the problem with the law is that humans are making poor predictions. But what if the problem is the law itself? What if the most accurate, least biased applications of our criminal laws would still result in racial inequities and mass incarceration? In these circumstances, replacing humans with algorithms won't ever do the trick. Optimizing unjust laws will simply produce optimized injustice — and worse, it will hide that injustice behind a mask of scientific objectivity.

Perhaps surprisingly, these same algorithms could be recruited as an ally in efforts to transform our legal systems. Rather than accidentally revealing a law's negative outcomes when applying that law, machine learning algorithms could be repurposed to study laws and project their effects. This article coins a phrase — "algorithmic exposure" — for a process wherein standard machine learning techniques can be used to assess a law's fairness and efficacy. Constructing a legal algorithm requires testing a world of possible predictions within a given legal framework. Machine learning is a process of trial and error that involves exploring, primarily through automation, countless ways that a particular prediction can be made.[244] Machine learning could be used to generate a range of predictive models that represent the scope of what predictions are permissible under a law. These predictive models are a sandbox for discovery. By charting the many ways that a prediction can be made under a law, algorithms can reveal the scope of what a law can achieve. Three insights about legal prediction can be systematically extracted: accuracy, distribution of outcomes, and strength of predictive factors.

Apr 26, 2021); James Austin, Sarah L Desmarais & John Monahan, *Open Letter to the Pretrial Justice Institute.pdf*, http://www.jfa-associates.com/publications/Open%20Letter%20to%20the%20Pretrial%20Justice%20Institute.pdf (last visited May 19, 2021).

242 Miller, *supra* note 10. (collecting research).

243 Sharad Goel et al., *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment*, SSRN Electron. J. (2018), https://www.ssrn.com/abstract=3306723 (last visited Sep 25, 2019).

244 *Id.* at 21.

Algorithms' much-ballyhooed strength is that they exploit huge datasets and an array of statistical methods to outperform humans at making predictions.[245] But because algorithms are the best available option for prediction, their shortcomings are telling. Limits on algorithms' predictive accuracy reveal the limits of predictive accuracy possible under a given law. Algorithms can also reveal the possibilities for how a law's outcomes and errors can be distributed across the population. And the process of constructing these models reveals the comparative strength and relevance of different predictive factors.[246] This knowledge can reinforce legal critiques by revealing how a law falls short of its intended purpose, when a law systematically burdens certain groups, or how a law works in arbitrary, inefficient, or redundant ways.[247]

Repurposing predictive algorithms as tool for critiquing law dovetails with a recent turn in algorithmic-fairness scholarship that encourages the field to locate unfairness at structural levels and engage with questions of political economy.[248] As part of a new wave of algorithmic scholarship, this Article seeks to center questions of power, marginalization, and structural inequality within algorithmic discourse.[249] This Article follows up on a call that I and co-authors previously made to encourage data scientists and other researchers to shift their research focus away from using algorithms to study only marginalized populations and toward using algorithms to study people and institutions of power and authority.[250]

---

[245] Goel et al., *supra* note 12 at 2.

[246] This disparate treatment can happen even when a law is written in neutral language, has not been designed to be discriminatory, and is being administered by court actors who intend to treat people equally.

[247] Conventional algorithmic fairness approaches have had too narrow a view of both law and algorithms. Algorithms' potential has been limited to the role of optimizing existing legal practices. And the scope of algorithmic fairness has been confined to scrubbing a decision-making process clean of overt human bias. Interventions have looked to tweak the algorithm in its current task but have not looked to problems with the task itself. This approach can also synthesize seemingly isolated studies of algorithmic fairness by folding one-off critiques of a particular algorithm's predictions or outcomes into a broader analysis of the background legal process.

[248] *E.g.*, Ben Green & Salomé Viljoen, *Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought* 13 (2020); Chelsea Barabas, *Beyond Bias: Re-Imagining the Terms of 'Ethical AI' in Criminal Law*, Geo J Mod Crit Race Persp (2019), https://www.ssrn.com/abstract=3377921 (last visited Oct 27, 2021).

[249] *See* Frank Pasquale, *The Second Wave of Algorithmic Accountability*, Law and Political Economy (2019), https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/ (last visited Jul 31, 2020); Ben Green, *Algorithmic Imaginaries: The Political Limits of Legal and Computational Reasoning*, LPE Project (2021), https://lpeproject.org/blog/algorithmic-imaginaries-the-political-limits-of-legal-and-computational-reasoning/ (last visited May 19, 2021).

[250] Chelsea Barabas et al., *Studying Up: Reorienting the study of algorithmic fairness around issues of power* 9 (2020).

The payoff is not just academic. This article uses pretrial incarceration doctrine as a lens for examining algorithmic exposure's potential as a tool for political advocacy, legislation, and agency policymaking. Across these domains, predictive algorithms should be repurposed as a routine, best practice for diagnosing the efficacy and fairness of predictive laws.[251] Advocacy groups could use algorithms to clarify their policy positions, educate the public, and galvanize support for their cause. Within the bail and pretrial reform movements, algorithmic exposure can help orient advocacy away from purely procedural reforms and toward substantive doctrinal change. In a political environment in which legal reforms must be evidence-based — and evidence-based reforms are typically minor tweaks to current practices — algorithms can be repurposed to provide empirical support for a more fundamental reshaping of law and policy.[252]

Legislatures and other rulemaking bodies could use algorithms as a diagnostic tool for assessing and calibrating potential pretrial reforms. Algorithmic exposure can enrich the public and legislative debate over predictive laws by breaking prediction down to its constitutive elements.[253] Legislatures and administrative agencies could subject predictive laws and regulations to algorithmic audits to ensure that the laws are able to achieve their goals and are producing fair outcomes. The process of writing predictive laws could include the use of predictive algorithms from start to finish.[254] Algorithms could be used to test and compare draft legislation to see which versions of a law best adhere to legislative goals and priorities.

Independent of statutory changes, prosecutors could use algorithmic analysis of pretrial doctrine to justify less carceral pretrial policies — both in court and to the public. In recent years, progressive prosecutors have been elected in major cities across the country.[255] These district attorneys have turned away from tough-on-crime rhetoric, denouncing the racism pervasive to the criminal legal system and rejecting incarceration as the only means to protect the public.[256] But the policies that these prosecutors have adopted for money bail and pretrial incarceration may be at odds with their broader message. Although progressive prosecutors talk of a clean break with their predecessors' track record on pretrial incarceration, their approach has been a gentler applica-

---

[252] *See* Erin Collins, Shifting the Evidence-Based Paradigm (forthcoming) (on file with the author).

[253] Bernard E. Harcourt, *The Systems Fallacy: A Genealogy and Critique of Public Policy and Cost-Benefit Analysis*, 47 J. Leg. Stud. 419–447 (2018).

[254] Insights gained through the algorithmic exposure of law do not then require that the predictive law be administered by algorithms. There's no iron rule that software must be integrated into the legal process. A law could be made simpler for the sake of honest and transparency while maintaining predictive accuracy.

[255] Allison Young, *The Facts on Progressive Prosecutors*, Center for American Progress , https://www.americanprogress.org/issues/criminal-justice/reports/2020/03/19/481939/progressive-prosecutors-reforming-criminal-justice/ (last visited May 14, 2021).

[256] *Id.*

tion of pretrial incarceration doctrine, rather than an alternate paradigm.[257] Algorithms could help progressive prosecutors reshape their pretrial policies by using data from the prosecutor's own district to show how the local community would fare under current, past, and potential policies.[258] If a prosecutor's office chose to turn away from a purely predictive model of pretrial incarceration and adopt pretrial policies more aligned with their broader vision of reform, this data could justify that change in court and in the bully pulpit. And if a prosecutor's office fell short of that commitment, advocacy groups could use these same tools to hold the office accountable.

The Article has four parts. Part I examines a tension at the heart of contemporary algorithmic fairness debates. As algorithms proliferate through our legal systems, they somehow seem to be both responsible for untold harm and the best option for any predictive-legal problem. This dilemma stems from an error in causal reasoning. Harm that is assumed to stem from an algorithm is often harm that an algorithm has revealed but not caused.

Part II demonstrates how machine learning can be used to expose how a predictive law works and what outcomes it produces. In part because algorithms are the most accurate method of prediction available, they can reveal the limits of what a predictive law can achieve. Algorithms can reveal how accurately an outcome can be predicted and the possible distributions of outcomes and errors within the population. They can also reveal what factors are useful for this prediction and the predictive strength of those factors. These insights could be used to better understand and evaluate laws by showing when a law falls short of its intended purpose, when a law systematically burdens certain groups, and how a law works in arbitrary, inefficient, or redundant ways.

Part III argues that algorithmic exposure should be adopted as a routine, best practice for diagnosing the efficacy and fairness of predictive laws. Pretrial incarceration doc-

---

[257] Rachael Rollins On High Bail: "That Is Not How We Operate," , News (2020), https://www.wgbh.org/news/local-news/2020/09/24/rachael-rollins-on-high-bail-that-is-not-how-we-operate (last visited May 19, 2021); Chris Palmer, *Tensions are boiling over between Philly DA Larry Krasner and bail reform advocates*, https://www.inquirer.com , https://www.inquirer.com/news/philadelphia/philadelphia-da-larry-krasner-cash-bail-reform-advocates-20200729.html (last visited May 19, 2021); Colin Doyle, *Chesa Boudin's New Bail Policy is Nation's Most Progressive. It Also Reveals Persistence of Tough-on-Crime Norms.*, The Appeal Political Report , https://theappeal.org/politicalreport/chesa-boudin-cash-bail-predictions/ (last visited May 19, 2021). This approach to pretrial incarceration emerged in the tough-on-crime Nixon era, was codified into law in the 1980s, and has long been opposed by scholars and activists as violating the presumption of innocence and harming communities in a racially discriminatory manner — all without improving public safety.

[258] At its most accurate, preventive pretrial incarceration flips the Blackstone principle on its head: to prevent one person from committing a violent crime, ten legally innocent people must be incarcerated. Colin Doyle, *All Models Are Wrong, But Are Risk Assessments Useful?*, *in* American Society of Criminology DCS Handbook on Corrections and Sentencing: Pretrial Justice (2021).

trine serves as a lens for exploring algorithmic exposure's potential within political advocacy, legislation, and agency policymaking. Advocacy groups, legislatures, and government agencies could all repurpose legal algorithms to inform their work of galvanizing political support, crafting legislation, setting policy, and litigating cases.

Part IV takes a step back to address both concerns and opportunities with this approach. Algorithmic exposure may be impeded by biased data, poor data quality, and policies against data disclosure. The process itself may be coopted. And longstanding critiques of empirical approaches to legal analysis apply here as well.

In closing, the article gestures toward a broader vision, of which this article plays one small part. We are at the cusp of dramatic empirical and technological change for our legal system. The future of criminal law and the prospects for racial justice are tied up with the future of machine learning and artificial intelligence. When used to optimize extant practices, legal algorithms consolidate and preserve entrenched power, inequities, and ideology. But this is not the only way that algorithms can be used. There is opportunity yet to harness algorithms' potential to expose inequities that the law creates and sustains.

Background: Legal Predictions

Prediction guides many of the decisions that judges, police, bureaucrats, and other legal actors must make every day. Before issuing a preliminary injunction, a judge must predict whether the plaintiffs will win their case on the merits.[259] Police must have probable cause for many arrests, searches, and seizures to be constitutionally permissible.[260] State unemployment agencies use predictions of fraud to grant or deny people unemployment benefits.[261] Child welfare agencies triage investigations of suspected neglect based on predictions of which claims will be substantiated,[262] while public housing authorities manage waitlists for housing based on predictions of who will use public housing for the shortest length of time before living independently.[263]

Across legal systems nationwide, algorithmic predictions are replacing or informing predictions traditionally made by humans. Today, algorithms can deny a person government food benefits,[264] send a social worker to investigate a home, or ban a person

---

[259] *Preliminary Injunction*, LII / Legal Information Institute, https://www.law.cornell.edu/wex/preliminary_injunction (last visited Aug 8, 2021).

[260] U.S. Const. amend. IV.

[261] Charette, *supra* note 1.

[262] Eubanks, *supra* note 2.

[263] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018).

[264] Lydia X. Z. Brown et al., Challenging the Use of Algorithm-Driven Decision-Making in

from flying on commercial airlines.[265] In many places, criminal procedure is now algorithmic from start to finish. Based on predictions of wrongdoing, algorithms encourage police to investigate,[266] judges to incarcerate,[267] probation to surveil,[268] and parole boards to deny release.[269]

Predictive algorithms arrived in law with much fanfare from policymakers and academics. Algorithms were seen as a way to boost predictive accuracy and efficiency, while reducing bias and error.[270] Law, particularly criminal law, is riddled with race and class inequities.[271] To the extent that these inequities are the result of imperfect human decision-making within a legal process, algorithms may provide a helpful course correction. Algorithms tend to be both less biased and more accurate than humans making the same predictions.[272] When predicting a person's likelihood of committing a crime if released on bail, a judge may consider the person's apparent race, gender, class, or cultural background. Through implicit bias, this can happen even when a judge is trying *not* to consider that information. But an algorithm will only consider the factors it has been programmed to consider.[273] Judges' predictions are likewise distorted by cognitive biases. Decades of empirical research demonstrate how humans do not reason statistically and tend to systematically overestimate or underestimate the likelihood of

---

Benefits Determinations Affecting People with Disabilities 31 (2020).

265 Spencer Ackerman, *No-fly list uses "predictive assessments" instead of hard evidence, US admits*, The Guardian, August 10, 2015, https://www.theguardian.com/us-news/2015/aug/10/us-no-fly-list-predictive-assessments (last visited Sep 21, 2021).

266 Andrew G. Ferguson, The rise of big data policing: surveillance, race, and the future of law enforcement (2017); Rashida Richardson, Jason Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N. Y. Univ. Law Rev. 15, 19 (2019).

267 Doyle, *supra* note 25.

268 Cade Metz & Adam Satariano, *An Algorithm That Grants Freedom, or Takes It Away*, The New York Times, February 6, 2020, https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html (last visited Aug 8, 2021).

269 Angwin et al., *supra* note 5.

270 *E.g.*, Kamala Harris & Rand Paul, *To Shrink Jails, Let's Reform Bail*, N.Y. Times, July 20, 2017, https://www.nytimes.com/2017/07/20/opinion/kamala-harris-and-rand-paul-lets-reform-bail.html (last visited Aug 9, 2021); Jon Kleinberg, Jens Ludwig & Sendhil Mullainathan, *A Guide to Solving Social Problems with Machine Learning*, Harvard Business Review (2016), https://hbr.org/2016/12/a-guide-to-solving-social-problems-with-machine-learning (last visited Sep 4, 2018); Miller, *supra* note 10.

271 Becky Pettit & Bruce Western, *Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration*, 69 Am. Sociol. Rev. 151, 151 (2004).

272 Goel et al., *supra* note 12 at 2.

273 Ethem Alpaydin, Machine Learning (2016).

events.[274] To the extent that legal decision-making is undermined by humans' imperfect thinking, algorithms should be able to provide a helpful course correction.

But in recent years, algorithms have come under scrutiny. Data-driven tools can bypass some human shortcomings, but they can still follow unfair rules and produce unfair outcomes. Legal algorithms have made predictions using biased data, have produced inequitable outcomes, and have recommended that state actors take unjustifiable actions.[275] And when developers refuse to disclose their data and methods — as they often do — the entire process can be hidden from scrutiny.[276] A burgeoning field of study of algorithmic fairness, accountability, and transparency seeks to address these concerns.[277] In the field of algorithmic fairness, a common pattern is to identify an "unfair" algorithm in some real-world domain, document its harms, and explore ways to produce fair results.[278]

Although there are a variety of ways that algorithmic harm might be redressed, the field has settled on three techniques: recalibrate the algorithm, restrict the algorithm, or remove the algorithm. Recalibrating an algorithm involves changing how the tool works internally: training the tool on different data, generating predictions based on different features, or following different statistical processes for making predictions.[279] Restricting the algorithm means using the algorithm for a different decision-making process.[280] And if the algorithm can't be recalibrated and can't be repurposed, removing the algorithm from the field is always an option.[281]

In some circumstances, this menu of options may be sufficient. But in other cases, these options cannot fully address the unfairness they target. Pretrial risk assessments are one such domain. In recent years, actuarial risk assessments have become the hallmark of both bail reform and algorithmic fairness discourse.[282] The tools encourage judges to incarcerate or release a defendant pretrial based on predictions of whether

274 See generally Daniel Kahneman, Thinking, fast and slow (2012). Unlike humans, algorithms make predictions without cognitive biases. Goel et al., supra note 12 at 2.

275 Sandra G. Mayson, *Bias In, Bias Out*, 128 Yale Law J. 2218, 2221–22 (2019).

276 Wexler, *supra* note 6 at 1349.

277 Home :: FAT ML, , https://www.fatml.org/ (last visited Aug 8, 2021).

279 *E.g.*, Richard Berk et al., *Fairness in Criminal Justice Risk Assessments: The State of the Art*, ArXiv170309207 Stat, 3 (2017).

280 Sandra G. Mayson, *supra* note 42 at 2286.

281 Shira Ovide, *A Case for Banning Facial Recognition*, The New York Times, June 9, 2020, https://www.nytimes.com/2020/06/09/technology/facial-recognition-software.html (last visited Aug 8, 2021).

282 Colin Doyle, Chiraag Bains & Brook Hopkins, Bail Reform: A Guide for State and Local Policymakers 14 (2019).

the person will be arrested or miss a court date.[283] Critics of the tools contend that risk assessments are unjustified because they rely on racially biased data, produce racially inequitable outcomes, and have limited accuracy.[284] Conventional approaches to fixing these algorithms have proven insufficient.[285] Neither recalibrating, repurposing, nor removing the algorithms works. The algorithms cannot be recalibrated using unbiased policing and court data, because no such data exists.[286] Repurposing or removing the algorithms only does so much, because pretrial incarceration decisions still need to be made. Only now, judges would have to make these decisions without algorithmic assistance. As supporters of risk assessment tools are quick to remind us, algorithms ought to make less biased and more accurate predictions than judges.[287] So long as these predictions are being made, algorithmic predictions should be preferable to human predictions.

What to do? The academic literature and public policy debates are replete with permutations of this critique and defense.[288] The discussion always stalls at the same impasse: Risk assessments are somehow both unjustifiable and the best option available. The usual approaches to algorithmic fairness cannot resolve the deadlock.[289]

Consider the controversy over racial disparities with pretrial risk assessment algorithms. In 2016, a ProPublica investigation led with the headline "There's software used across the country to predict future criminals. And it's biased against Blacks."[290] The report found that a pretrial risk assessment algorithm had different error rates for Black people than white people.[291] The tool incorrectly predicted Black people as being at high risk of future crime much more frequently than it incorrectly predicted white people as being at high risk. And it incorrectly predicted white people as being at low

---

283 Doyle, Bains, and Hopkins, *supra* note 48.

284 *Id.* at 14–16.

285 Chelsea Barabas, Karthik Dinakar & Colin Doyle, Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns (2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf (last visited Oct 3, 2019).

286 Ngozi Okidegbe, *Discredited Data*.

287 Goel et al., *supra* note 12 at 2.

288 Sandra G. Mayson, *supra* note 42 at 2227–33.

289 *Id.* at 2248–49. Both the critique and the defense of risk assessments are incomplete in their own ways. The critics have held back from following their argument to its necessary conclusion. If algorithms are both unjustifiable and better than any available alternatives, then the critics' concerns would seem to precede the introduction of algorithms. Likewise, by only comparing legal algorithms' performance to humans, the defenders have offered only a partial justification, ignoring the possibility that neither humans nor algorithms can produce predictions capable of justifying important legal decisions.

290 Angwin et al., *supra* note 5.

291 *Id.*

risk of future crime much more frequently than it incorrectly predicted Black people as being at low risk. The report galvanized progressive groups against pretrial risk assessment algorithms,[292] which they still oppose today.[293] The report — and the accompanying dataset — also inspired a wealth of empirical research.[294] Using this data, researchers found that disparate error rates would occur in any predictive system in which different racial groups had different baseline rates of being arrested.[295] In other words, any legal system of purely preventive pretrial incarceration would produce these racial disparities, whether the legal system relied upon algorithms or relied upon humans to make these predictions. It's now understood that the problem of racial disparities in error rates stems from the legal process of preventive pretrial incarceration — not from risk assessment algorithms.[296] What has not been fully appreciated is that a risk assessment algorithm produced the information that led to this insight.[297]

There's a pattern here. Because an algorithm had brought unfairness to light, unfairness was attributed to the algorithm. But oftentimes the algorithm is not the source of the unfairness. Rather, the algorithm is reflecting unfairness within culture, law, and society.[298] Accordingly, the unfairness brought to light by legal algorithms may often be symptomatic of unfairness in underlying law.[299] As the ProPublica example illus-

---

[292] The Leadership Conference on Civil and Human Rights, *supra* note 9 at 100.

[293] Coalition letter on the use of the PATTERN risk assessment in prioritizing release in response to the COVID-19 pandemic, American Civil Liberties Union , https://www.aclu.org/letter/coalition-letter-use-pattern-risk-assessment-prioritizing-release-response-covid-19-pandemic.

[294] *E.g.*, Eugenie Jackson & Christina Mendoza, *Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not*, 2 Harv. Data Sci. Rev. (2020), https://hdsr.mitpress.mit.edu/pub/hzwo7ax4 (last visited Aug 14, 2020); Anthony W Flores, Kristin Bechtel & Christopher T Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks."*, 80 9 (2016); Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Alexandra Chouldechova, *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, ArXiv161007524 Cs Stat (2016), http://arxiv.org/abs/1610.07524 (last visited Aug 14, 2020).

[295] Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Berk et al., *supra* note 45. Absent an intervention that explicitly adjusted predictions based on race. Kleinberg, Mullainathan, and Raghavan, *supra* note 8.

[296] Sandra G. Mayson, *supra* note 42 at 2224–25.

[297] Some critics doubt the value of this insight, preferring alternative definitions of fairness. Flores, Bechtel, and Lowenkamp, *supra* note 60.

[298] Looking "beyond the algorithm" has its own pitfalls. Problems that are structural or systemic are — by definition — pervasive and hard to change. This conception of unfairness risks becoming too abstract and responsibility for the unfairness risks becoming too diffuse.

[299] A new wave of scholarship approaches algorithmic fairness differently by locating unfairness outside the algorithms themselves. In this approach, algorithmic unfairness is often a reflection of deeper structural and cultural problems. Second wave research into pretrial risk as-

trates, legal algorithms have inadvertently been revealing important information about underlying laws. The next Part of this Article examines how we might exploit this hidden talent and repurpose algorithms as a diagnostic tool.

Concept: Algorithmic Exposure

By revealing information about predictions that the law requires, algorithms can teach us about those laws.[300] Consider the following syllogism: Some laws justify legal judgments based upon predictions. Algorithms can reveal information about these predictions. Therefore, algorithms can reveal information about these laws.[301] Like all syllogisms, the conclusion is true only if the premises are true. The first premise — that some laws justify legal judgments based upon predictions — hardly needs proving, as many laws explicitly require this.[302] The second premise — that algorithms can reveal information about these predictions — is the conceptual claim of this article.[303]

Whether done by humans or machines, prediction is fundamentally the same. Patterns from the past are used to anticipate the future. When making predictions, both humans and algorithms attempt to find an underlying pattern that connects information about the present with outcomes in the future.

But machines and humans make predictions in different ways. Humans rely upon intuition, informed by their background experiences and knowledge. With humans, it's

---

sessments has questioned whether these tools can justifiably be relied on to promote criminal justice reform or make incarceration decisions, often concluding that these algorithms entrench and obscure harmful penal ideologies. Chelsea Barabas, *Beyond Bias: Re-imagining the Terms of "Ethical AI" in Criminal Law* 40; Barabas et al., *supra* note 17; Ben Green, *``Fair" Risk Assessments: A Precarious Approach for Criminal Justice Reform* 5 (2018); 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. Univ. Law Rev. 42 (2019). Rather than try to optimize how the algorithms work, this scholarship asks whether an algorithm is needed and what interests it serves. This Article dovetails with second-wave scholarship by reorienting the algorithmic gaze away from marginalized populations and toward institutions of power, asking what machine learning algorithms can reveal about the law itself. Barabas et al., *supra* note 17.

[300] As these tools have proliferated and been subject to research and critique, the assumption has *not* been that they can offer much insight about the law itself. The assumption has been that analysis of how an algorithm applies the law explains only how the algorithm works — not how the law itself works. There have been some exceptions. *E.g.*, Sandra G. Mayson, *supra* note 42; Andrew Ferguson, *The Exclusionary Rule in the Age of Blue Data*, 72 Vanderbilt Law Rev. 561, 594 (2019).

[301] To put it another way:

[302] *See supra* Part I, cataloguing many of these laws.

[303] This Part demonstrates the validity of that premise. Parts III and IV examine potential applications and implications.

hard to know how predictions are being made and whether the results are accurate or equitable. Compared to human prediction, machine prediction is much less mysterious. Machines require more precise programming and information. Predictive models must obey defined rules to produce defined outcomes. With statistical models, it's often possible to observe how a prediction has been made and what outcomes it has produced. In these circumstances, the rules that a law follows and the range of possible legal outcomes can become subject to richer analysis.

To date, machine learning has been used to optimize legal predictions. But machine learning could also be used to explore and find the limits of legal predictions. Machine learning is a process of testing, through brute-force automation, many ways that a particular prediction can be made. Rather than test many predictive models and discard all but the most optimal model, machine learning could be used to uncover the range of ways that a prediction can be made under the rules of a particular law. This exploratory process could regularly and systematically reveal three insights about a legal prediction:

1. How accurately an outcome can be predicted
2. The potential distributions of outcomes and errors
3. The relevancy and strength of factors used to predict a particular outcome

These algorithmic insights could be used to better understand and evaluate laws by showing how a law meets or falls short of its intended purpose; how a law systematically affects different groups; how a law works in arbitrary, inefficient, or redundant ways; and how seemingly neutral legal predictions include value judgments that may conflict with other legal principles and goals.

This Part starts by comparing human and machine predictions. It then explains how machine learning can be used to expose how a predictive law works and what outcomes it produces. The Part concludes by examining the insights that can be routinely derived from this process and identifying how these insights can be used to assess laws.

Understanding Prediction

Whether prediction is made by humans or algorithms — some fundamental elements of prediction are the same. Patterns from the past are used to anticipate the future. When making predictions, both humans and algorithms attempt to find an underlying pattern that connects information about the present with outcomes in the future. Prediction is only possible for situations in which the world follows regular patterns. Dark clouds foretell rain; obesity portends high blood pressure, and SAT scores indicate college prospects. To the extent that factors "X1, X2, X3 … X*n*" reliably preceded outcome "Y" in the past, both humans and machines will predict "Y" when factors "X" arises again.

An important way in which algorithmic and human processes differ is what the computer science community has termed "general intelligence." General intelligence is a type of intelligence that can "possess a reasonable degree of self-understanding and autonomous self-control, and ha[s] the ability to solve a variety of complex problems in a variety of contexts, and [can] learn to solve new problems that [it] didn't know about at the time of their creation."[304] Although machine learning and artificial intelligence have progressed by leaps and bounds in recent years, we are still far from creating machines that have an artificial general intelligence. In contrast, human-oriented systems — like law — depend upon people's capacity for general intelligence. When given any new task, people bring a wealth of background knowledge, habits, common sense, and intuition to bear — for better or worse.

Laws tend to be written with humans' general intelligence in mind. Because humans can rely on general intelligence, the rules for prediction can be vague and under-specified. In law, the rules that must be followed in making predictions are often open-ended or include more factors than one person can juggle.[305] When multiple factors are considered, it's hard to know the proper weight to be given to each factor.[306] Even when the rules for prediction are clear, there will always be uncertainty about a person's fidelity to those rules. Bias — including subconscious bias — and human error can slip in and distort a person's predictions. As a result, we usually don't know how legal predictions are being made. We don't know what a person considers when making a prediction — oftentimes the person making the prediction can't be sure. In the moment, it's hard to identify how accurate or fair those predictions wind up being. Even in circumstances in which outcomes of predictions are observed, it's difficult to discern how to improve those predictions.

Consider an administrative agency like a state's child welfare services. The agency is tasked with investigating alleged child neglect and abuse. But the number of potential cases might overwhelm the agency's staffing capacity. Not every case can be investigated. Some set of rules must be adopted to decide which cases deserve investigation. In a purely human-driven system, the rule might be as simple as "prioritize and investigate the highest risk cases." This might not be an optimal rule — perhaps one with great specificity would produce better results — but it is a functional rule. Bureaucrats working for the agency can, upon reading intake forms and case files, sort cases into what seem to them to be high and low priority. Each state actor in each case must make a subjective determination of whether the case before them is "high risk" or not. In such a system, an outside observer can't ascertain what "high risk" means, what factors the state actor has considered, or how accurate the predictions are. Even if a statute defines high risk or lists specific factors for bureaucrats to consider, people are still influenced by extraneous — sometimes impermissible — information.[307]

---

[304] Artificial Brains VI.

[305] *E.g.*, Cal. Const. Art. I, § 12.

[306] As is shown later, oftentimes the factors themselves bear little relationship, statistically speaking, to the outcome being predicted.

[307] No doubt, such a system may be deeply problematic. Different bureaucrats may have different ideas of what constitutes high risk, resulting in inconsistent decision-making. Some factors that they consider important — like a parent's income or a mental health history — may not be relevant or may be unfair to be used against someone. Bias against marginalized

Because computer systems do not have general intelligence, similar shortcuts are not available when software is programmed to make predictions. If the same child welfare service agency wanted to replace their human-based triage system with an algorithmic system — as many agencies have in recent years[308] — the rules for prediction would need to be more clearly specified.[309] A statistical model cannot be told to simply "prioritize and investigate the highest risk cases." Compared to humans, the algorithmic prediction process is, by mechanical necessity, more rigorous and transparent.[310] When algorithms are introduced, an opaque system of human prediction is replaced with a predictive system that must obey defined rules to produce defined outcomes. Algorithms reduce ambiguity in the predictive process because they require specified predictive variables and specified outcomes of interest.[311] Algorithms reduce uncertainty about outcomes because training the algorithms requires measuring their accuracy against test data.[312] Because machines make predictions in a methodical, testable way, they can provide a window into the ways that a given prediction can be constructed and the results that can be produced. With machine prediction, both the process and the outcomes of predictions can become knowable.[313]

Building predictive models require some setup. Although it's widely understood that machine learning depends upon large-scale datasets, building models is not yet quite as simple as pulling up an Excel sheet of court data and pressing "go." Decisions need to be made before the predictive model can be trained on the dataset and run in the field. In general, the data must be cleaned and split, the outcome of interest must be speci-

---

people may result in Black households or single-parent households being unfairly singled out for investigation. But such a system can still operate, however poorly. Human beings' general intelligence can fill in the gaps in the rules, and humans can make their best guesses under the circumstances.

308 *See e.g.*, Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. Times, January 2, 2018, https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html (last visited Aug 10, 2021); The Allegheny Family Screening Tool, , https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx (last visited Aug 10, 2021).

309 James et al., *supra* note 13 at 21–24.

310 At least to those building the system. See *infra* X for a discussion of transparency and data disclosure concerns.

311 James et al., *supra* note 13 at 21–24.

312 *Id.* at 29–31.

313 Some learning methods are less interpretable than others, particularly deep learning methods. The complexity and uninterpretability of a model does not determine its success at prediction. People may become enamored with complicated methods like convolutional neural networks, but for many problems, neural networks can be inferior to even simple statistical methods. The effectiveness of using algorithms to expose law depends, in part, on the interpretability of the learning method chosen because less interpretable methods yield less information about how a prediction is made.

fied, predictive variables must be defined and labeled, and statistical methods must be chosen and tested.[314] The first step is to split the data into a training dataset and a testing dataset.[315] The model will learn how to predict accurately using the training data. After the model has been built, it must be tested using the testing data to see how well the model can make predictions with data that it has not been trained on.[316]

An output variable must be specified.[317] The output variable — interchangeably referred to as the "outcome of interest" or the "dependent variable" — is the outcome that the model will try to predict. This outcome needs to be defined in the dataset. Each case in the training dataset ought to specify whether the outcome of interest been met. Without an outcome specified with the training examples, the predictive model will not be able to predict the outcomes based on the other variables.

Input variables must also be specified.[318] In machine learning, input variables are often called "predictive variables" or "features," and in statistics, they are often called "independent variables." They generally refer to the same thing: the information in a dataset that may help predict the outcome the outcome of interest. At an early stage in building a model, many features might be considered as it's not yet known what features will correlate with the outcome of interest. The process of constructing and building a machine learning model requires discovering what variables are predictive of the outcome of interest.[319] Before a model can be trained, these potential predictive variables need to be collected and identified within the dataset.[320]

The next step is choosing a statistical learning method.[321] Within legal domains, the methods that tend to be used are often regression and decision tree models,[322] al-

---

[314] John D. Kelleher & Brendan Tierney, Data Science 97–98 (2018).

[315] Giorgos Myrianthous, *How to Split a Dataset Into Training and Testing Sets with Python*, Medium (2021), https://towardsdatascience.com/how-to-split-a-dataset-into-training-and-testing-sets-b146b1649830 (last visited Aug 10, 2021).

[316] The aim is not to build a model that just predicts the training data well. The aim is to build a model that can be effective at predictions in the field. A good model is effective at making predictions using new data that the model has not been trained on.

[317] James et al., *supra* note 13 at 15.

[318] *Id.* at 15.

[319] Randy Au, *Data Science foundations: Know your data. Really, really, know it*, Towards Data Science 105–10 (2019), https://towardsdatascience.com/data-science-foundations-know-your-data-really-really-know-it-a6bb97eb991c (last visited Feb 21, 2019).

[320] With machine learning, most of the human-level work is the cleaning and labeling of datasets so that researchers can construct a model that produces meaningful results.

[321] James et al., *supra* note 13 at 21.

[322] Doaa Abu Elyounes, *Bail Or Jail? Judicial Versus Algorithmic Decision- Making in The Pretrial System*, 21 Colum Sci Tech Rev 376, 381 (2020); Jon Kleinberg et al., *Human Decisions and Machine Predictions**, Q. J. Econ., 239 (2017), http://academic.oup.com/qje/

though this may change in the future, particularly since machine learning is a rapidly changing field. For a given project, researchers often explore multiple statistical learning methods, as different methods present different tradeoffs. Even after choosing a particular statistical method, researchers may explore different variations on the same method — often with the help of automated tools.[323] There is no one-size-fits-all perfect method.[324] One method may work very well on a particular dataset, while a different method may work better on a very similar dataset.[325] Ultimately, the "right" choice of a statistical learning method is a domain-specific issue that depends upon the data being used and the problem being solved. Once the data is cleaned, the input and output variables are specified, and a statistical method is chosen, the machine learning can begin.

Assessing Law

These predictive models are a sandbox for discovery. It's an opportunity to explore the universe of possibilities that a legal framework allows.[326] By exploring the various ways that a prediction can be made in accordance with the predictive factors prescribed or allowed by underlying law, algorithmic exposure can reveal the limits of a given legal doctrine or theory.[327] Three insights about legal prediction can be regularly and sys-

---

article/doi/10.1093/qje/qjx032/4095198/Human-Decisions-and-Machine-Predictions (last visited Sep 4, 2018).

[323] Jeremy Jordan, *Hyperparameter tuning for machine learning models.*, jeremyjordan.me (2017), https://www.jeremyjordan.me/hyperparameter-tuning/ (last visited Aug 10, 2021). The goal is the same: we assume that there is a relationship between the input data and the output data. Different statistical learning methods make different assumptions about that relationship. It's often a process of trial and error to determine which method best fits the project at hand.

[324] Different methods have different benefits and drawbacks. Some methods are more interpretable — that is, humans are better able to understand how and why the model makes predictions. And other methods are more flexible: the model is better able to adjust its shape to fit the data. But flexibility often comes at the cost of interpretability, and vice versa. James et al., *supra* note 13 at 25.

[325] *Id.* at 29.

[326] The measurement possibilities extend beyond calculating the results of one specific prediction. Experimenting with different models can reveal how outcomes change when certain factors are weighted differently or removed altogether. With machine learning to inform the process, researchers can learn how changes to the legal rules guiding a prediction would affect the outcomes produced.

[327] The outcomes across many different predictive models can vary, but only by so much. In legal systems that ask prediction to do too much, it can be helpful to learn what cannot be achieved through prediction.

tematically extracted from these models: accuracy, distribution of outcomes, and strength of predictive factors.

Accuracy. Algorithms can reveal how accurately a legal outcome can be predicted as measures of accuracy are intrinsic to constructing a predictive model.[328] The very process of learning a model is iteratively adjusting how the model considers predictive factors to arrive at a model that most accurately predicts outcomes.[329] Building a model that optimizes accuracy necessarily reveals the limits on our ability to accurately predict an outcome based on the factors being used. Because machine learning models outperform humans at prediction, the limits on a model's accuracy represent the accuracy limits for any form of prediction.[330] Therefore, if a machine learning model can predict who will commit employment fraud with a high degree of accuracy, humans will be unlikely to outperform the statistical model. And if a machine learning model cannot predict who will commit employment fraud with a high degree of accuracy, then humans will still be unlikely to outperform the statistical model.[331]

Distribution of Outcomes and Errors. The machine learning process reveals how a law's outcomes and errors will be distributed across the population.[332] If the dataset is labeled for any characteristic — such as the race, gender, or age of a person — the model can produce information about the outcomes and error rates across different groups. The value here rests in the quantity of models that can be produced. The process doesn't just reveal one legal prediction's outcomes and errors. Rather, the process can reveal the distribution of outcomes and errors of the full range of different predictions permissible under law.

Strength and Relevancy of Predictive Factors. The process of constructing a machine learning model can also reveal what predictive variables can be relevant to the prediction at hand and the strength of different predictive variables that are considered. Both machine learning and human intuition are processes of pattern recognition.[333] The

---

[328] James et al., *supra* note 13 at 29.

[329] Id.

[330] Goel et al., *supra* note 12 at 2.

[331] An assumption built into this claim is that there is an adequate dataset for the machine learning model to be trained upon. Without relevant data, statistical models cannot predict anything and therefore cannot outperform humans. One might expect that a combination of judges and algorithms would fare better than algorithms on their own, but that's not case. Judicial overrides have been found to produce worse outcomes than algorithms acting alone. Studies of judges, probation officers, and other criminal justice professionals all reveal that human overrides tend to decrease accuracy. *Id.* at 3–4.

[332] James et al., *supra* note 13 at 20.

[333] They both use patterns from the past to predict outcomes in the future. For unstructured prediction, humans rely upon intuition. Consider the example of a multi-factor test in law. A multi-factor test is not a form of structured judgment that walks a state actor through the process of arriving at a prediction. Instead, a multi factor test anchors a judge's perception

question that both humans and machines ask is, how does a change in factor "X" affect output "Y?"[334] Statistical correlation reveals how the presence of "X" relates to outcome "Y." If there's no relationship between factor "X" and output "Y" — in other words, if changes in "X" do not result in changes in "Y" — then that factor is not helpful for making predictions. When humans make predictions, we can guess that that a relationship between "X" and "Y" isn't there. When we attempt the same task with statistical learning, we can learn definitively whether a relationship exists.[335] Likewise, through statistical learning we can learn the relative strength of different factors.[336] This can reveal what factors, compared to other factors, are the most helpful at predicting the outcome. It can also reveal which factors are not independent of each other but are both capturing the same signal.

Insights about accuracy, distribution of outcomes, and strength of predictive factors can be used to ascertain a law's systemic potential. These insights are not exclusive of one another but can work together to reveal: how a law falls short of its intended purpose; how a law systematically affects certain groups; how a law works in arbitrary, inefficient, or redundant ways; and how seemingly neutral predictions include value judgments that may conflict with other legal principles and goals.

How a law falls short of its intended purpose. Algorithms can expose how effective a law is at achieving its own objectives. Accuracy is the essential insight here. Insights into accuracy can reveal how frequently legal decisions will be made based on incorrect predictions. Consider a preventive incarceration law that seeks to imprison people who would harm others and seeks to release people who would not harm others. The ideal application of this law would correctly identify both sets of people: those who would harm others and those would not. In this case, algorithms can reveal how frequently any application of this law makes errors in both directions: either mistakenly identifying people as dangerous or mistakenly identifying people as not dangerous.

How a law systematically affects certain groups. If the errors in prediction fall upon certain groups rather than others, the law may violate commitments to equal treatment and distributional fairness. Through algorithmic construction, one can see the distribution of outcomes for not just one particular application of a law or legal doctrine but the whole range of possibilities. If particular ranges of outcomes are better than others,

---

to certain factors, and the judge makes an intuitive prediction based on a mental consideration of those factors. In contrast, predictive models rely upon the statistical relationship between inputs and outputs. Kelleher and Tierney, *supra* note 79 at 104–05. Intuition is unavailable with machine learning, and statistical correlation replaces it.

334 With humans, it's intuition and experience. With machine learning models, it's statistical correlation.

335 *See* Rahil Shaikh, *Feature Selection Techniques in Machine Learning with Python*, Towards Data Science (2018), https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e (last visited Aug 10, 2021).

336 *Id.*

an algorithm can reveal which applications of a legal doctrine should be foreclosed and which should be permitted.

How a law works in arbitrary, inefficient, or redundant ways. By revealing the relevancy and strength of different predictive factors included in the law, algorithmic exposure can reveal whether the predictors included in the law have a statistical relationship with the outcome of interest. Algorithms can reveal that a law relies upon irrelevant or misleading predictive factors. Accuracy matters here, too. If a certain type of prediction cannot be made accurately and reliably, then a law depending on that prediction may produce arbitrary results.

How seemingly neutral legal predictions include value judgments that may conflict with other legal principles and goals. Some factors that are useful for prediction may conflict with other values in the law. It may be unfair to consider these factors at all or it may be unfair to weigh them so heavily. For example, when predicting whether a parent has committed child abuse or neglected their child, a useful predictive factor is whether that parent was involved in the child welfare system when they were a child.[337] But using an immutable aspect of a person's childhood to predict that person's current actions may betray a value of the legal system to treat people as moral actors with independence and dignity. Algorithms cannot resolve these value-laden questions, but they can reveal how these normative commitments are hidden within seemingly neutral legal predictions.

Exposing Law

Machine learning is a process that is often used to optimize a predictive model.[338] With machine learning, computer software independently tests out many ways of making a prediction to arrive at an optimal way of making that prediction.[339] It's a process

---

[337] Eubanks, *supra* note 2.

[338] Machine learning used for predictions is called "supervised learning." This Article is only concerned with applications of supervised learning. Other popular applications of machine learning include unsupervised learning and reinforcement learning. With unsupervised learning, there is no outcome of interest being predicted. In this context, machine learning is used to better understand the relationship of clusters of data to one another. What is Unsupervised Learning? | IBM, , https://www.ibm.com/cloud/learn/unsupervised-learning (last visited Sep 17, 2021). Reinforcement learning is a method for allowing a computer agent to interact within an environment and learn through a process of trial and error based upon rewards and punishments for its actions. Reinforcement Learning 101. Learn the essentials of Reinforcement… | by Shweta Bhatt | Towards Data Science, , https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292 (last visited Sep 17, 2021).

[339] James et al., *supra* note 13 at 15. In today's culture, machine learning can seem synonymous with magic. Machine learning is how a website can identify your face from a photograph, how streaming services predict the next show you'd like to watch, and even how cars

of almost brute-force experimentation.[340] In iteration after iteration, the model's reliance on predictive factors is slightly adjusted, and then the model's predictive performance on the training data is measured according to some criterion — often accuracy.[341] After testing many variations, the best performing model is chosen.

But optimization isn't everything. Machine learning has more to offer than just an optimized legal prediction — it can reveal the full range of predictions that the law allows.[342] Within many fields, machine learning is used not to optimize but to explore. This can also be done in law. Rather than generate a range of predictive models to arrive at one optimal version of the law, machine learning could be used to generate a range of predictive models that represent the scope of what is possible under the law. The advantage of machine learning is its capacity. With modern computing power, we can generate countless variations of a prediction. Any particular version of a predictive model represents just one way of making a prediction under the law. But taken in total, the many versions of the model could represent the range of possible ways of making a certain legal prediction. Researchers could use this collection of models to evaluate the results of all the ways that a law can be applied.

A graphical example can help illustrate the concept. Imagine a law that requires judges to consider at sentencing a convicted person's risk for committing a violent crime in the future. Under this law, judges must consider four factors when making this prediction: age, prior arrests, prior convictions, and prior length of imprisonment. The hypothetical true relationship of these four predictive factors with violent crime risk is depicted with the four graphs below:

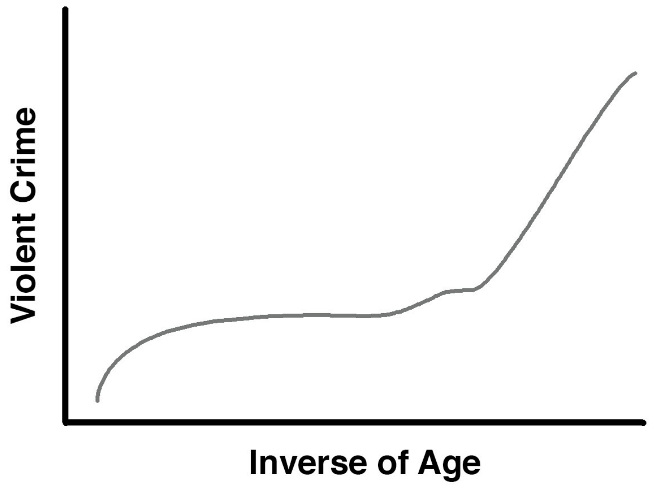
---

can learn to drive themselves. Although exceptional applications of machine learning capture public attention, most uses of this technology are more mundane, particularly in law.

340 *See* Andreas Stöckl, *Watching machine learning models fitting a curve!*, Medium (2021), https://towardsdatascience.com/watching-machine-learning-models-fitting-a-curve-c594fec4bbdb (last visited Aug 10, 2021).

341 Ethem Alpaydin, Machine Learning 38 (2016). An algorithms determines the process for exploring potential models. Some algorithms are exhaustive in their exploration of different models, while others seek to save time and computing power by exploring a more limited range of potential models.

342 This is how machine learning models can reveal information about prediction more generally even when a prediction in law is not clearly specified. For a specific prediction, an algorithm must have a clearly defined outcome of interest and clearly defined predictive factors — even when the law is vague about both. But the process of constructing a predictive model is much more expansive than making one specific prediction. Constructing a predictive model can include considering a multitude of possible variables, and the full range of weights that each of these variables could be given.

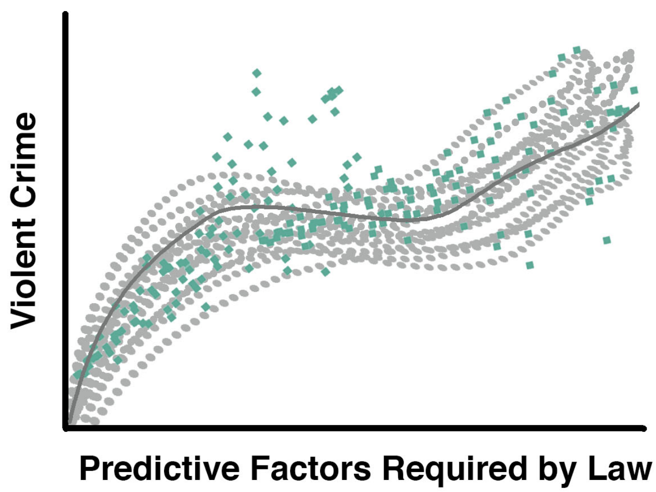


These four lines represent the hypothetical true function of how each of these factors correlate with violent crime. To simplify our example, we can reduce the dimensionality of our example and collapse those four graphs into one graph.

This x-axis on this graph represents all four factors. The line represents the hypothetical true function of how these combined four factors correlate with violent crime. In the real world, we don't know what the real relationship is between predictive factors and outcomes. Prediction is a process of estimating that hypothetical true function. Humans do this with intuition. To apply this hypothetical law, judges would rely upon their background and experience to estimate how likely a person is to commit a violent crime based on the factors the law requires them to consider. We can demonstrate in our graph three different judges' understanding of how these factors relate to recidivism.

These lines depict how different judges would weigh different factors differently. None of these predictions exactly matches the hypothetical true function, but each judge in this example would be faithfully applying the law and trying to capture the same real, unknown relationship. Statistical models can be used for the same task. Constructing a predictive model would require data on the predictive factors and case outcomes, depicted in the scatter plot below.

This graph shows a scatterplot superimposed over the line that represents the hypothetical true function of how each of these factors correlate with violent crime. Just as judges rely on experience to inform their predictions, a predictive model would use the available data to estimate the relationship between the factors and the outcome of interest. Optimizing a model means adjusting the shape of the model to better fit the data.[343] The conventional way of using machine learning to optimize legal prediction generates many models. Each is evaluated based on how well it performs. The model that best approximates the true function is selected and the rest are discarded.



The left-hand graph illustrates how the machine learning process generates many models. The right-hand graph pares down the many models so that only the one optimized model remains. Rather than select one optimized model, this Article suggests that we should linger with the left-hand graph and explore a range of models that together represent the range of predictions that are permissible under a particular law.[344]

---

[343] The choice of a particular model is often based on an estimation of the shape of the real function. Linear models will draw a straight line through the data. Logistic models bend. Flexible models make few assumptions about the shape of the real function. Data scientists commonly explore a variety of models to determine what works best for a particular dataset and prediction.

[344] There are many ways that this kind of exploration could be structured. A follow-up technical paper examines a series of options. *Statistical Methods for Exposing and Exploring Law with Algorithms* (on file). For the purposes of this Article, it's sufficient to show that this process *can* be structured. One way of setting the boundaries of this exploration would be 1) restrictting models to consider only the predictive factors allowed under law and 2) only include models that satisfy a performance criterion such as accuracy. The purpose of the performance criterion would be to limit the range of models considered to exclude models that are so inaccurate that they would represent a misapplication of the law.

In many cases, it will be difficult to decide exactly how poorly a model must perform to be a misapplication of the law. Even attempting to draw a line may appear to be forcing a quantitative response to a humanistic inquiry, demanding a level of statistical specificity that the law does not typically provide. Although the law often uses probabilistic language, such as probable cause, courts have long resisted formalizing these terms in a mathematical way.

Application: Pretrial Incarceration Laws

Algorithmic exposure should be adopted as a routine, best practice for diagnosing the efficacy and fairness of predictive laws. To date, courts, legislatures, advocacy groups, and lawyers have flown blind with predictive laws. We don't know the error rates these laws have, the outcomes they're capable of producing, or the relevancy of the predictive factors included in the law. And so it follows that we have not been able to discern the range of fair and unfair applications of these laws and have not been able to ascertain the legitimacy of legal judgments that have been justified by these predictions. Legal actors and institutions would benefit from knowing how a predictive law works and what outcomes it produces. And there are many ways in which legal, governmental, or private actors might be able to apply that knowledge.

This Part uses contemporary pretrial incarceration doctrine as a lens for exploring algorithmic exposure's potential within political advocacy, legislation, and agency policymaking. Advocacy groups, legislatures, and government agencies could repurpose legal algorithms to expose a legal doctrine's systemic potential and inform their practices of galvanizing political support, crafting legislation, setting policy, and litigating cases. Beyond pretrial doctrine and the applications examined in this part, there are more opportunities for algorithmic exposure worth exploring. This Part's goal is not to be exhaustive but to begin a discussion of how the method might be deployed in practice and what it might help achieve.

Pretrial incarceration is a prime example of a predictive legal doctrine ripe for algorithmic exposure. The decision of whether to jail someone pretrial is a weighty legal decision that is made routinely and systematically and is justified purely by a prediction about the future. Nearly every state legally allows judges to incarcerate some people pretrial based on upon predictions of dangerousness.[345] The justification for pretrial in-

---

Andrew Manuel Crespo, *Probable Cause Pluralism*, SSRN Electron. J. (2019), https://www.ssrn.com/abstract=3342902 (last visited Jul 29, 2021).

Fortunately, the decision over what range of models to exclude is unlikely to influence a global understanding of the law. Although the range of models that permissibly represent the law may be contested, the insights described in the following section can be derived from models that fall squarely in the permissible range — indeed largely from models that outperform humans making the same predictions. Thus, a precise determination of what models to exclude is unnecessary. Although the process described in this Article requires exploring a range of ways to apply the law, the poorest performing models are, by nature, the least helpful. The goal is to use machine learning to explore the limits of what a predictive law can accomplish: how accurate it can be, how its outcomes are distributed, and what predictive factors are most important. Poorly performing models represent the very margins of what is allowable under the law, not how the law ought to be applied. These models don't speak to the limits of predictive accuracy and can only provide redundant information about the strength and relevancy of predictive factors. Depending on the domain, poorly performing model may provide some information about potential distributions of outcomes when predictions are made poorly.

[345] KY R. Crim. P. 4.06. Most jurisdictions allow judges to incarcerate only people facing cer-

carceration is purely predictive and forward-looking, not based on the person's guilt in the current case. Looking to first principles of criminal law, people are jailed pretrial not because of retribution or just deserts but purely for the purpose of incapacitation, to prevent those people from harming the community at large.

Because the pretrial field is flush with datasets and algorithmic tools, pretrial laws are already ripe for algorithmic exposure. In the span of just a few years, pretrial algorithmic risk assessment tools have spread across the country to over 1,000 counties in all but four states.[346] If a pretrial risk assessment tool is already in use in a jurisdiction, then the existing models and data could jumpstart the process of algorithmic exposure.[347] Although dozens of different actuarial pretrial risk assessments exist, they are fairly uniform in design and loosely follow state law for pretrial incarceration predictions. Risk assessment tools use historical court and police data to label a particular defendant as low-to-high risk based on the rate at which people with similar characteristics were arrested or missed court dates while on pretrial release.[348] To make these pre-

---

tain, serious violent charges. *E.g.*, Cal. Const. Art. I, § 12. In some states, judges can incarcerate someone only to prevent grave physical harm. *Id*. Defendants awaiting trial in a criminal case may be released on personal recognizance, released on certain conditions, or detained in jail. For a broader background on general bail practices, see generally Criminal Justice Policy Program, Harvard Law Sch., Moving Beyond Money: A Primer on Bail Reform (2016). Most defendants are ordered to be released pending trial. *E.g.*, Cal. Penal Code § 1270 (West 2017) ("Any person who has been arrested for, or charged with, an offense other than a capital offense may be released on his or her own recognizance by a court or magistrate who could release a defendant from custody upon the defendant giving bail."). A person released on recognizance promises to return for future court dates. A person conditionally released must fulfill additional requirements such as posting a money bond, checking in with a pretrial services agency, maintaining employment, staying away from the victim or witnesses, or refraining from using alcohol or drugs.

346 Some states, like New Jersey, have adopted a uniform risk assessment for every court in the state. Public Safety Assessment New Jersey Risk Factor Definitions, (2018), https://njcourts.gov/courts/assets/criminal/psariskfactor.pdf?c=99i (last visited Aug 14, 2020). But this is the exception. The decision to adopt pretrial risk assessments is more often made at the county level, which can result in a patchwork of risk assessments in use across a single state. In California alone, over a dozen different pretrial risk assessment tools are used, while some counties do not use pretrial risk assessments at all. National Landscape, , Mapping Pretrial Injustice (2020), https://pretrialrisk.com/national-landscape/ (last visited Aug 13, 2020).

347 But in many cases, neither the model nor the dataset are available as the third-party developers who create these tools typically assert trade secret protections to shield the underlying data and code from public view. Most risk assessment models would need to be adjusted to more closely match state law's preventive incarceration requirements. Doyle, Bains, and Hopkins, *supra* note 48.

348 Public Safety Assessment- Risk Factors and Formula, . Consider a modern risk assessment tool like the Public Safety Assessment: Every defendant receives a "new criminal activity" risk score between 1 (lowest risk) and 6 (highest risk) and is either flagged or not flagged for "new violent criminal activity." *Id*. Most pretrial risk assessments also recommend that a judge incarcerate or release a person based on these scores. After calculating a person's risk

dictions, the tools' developers build statistical models based on factors that correlate with these outcomes.[349]

Pretrial incarceration laws are also an apt example because this is a unique historical moment for criminal law reform and pretrial reform in particular. There is now broad political support to end mass incarceration, including mass pretrial incarceration. The pretrial incarceration levels in this country are staggering: On any given day, American jails imprison nearly half a million people who have not been convicted of a crime.[350] The astonishing truth is that there are more legally innocent people behind bars in America today than there were convicted people in jails and prisons in 1980.[351] Across the country, increases in pretrial incarceration rates are "responsible for all of the net jail growth in last twenty years."[352] With just over 4% of the world's population, the United States has almost 20% of the world's pretrial jail population.[353] Pretrial reform has attracted the support of the media, politicians of both parties, professional organi-

---

score, that score is filtered through a "decision-making framework" or "decision-making matrix" that encourages judges to lock up people with high risk scores and free most of the rest. Pretrial Release Recommendation Decision Making Framework (DMF), (2018).

349 These characteristics often include age, history of arrest, history of convictions, and time spent in jail or prison. Megan T Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and The Double-edged Sword of Youth*, 96 Wash. Law Rev. 1 (2018). Some tools consider only a person's age and criminal history. For example, the Laura and John Arnold Foundation's Public Safety Assessment (PSA) looks to nine risk factors that include age, a defendant's criminal history, and a defendant's history of missed court appearances. Laura and John Arnold Found., Public Safety Assessment: Risk Factors and Formula 2 (2016). The full set of factors includes age at current arrest, current violent offense, pending charges, prior misdemeanor conviction, prior felony conviction, prior violent conviction, prior failure to appear in past two years, prior failure to appear longer than two years ago, and prior sentence to incarceration. *Id*. Based on these factors, the PSA ranks the person on a six-point scale from low to high risk for two pretrial risks, "failure to appear" and "new criminal activity." *Id*. at 3. Other tools are more eclectic and include personal information such as owning a cellphone or renting, rather than owning, a home. The Colorado Pretrial Assessment Tool Revised Report, (2012).

350 Todd D. Minton & Zhen Zeng, Bureau of Justice Statistics, Jail Inmates at Midyear 2014, at 1 (2015), https://www.bjs.gov/content/pub/pdf/jim14.pdf; Roy Walmsley, World Pre-trial/Remand Imprisonment List 1 (3d. ed. 2016).

351 Prisoners in 1980, (1981), https://www.bjs.gov/content/pub/pdf/p80.pdf (last visited Aug 13, 2020).

352 Peter Wagner & Wendy Sawyer, Prison Policy Initiative, *Mass Incarceration: The Whole Pie 2018* (Mar. 14, 2018), https://www.prisonpolicy.org/reports/pie2018.html.

353 Walmsley, *supra* note 115, at 13; Michelle Ye Hee Lee, *Does the United States Really Have 5 Percent of the World's Population and One Quarter of the World's Prisoners*, Wash. Post (Apr. 30, 2015), https://www.washingtonpost.com/news/fact-checker/wp/2015/04/30/does-the-united-states-really-have-five-percent-of-worlds-population-and-one-quarter-of-the-worlds-prisoners/?utm_term=.d10281e3c39c.

zations, and the public at large.[354]

And yet this opportunity to transform our pretrial legal systems may not be realized. The typical approach for pretrial reform has been to facilitate a better application of pretrial incarceration doctrine — not to reexamine or change the doctrine itself. The bail reform movement has implicitly assumed, at its peril, that the problem with pretrial incarceration is with how the doctrine is applied, not with the doctrine itself. This can be seen with the two hallmarks of contemporary bail reform: procedural protections and algorithmic risk assessments. The purpose of procedural protections is to ensure that incarceration decisions are not made hastily but after a thorough adversarial process. And the purpose of risk assessments is to optimize incarceration decisions by helping judges make more accurate, consistent predictions of dangerousness.[355] These reforms are both designed to help judges apply the law better. To be fair, all things being equal, procedural protections and risk assessments ought to produce better outcomes than the status quo. But all other things don't need to be equal. Reducing the scope of legal change to methods for optimizing pretrial doctrine ignores the possibility that pretrial incarceration doctrine itself should change.

The current crisis in mass pretrial incarceration and the groundswell of support for bail reform and racial justice should invite a closer critique of pretrial incarceration doctrine itself. While it may seem self-evident that people should be incarcerated pretrial based on predictions of dangerousness, the doctrine is of recent vintage and its moral and legal legitimacy has been hotly contested over time. Pretrial incarceration based on dangerousness assessments, a policy first proposed as legislation by the Nixon administration,[356] swept the country in the 1970s and 80s. It permitted courts, for the first time in American history, to legally jail people awaiting trial based on a public safety rationale.[357] Nixon's approach turned away from the conclusions of the Johnson administration's seminal 1967 report "The Challenge of Crime in a Free Society," which had

---

354 Doyle, Bains, and Hopkins, *supra* note 48 at 11–12.

355 These algorithmic tools should outperform judges at making predictions. Predictive models driven by millions of data points can more accurately predict recidivism than a judge making quick judgments based on limited information. Laura and John Arnold Foundation, Research Summary: Developing a National Model for Pretrial Risk Assessment 2 (2013). If a jurisdiction uniformly adopts these tools, then pretrial decisions could be more consistent and less influenced by the whims or prejudices of individual judges. And if the risk assessment tools contain or reflect bias along racial, class, gender, or other lines, then the tools have some potential to be analyzed and adjusted — unlike judges, whose biases remain hidden within the inaccessible "black box" of their minds. Pretrial risk assessments have become a divisive issue, and the critiques of these tools are varied. For a helpful summary of a variety of these critiques, see Sarah L Desmarais & Evan M Lowder, Pretrial Risk Assessment Tools 12.

356 Guy M. Blynn, *Pre-Trial Detention*, New York Times, February 9, 1969, https://www.nytimes.com/1969/02/09/archives/pretrial-detention.html (last visited Aug 10, 2021).

357 Kellen Funk, *The Present Crisis in American Bail*, Yale Law J. Forum 1098, 1104–05 (2019).

considered pretrial detention as way to reform bail but had determined that it "might well create more of a problem than the imposition of money bail, in the light of the difficulty of predicting dangerousness."[358] Legal scholars, including Laurence Tribe, Caleb Foote, and Ronald Dworkin, opposed preventive pretrial incarceration as violative of due process and human dignity.[359] In a 1987 opinion, *United States v. Salerno*, the Supreme Court found the federal government's new preventive detention scheme to be constitutionally permissible.[360] In dissent, Justice Thurgood Marshall chastised the court for "disregard[ing] basic principles of justice."[361] He warned of "the coercive power of authority to imprison upon prediction" and "the dangers which the almost inevitable abuses pose to the cherished liberties of a free society."[362] In the years following *Salerno*, the debate over preventive pretrial incarceration has petered out.[363] Algorithmic exposure may be one avenue for reigniting this debate and recentering the legitimacy of preventive pretrial incarceration doctrine within broader discussions of pretrial reform.

This Part explores a variety of practical ways that legal actors and institutions might use algorithmic exposure to challenge and transform pretrial incarceration doctrine. By using algorithms to reveal the systemic limits of contemporary pretrial laws, advocacy groups could refine their policy positions and galvanize the public. Likewise, legislatures could use algorithms as a diagnostic tool for assessing and calibrating potential pretrial reforms. Independent of statutory changes, prosecutors could use algorithmic analysis of pretrial doctrine to justify less carceral policies — both in court and to the public.

---

[358] The Challenge of Crime in a Free Society: A Report by the President's Commission on Law Enforcement and Administration of Justice., 98 (1967).

[359] Laurence H Tribe, *An Ounce of Detention: Preventive Justice in The World of John Mitchell*, 56 Va. Law Rev. 38 (1970); Caleb Foote, *Comments on Preventive Detention*, J. Leg. Educ. 9.[Insert Dworkin from book].

[360] United States v. Salerno, 481 U.S. 741 (1987).

[361] *Salerno* 481 U.S. at 755.

[362] *Id.* at 766–67.

[363] Although some scholars have kept the torch burning. *See* Robin Steinberg, *Freedom Should Be Free: A Brief History of Bail Funds in the United States* 19; Hegreness, *America's Fundamental and Vanishing Right to Bail*; Shima Baradaran, *Restoring the Presumption of Innocence*, 72 55; Jeff Thaler, *Punishing The Innocent: The Need For Due Process And The Presumption Of Innocence Prior To Trial*, Wis. Law Rev. 45; Albert W Alschuler, *Preventive Pretrial Detention and The Failure of Interest-balancing Approaches to Due Process* 61.

Advocacy

Algorithmic exposure can provide advocacy groups with a better understanding of how a predictive law works and what outcomes it produces. Advocacy groups can use these empirical insights to clarify their policy positions, educate the public, and galvanize support for their cause. With money bail and pretrial reform movements, algorithmic exposure can help orient advocacy away from purely procedural reforms and toward substantive doctrinal reform.

Clarifying Policy Positions. Advocacy groups can use algorithmic exposure to clarify their policy positions on pretrial reform by discerning how current pretrial incarceration laws accord with their values. Many advocacy groups tacitly accept pretrial incarceration justified by predictions of dangerousness as the legal backdrop for reform.[364] Instead of targeting the substantive doctrine of pretrial incarceration, these groups have focused their policy guidance and advocacy on promoting procedural protections for pretrial incarceration decisions.[365] These procedural protections are important. Rather than incarcerate people pretrial through imposing unaffordable money bond amounts at a perfunctory hearing, jurisdictions should have robust hearings in which the prosecution must prove its case and defense counsel can cross-examine witnesses and present evidence.[366] But the current trend in bail reform is strictly procedural.

---

364 Some groups have already taken a strong stance to oppose or support pretrial incarceration based on generalized predictions of dangerousness. Many community bail funds take an explicitly abolitionist stance that opposes incarceration on any grounds.

365 Doyle, Bains, and Hopkins, *supra* note 48.

366 The standard way that people are incarcerated pretrial is still money bail. The study of Cook County, Illinois in the appendix is an example of this problem. The county has robust preventive detention procedures and requires judges to determine that a defendant has the ability "to pay the amount necessary to secure his release" before imposing bail. General Order No. 18.8A, Circuit Court of Cook County, (July 17, 2017), http://www.cookcountycourt.org/Portals/0/Orders/General%20Order%20No.%2018.8a.pdf. But judges in Cook County continue to impose unaffordable bond amounts that result in detention. The Coalition to End Money Bail, Shifting Sands: An Investigation Into The First Year of Bond Reform in Cook County 6 (2018), https://www.chicagobond.org/reports/ShiftingSands.pdf (last visited Sep 20, 2018). At an initial hearing following arrest, a judge — or often a magistrate — can condition a person's release from jail upon the payment of a certain bond amount. Doyle, Bains, and Hopkins, *supra* note 48. Those who can afford bond, or who can at least afford a bail bond company's fee, leave jail. Those who can't pay are locked up until their case is over. In the process, people may be detained pretrial without having an adequate hearing and without having been represented by defense counsel. In recent years, a series of civil rights lawsuits have challenged bail practices in states across the country, and courts have tended to find these practices to be unconstitutional. Funk, *supra* note 122 at 1111. These cases have resulted in federal courts issuing preliminary injunctions against local governments and have prompted many states and counties to consider changing their laws and policies. Civil Rights Corps, *Challenging the Money Bail System* (2020), https://www.civilrightscorps.org/work/wealth-based-detention (last visited Aug 13, 2020).

Advocacy groups ought to expand their focus to reckon with underlying pretrial incarceration doctrine. Algorithmic exposure can help, by showing how a pretrial incarceration law operates and what effects it produces — and can produce. This process would enable advocacy groups to better evaluate how a law could conform with their values — or whether their values and a legal doctrine are incompatible. There's reason to think that algorithmic exposure could help some advocacy groups reconcile a particular tension in their current policy platforms: their opposition to pretrial risk assessment tools and their acceptance of preventive incarceration on predictions of general dangerousness as a background legal doctrine. A core value for progressive advocacy groups is racial justice. A commitment to this principle has led advocacy groups to critique and oppose pretrial risk assessment algorithms because of their racial bias and poor accuracy.[367] But they have not been so outspoken about the legal doctrine that these algorithms apply. Algorithmic exposure can show how their emphasis may be misplaced. The racial inequities and inaccuracy attributed to risk assessment tools are problems inherent to the legal doctrine that these algorithms apply.[368]

By revealing the distribution of outcomes possible under pretrial incarceration laws, algorithms could reveal if a jurisdiction's pretrial incarceration laws result in the disproportionately incarceration of Black people on weak predictions of future violence.[369] There's reason to suspect that in most places the distribution of both adverse outcomes and errors would disproportionately falls upon Black people[370] This consequence is a mathematical necessity given how preventive incarceration laws function and given what demographic groups tend to have criminal records. Using data from pretrial risk assessment algorithms, researchers have already found that disparate error rates will occur in any predictive system in which different racial groups had different baseline rates of being arrested.[371] It will occur without any personal bias or discriminatory intent on

[367] More than 100 civil rights and community groups, including the ACLU and the NAACP, have signed a statement opposing pretrial risk assessment tools. The Leadership Conference on Civil and Human Rights, *supra* note 9. Scholars have likewise written open letters detailing the harms that these tools perpetuate and encouraging jurisdictions to consider alternative pretrial reforms. Barabas, Dinakar, and Doyle, *supra* note 51.

[368] In fact, risk assessments may be the least biased and most accurate application of the doctrine.

[369] In practice there should be some local variation. Pretrial incarceration laws, judicial practices, and criminal laws all differ — not to mention differences in the population at large, including rates of criminal offending, socioeconomic circumstances, racial composition, and more.

[370] In the wake of ProPublica's *Machine Bias* report, Angwin et al., *supra* note 5. the distribution of outcomes for pretrial incarceration decisions was thoroughly examined in the literature. See Sandra G. Mayson, *supra* note 42 at 2333–38 (collecting sources and explaining the phenomenon).

[371] Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Berk et al., *supra* note 45. Absent an intervention that explicitly adjusted predictions based on race. Kleinberg, Mullainathan, and Raghavan, *supra* note 8.

the part of people or machines that apply these laws.[372] Higher rates of prior arrest within a subset of the population will lead to higher rates of pretrial incarceration.[373] Predictions of dangerousness will have more false positives for the group that has more positive outcomes and will have more false negatives for the group that has more negative outcomes.[374] In practice, this means that the distribution of errors for Black defendants will be different than the distribution of errors for white defendants. When mistakes are made with dangerousness predictions, Black peoples' dangerousness would tend to be overestimated and white peoples' dangerousness would tend to be underestimated.[375]

When applying pretrial incarceration laws, these mistaken predictions are common because predictions of future violence are highly speculative. Largely because pretrial violence is so rare, it is virtually impossible for judges or any statistical model to identify people who are more likely than not to commit a violent crime. Predicting the future is always difficult but is nearly impossible when predicting very rare, interpersonal events, like violence, within a short timeframe, like the pretrial period. Current pretrial risk assessments data can again be a useful demonstration.[376] With the leading tool on the market, the Public Safety Assessment, every defendant is flagged or not flagged for "new violent criminal activity." Within the tool's training data, of those who were flagged for "new violent criminal activity," only 7% went on to commit a violent crime on pretrial release.[377] In other words, the "new violent criminal activity" flag can be expected to correctly identify who will commit violence about 7% of the time and can be expected to incorrectly identify who will commit violence about 93% of the time.[378]

---

372 *Id.*

373 To put it another way, groups of people that have more members with criminal histories are more likely to be preventively incarcerated. *Id.*

374 In essence, to create a predictive algorithm with equal error rates for different groups, unless the algorithm may require explicitly considers and adjusts for group differences.

375 Angwin et al., *supra* note 5.

376 As a reminder, because these tools outperform humans at predicting, their limits represent the limits of our ability to make violence predictions. *See supra* Part II. B–C.

377 14B Measuring and Managing Pretrial Risk with the Public Safety Assessment: Assessor Training.

378 This pattern has proven to be fairly consistent in places where the PSA has been adopted, despite regional variations in pretrial incarceration rates, policing practices, crime rates, and more. Over different time periods, the percent of people flagged for NVCA who go on to be arrested for a violent crime on pretrial release has been 14% in New Jersey, 3% in Kentucky, and only 1% in Cook County, Illinois, which includes Chicago. Ethan Corey, *How a Tool to Help Judges May Be Leading Them Astray*, The Appeal (2019), https://theappeal.org/how-a-tool-to-help-judges-may-be-leading-them-astray/ (last visited Jan 8, 2021). Other studies of risk assessments have found similarly low rates. Megan T. Stevenson & Sandra G. Mayson, *Pretrial Detention and the Value of Liberty*, 39 (collecting studies).

Risk assessments make errors when predicting who will not commit violence, too. Risk assessments do not flag the overwhelming majority of people who go on to commit violence on pretrial release.[379]

Algorithmic exposure of the range of possible outcomes of pretrial laws could prompt advocacy groups to reconcile their policy stances on risk assessment tools and pretrial incarceration doctrine. If values of racial justice have led these organizations to critique and oppose risk assessments, these same values ought to the lead them to criticize pretrial incarceration laws that these algorithms apply.[380] Algorithmic exposure may also help advocacy groups evaluate pretrial incarceration laws' impact on other marginalized groups. To date, Black and white populations have been the primary subjects of empirical research on this question. Algorithmic exposure could reveal more details about how other population subgroups are treated under the law. So long as the dataset being used is labeled for characteristics of the population — age, race, gender identity, ethnicity, and so on — this analysis is possible.

---

[379] There is little reason to expect the quality of these predictions to improve. Recent decades have seen only slight improvement in our ability to predict violence, among any populations. Although pretrial risk assessments have only recently come to prominence, these tools are an offshoot of a broader field of risk prediction in the social sciences, particularly psychology, which has developed a rich literature on the subject. A leading meta-analysis of risk assessment tools concludes that "the ceiling of predictive efficacy may have been reached with the available technology." Min Yang, Stephen C. P. Wong & Jeremy Coid, *The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools.*, 136 Psychol. Bull. 740–767 (2010). Rare, interpersonal events are among the hardest to predict, and historical background information about people can only tell us so much about what they will do in the future. Given these limits, the authors conclude that risk assessments, "should not be used as the sole or primary means for clinical or criminal justice decision making that is contingent on a high level of predictive accuracy, such as preventive detention." Statistical models can modestly outperform unaided human prediction for predicting arrest and arrest for a violent offense. A meta-analysis of over half a century of psychology research concludes that "statistical prediction methods are, in general, more accurate than clinical prediction methods." (Ægisdóttir et al., 2006). But statistical methods are no crystal ball. This same meta-analysis acknowledges that these tools are, in general, only slightly more accurate than humans. For pretrial risk assessments, a recent empirical paper arrives at the opposite conclusion. Julia Dressel & Hany Farid, *The accuracy, fairness, and limits of predicting recidivism*, 4 SCI. ADV. eaao5580 (2018). Within a study framework that gave people specific factors to consider and provided feedback on the accuracy of their predictions, human subjects recruited on the internet were able to slightly outperform the COMPAS pretrial risk assessment tool. Some have criticized the study's framework for being too divorced from real-world circumstances, given that judges have more information in front of them and rarely know whether their predictions were correct or not. Goel et al., supra note 12. Perhaps the problem is not so much with the study design but with a legal system that doesn't provide judges feedback on their predictions. But whether this criticism is valid or not, the point remains that statistical tools are, at best, a modest improvement over human prediction.

[380] As a policy matter, heightened procedural protections still matter, but they can't bring pretrial incarceration doctrine in line with these groups' broader values of racial justice.

Reconciling this policy position, should also help to resolve internecine conflicts among advocates and researchers with shared goals. Too often, progressive advocates and sympathetic researchers are pitted on opposite sides of the algorithmic debate. Pretrial risk assessments are a case in point. Advocates oppose algorithmic risk assessments for the harm they reveal. Researchers support algorithmic risk assessments because they're superior to human prediction. Neither side gets it quite right.[381] The choice between harmful algorithms or worse humans is a false dilemma. If an algorithm seems to be causing harm but recalibrating the algorithm cannot fix that harm, then it is likely that the algorithm is not causing this harm but is revealing harm inherent to the law it applies. When a legal doctrine is the source of problems, the concern should be transforming the law — not choosing whether algorithms or humans should apply that law.[382]

In this way, algorithmic exposure can help to expand the scope of evidence-based or data-driven law reform. Technology and data are potent tools for political advocacy, particularly in a time of data-driven public policy. In the domain of criminal law, evidence-based criminal law reforms reign supreme. But scholars have noticed a troubling contradiction.[383] The shift to evidence-based policy standards now requires that any changes to the criminal law be justified through an empirical or data-driven lens. But existing criminal laws are not evidence-based. In fact, most of the laws that have driven mass incarceration in recent decades were adopted despite empirical evidence to the contrary.[384] Algorithmic exposure can provide a way to expand the scope of the "evidence-based" paradigm beyond tweaks or optimizations of existing practices to reckon with fundamental values.

Educating and Galvanizing the Public. These empirical insights aren't just useful for an organization's internal decision-making; advocacy groups can use this same information to educate and galvanize the public to pressure elected officials to change the law.

---

[381] As explored in detail in Part I, when humans are making biased and error-prone predictions, algorithms are an improvement — all other things being equal. But that phrase, "all things being equal," does a lot more work than supporters of algorithms let on. In politics and law, things are rarely in stasis — and all things never stay equal in the zero-sum game of limited political will, public engagement, and financial resources. Algorithmic reforms to law necessarily come at the expense of other legal reforms. For those who desire fundamental changes to our criminal legal systems, it's rational to consider algorithms a distraction that siphons away attention and resources from other opportunities. The conflict within the algorithmic fairness discourse between proponents and critics of algorithmic interventions ought to be understand against this political economic backdrop.

[382] To put it another way: When prediction can justify a legal process, algorithms are preferable to unaided human judgment. But algorithms can also reveal when prediction does not justify a legal process.

[383] Erin Collins, Shifting the Evidence-Based Paradigm (forthcoming) (on file with the author).

[384] Michael Tonry, *Predictions of Dangerousness in Sentencing: Déjà Vu All Over Again*, 48 Crime Justice 439–482 (2019).

Insights concerning accuracy and racialized distribution of outcomes can help justify the argument that pretrial incarceration based exclusively on predictions of violence is incompatible with either a commitment to racial justice or a commitment to due process of law.[385] Activists' similar arguments against algorithmic risk assessment tools have already proven politically effective.[386]

Insights into the accuracy and racial distribution of dangerousness predictions can be a powerful public education tool. The predictive turn in the criminal law has become deeply rooted within American culture. There is a sense that violence is predictable and that judges should be able to discern who should be incarcerated and who should be released pretrial. Particularly when machine learning algorithms are used, the public may overestimate the potential accuracy of these predictions. Although the phrase "preventive detention" conjures up ideas of rationality and science, the data collected here can expose the public to the fallout of these practices. The actual numbers are humbling, if not shocking.[387] Preventive incarceration that depends upon these predictions defies one of the central maxims in our legal tradition: William Blackstone's "[B]etter that ten guilty persons escape, than that one innocent suffer."[388] Pretrial incarceration based on these predictions generates — in some cases precisely — the opposite ratio: ten people must be incarcerated to prevent one from committing a violent crime on pretrial release. Recent empirical research indicates that the public may not approve of pretrial incarceration at this poor a level of accuracy.[389] The question of the legitimacy of pretrial incarceration has sometimes been framed as, "What level of accuracy for predictions of future violence is needed to justifiably incarcerate someone who has not been found guilty of a crime?" One potential answer is that whatever that level may be, it's not in the ballpark of what's achievable.

---

385 These insights invite other considerations as well. How much is pretrial incarceration justified by factors beyond someone's control? Given that age is the most predictive factor for violence, how does pretrial incarceration conflict with other values [insert commitment to rehabilitation / opportunities for kids]. How much does prediction create a feedback loop of prediction and punishment rather than open up opportunities for positive change?

386 Peter Krouse, *Ohio Supreme Court proposes bail reforms that don't include risk assessments*, cleveland.com (2020), https://www.cleveland.com/news/2020/01/ohio-supreme-court-proposes-bail-reforms-that-dont-include-risk-assessments.html (last visited Aug 14, 2020); Final Report of the Special Commission to Evaluate Policies and Procedures Related to the Current Bail System, (2019), https://d279m997dpfwgl.cloudfront.net/wp/2020/01/0102_bail-reform-report.pdf (last visited Jan 29, 2020).

387 This information may not be so new or shocking to impacted communities. In this way, the data can reinforce or corroborate community knowledge. There is a risk in using empirical data in this way: the risk that a sufficient condition could become a necessary one. Many groups, including bail funds, have argued that we shouldn't need empirical data to prove what we already know through the lived experiences of communities that are policed and punished.

388 Alexander Volokh, *n Guilty Men*, 174 Univ. Pa. Law Rev. (1997).

389 Stevenson and Mayson, *supra* note 139.

Accuracy and racial inequity are not separate issues but are deeply interconnected. Incarcerating people upon weak justifications may only be possible when the people being targeted are already marginalized. As Caleb Foote remarked decades ago, preventive detention's function is to be "an acceptable public-relations rationale" for "conceal[ing] and perpetuat[ing] a discriminatory system of justice in the face of growing unwillingness from those difficult lower classes to continue to be treated more as objects than as humans."[390] Prediction has been a way to sweep discriminatory treatment and replication of systemic inequality under the rug. As descriptive machines that reproduce what they've been fed, algorithms can lay bare the structural inequities have been reproduced within seemingly objective predictions.[391]

Legislation

Algorithmic exposure of law can allow for a more informed, nuanced approach to legislation and rulemaking. Society does not need to blindly construct predictive laws and discover the consequences later. Legislatures and administrative agencies could subject predictive laws and regulations to algorithmic analysis to ensure that the laws are able to achieve their goals and are producing fair outcomes. The process of writing these laws could aided by predictive algorithms from start to finish.[392]

The legislative processes explored in this section illustrates how algorithmic exposure can enrich the public and legislative debate over predictive laws by demystifying prediction and prevention, particularly within the domain of criminal law. By breaking a prediction down to its constitutive elements, algorithms can reveal the contingent, political nature of a predictive law. Predictive laws have often been characterized as an objective forecast coupled to a neutral cost-benefit analysis,[393] But a predictive-legal process cannot be objective because defining the system to be analyzed and the variables to be considered is always a normative, political choice — not a purely quantitative one.[394] As Ruha Benjamin reminds us, "even just deciding *what problem* needs

---

[390] Foote, *supra* note 124.

[391] For a discussion of the tension between empirical knowledge gained through algorithms and community knowledge gained through lived experience, see infra Part IV B. Knowledge and Values.

[392] Insights gained through the algorithmic exposure of law do not then require that the predictive law be administered by algorithms. There's no iron rule that software must be integrated into the legal process. A law could be made simpler for the sake of honest and transparency while maintaining predictive accuracy.

[393] Harcourt, *supra* note 20.

[394] *See Id.* Bernard Harcourt identifies this objectivity trap as the systems fallacy and traces its origins to the development of operations research during World War II and the rise of systems analysis in the mid-twentieth century.

solving requires a host of judgments."[395] Questions of scope, factors worth considering, and even objectives worth prioritizing are not self-evident.[396] Algorithmic exposure lays bare these considerations by showing what factors and outcomes of interest were or could have been considered in the predictive process.

Legislatures can incorporate algorithmic exposure into the legislative process as a kind of impact assessment for predictive laws. Impact assessments are nothing new. The legislative processes for state and federal governments often require impact assessments, including those that measure the environmental,[397] economic,[398] or racial[399] impact of proposed regulations and laws. A predictive impact assessment could report to the legislature a predictive law's accuracy, its distribution of outcomes and errors across different population groups, and the relevancy and strength of the predictive factors included in the law. Assessments could be performed on competing versions of proposed legislation so that legislators and the public could compare their respective performance.

To show how predictive impact statements could be useful — and not just another bureaucratic report among many produced in a legislative session — let's consider what an impact report might return about pretrial incarceration law's predictive factors and how a legislature might respond. Current pretrial incarceration statutes include many factors, few of which have been vetted to see if they correlate with pretrial violence or recidivism.[400] The report may find that some factors required by law do not have predictive value.[401] An impact report might also find that some factors have an outsized influence. Even if a pretrial incarceration statute lists four factors, it's unlikely that each

---

[395] Ruha Benjamin, Race after technology: abolitionist tools for the new Jim code 11 (2019). Insert reference to point brought up in Ben Green's LPE blog post.

[396] Harcourt, *supra* note 20.

[397] Biosafety Unit, *What is Impact Assessment?* (2010), https://www.cbd.int/impact/whatis.shtml (last visited Aug 11, 2021).

[398] Economic Impact Statement Act - American Legislative Exchange Council, , https://www.alec.org/model-policy/economic-impact-statement-act/ (last visited Oct 21, 2021).

[399] Racial Impact Statements, , The Sentencing Project , https://www.sentencingproject.org/publications/racial-impact-statements/ (last visited Oct 21, 2021).

[400] A slightly more complicated issue is that there can be some factors that have only slight predictive value. These factors do not dominate any statistical model, but they are not completely irrelevant on their own. They bear some statistical relationship to arrest for a violent crime. At the level of building a statistical model, the problem with these factors is that they tend to drag in more noise than signal when included in the model. That is to say: by including these factors in the model, the model becomes less accurate, even though those factors are predictive. The problem is that their relationship with the outcome isn't clean enough. We could make better predictions by considering fewer factors that have a stronger relationship with the outcome.

[401] This concern is not hypothetical. The factors used within many predictive laws were not derived from a rigorous process. A common way of drafting these laws is BOGSAT: a bunch of guys sitting around a table.

of the four factors is responsible for 25% of a prediction of future violence. Some factors will be stronger indicators of violence, and some will be weaker. Knowing which factors are most important may shape public and policymaker attitudes toward prediction, because these factors may compete with other values.

For example, policymakers may reconsider how to structure preventive incarceration legislation if a predictive impact report reveals that the law instructs judges to lock people up mostly because they are young. From a purely predictive standpoint, age is a helpful factor to include in a preventive incarceration law, as age is often the strongest factor when predicting recidivism risk.[402] In other words, compared to all other information, age is the best indication of whether someone will be rearrested for a violent crime on pretrial release. Although age statistically correlates with crime, other values might override the legislature's goal of maximizing the accuracy of preventive incarceration. Perhaps legislators want the law to reflect the importance of giving young people second chances or the overall benefit of prioritizing rehabilitation over incarceration given the plasticity of juvenile brain development.[403] If that were the case, a legislature may choose to modify a pretrial incarceration law that only considers recidivism risk to make age-related exceptions or to include alternative procedures for young people. The benefit of the predictive impact assessment is that it can alert the legislature to how a seemingly neutral law conflicts with other values and goals.

Algorithmic exposure could also be incorporated iteratively into the legislative drafting process. Algorithms could be used to experiment with draft legislation and test a variety of models to see which best adhere to legislative goals and priorities.[404] In this way, algorithmic exposure could be used not just to condone or condemn a law but to explore how it might be reshaped. Formulas for prediction could be tested and modified to capture a subset of models that perform more efficiently or better reflect moral or legal values.

One example of this problem is bail statutes. Following the first set of bail setting formulas pioneered by the Vera Institute in the 1960's, many states adopted laws that listed factors that judges ought to consider when setting bail, under the impression that these factors were predictive of future court appearance. The process of building pretrial risk assessment algorithms has already shown that many of these factors have a weak or non-existent relationship with future court appearance.

[402] *See* Stevenson and Slobogin, *supra* note 114 at 1–2.

[403] Megan Stevenson and Christopher Slobogin have more deeply explored the question of risk assessment policy on this issue. Megan T Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 26.

[404] In some circumstances, changes to how predictions are made could mitigate many distributional harms. If the errors in prediction fall upon certain groups rather than others, the law may violate principles of distributional fairness. Algorithmic exposure can reveal not just a particular instantiation of a legal doctrine or theory but the whole range of possibilities. If a particular range of outcomes are impermissible, an algorithm can reveal which applications of a legal doctrine should be foreclosed to prevent these outcomes.

Among other possible uses, this iterative process can be a way for legislatures to work through difficult decisions about predictive laws' dependence upon biased data.[405] The public and public officials today are concerned with biased data distorting predictions of risk. Whether a prediction of pretrial dangerousness is made by a judge or an algorithm, these predictions use data that reflect patterns of policing and prosecution. Decades of research have shown that, for the same conduct, Black and Hispanic people are more likely to be arrested, prosecuted, convicted, and sentenced to harsher punishments than their white counterparts.[406] As the classic computer science aphorism goes: Garbage in, garbage out.[407] Predictions made using biased data will produce biased results.

So what is a legislature to do about biased predictions? One proposal is to get rid of predictive incarceration laws. Another proposal is to explicitly consider race when making recidivism predictions, as a kind of affirmative action for crime prediction.[408] Neither option seems politically palatable.

Algorithmic exposure can provide an opportunity to test other ways of reducing the effect of biased data over predictive decision-making.[409] One idea worth testing is that biased data plays an outsized role at the margins of pretrial decision-making. The logic is this: Judges and algorithms are much more likely to be affected by bias and make the wrong predictions for borderline cases. A Black person with a few arrests on their record may only have those arrests because of policing practices in their neighborhood.

---

[405] Gone are the days when advocates and public officials proclaimed legal algorithms' objectivity and neutrality. Concerns over biased data have gone mainstream. Biased data is the algorithmic fairness concern that has found its way from academia to newspapers, documentary films, and presidential campaigns.

[406] See generally The Sentencing Project, Report of the Sentencing Project to the United Nations Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia, and Related Intolerance Regarding Racial Disparities in the United States Criminal Justice System (2018); Lynn Langton & Matthew Durose, U.S. Dep't of Justice, Police Behavior During Traffic and Street Stops, 2011 (2013); Stephen Demuth & Darrell Steffensmeier, The Impact of Gender and Race-Ethnicity in the Pretrial Release Process, 51 Soc. Probs. 222 (2004); Jessica Eaglin & Danyelle Solomon, Brennan Center for Justice, Reducing Racial and Ethnic Disparities in Jails: Recommendations for Local Practice (2015); Sonja B. Starr & M. Marit Rehavi, Racial Disparity in Federal Criminal Sentences, J. Pol. Econ. 1320 (2014); Marc Mauer, Justice for All? Challenging Racial Disparities in the Criminal Justice System (2010).

[407] Or as Sandra Mayson recasts it, Bias In, Bias Out. Sandra G. Mayson, *supra* note 42.

[408] Kleinberg, Mullainathan, and Raghavan, *supra* note 8; Deborah Hellman, *Measuring Algorithmic Fairness*, 106 Va. Law Rev. 811 (2020).

[409] This approach would (at least implicitly) adopt a counterargument to the claim that predictive laws should be abolished because of biased data is that the data coming in and the predictions coming out are not all garbage. This is a normative claim more than a technical one. The normative question that is left open is, What level of disparate treatment of marginalized groups is acceptable within predictive laws?

A similarly risky white person would have a clean record, because they don't live in a heavily policed area. There would be a difference in the data that does not reflect a difference in reality. Relying on this data, judges and algorithms would overestimate the risk of the Black person and underestimate the risk of the white person. In contrast, when a person has a lengthy history of violent crimes, biased data should have less of an impact. That data can meaningfully indicate predilection toward violence, despite any noise from biased policing or prosecution practices.[410]

Under this conception of biased data's impact, the impact of racial bias could be reduced if the law required greater certainty of future violence before an order of pretrial incarceration can be imposed. Absent a method for testing the theory, a legislature — and the public — would have to accept or reject it at face value as a credible explanation of biased data's impact. If accepted, the legislature would also have to guess what level of predictive certainty would best reduce these disparities.

In contrast, algorithmic exposure provides an opportunity to prospectively explore a range of potential variations on a legal rule and evaluate the results. The following graph shows the results of a simple experiment that explore how racial disparities in error rates can change based on how certain a prediction of recidivism must be before it can justify a decision to label someone as "high risk." For this test, 100 different pretrial risk assessment models were constructed and tested on pretrial data.[411] The only variation between the models is the threshold at which they labeled someone as high risk. On the left side of the graph, everyone was labeled high risk. Moving from left to right, the models become more stringent on who deserves to be labeled high risk. On the right side, no one was labeled high risk. The y-axis measures the false positive rate — the frequency with which the model mistakenly labeled someone as high risk.[412] The unbroken green line tracks the false positive rate as a function of threshold for Black defendants, and the dotted blue line tracks the false positive rate as a function of threshold for white defendants.

The difference between the two lines is the difference in the rate at which Black people and white people were mistakenly labeled high risk. The lines converge at the points where everyone is labeled "high risk" and where everyone is labeled "not high risk," but they diverge in varying amounts along the path between these two points. At all points, Black people were more frequently mistakenly labeled "high risk" than white people. But the difference narrows considerably across certain ranges.

---

[410] In other words, bias in the data can distort some predictions but it does not scrub all predictions of all value.

[411] The data included both case outcomes and a risk assessment tool's predictive performance and was used in ProPublica's report on risk assessment bias. Angwin et al., *supra* note 5. The risk assessment model was built to mirror the COMPAS risk assessment tool.

[412] For purposes here, the definition of a mistaken label of high risk was accepted as-is from the dataset, and a person was not considered high risk if they were not arrested within a two-year window following the assessment.

What this graph reveals is that — at least for the sample data — racial disparities in error rates for predicting recidivism can be reduced by changing how certain a prediction of recidivism must be before someone is incarcerated. Choosing the appropriate level of certainty is a value-laden question dependent upon potentially competing ideas of racial equity and acceptable error rates.[413] Algorithmic exposure can't answer that normative question, but it can make the stakes clearer. Any legislature contemplating changes to its predictive incarceration laws ought to have this kind of information available rather than draft pretrial incarceration laws without knowing their possible effects.

Policy

Professional researchers within state or federal administrative agencies are already capable of using algorithms as a diagnostic tool for law. These are institutions of expert authority that have access to government data and already use statistical methods for different analyses.[414] In the past decade, a growing scholarly literature in criminal law has explored the opportunity for these agencies to usher in and oversee a new era of empirical, expert-driven criminal law reform.[415] State administrative agencies, administrative offices of the court, and executive offices are primed to deploy algorithmic exposure as a method.

In the context of pretrial incarceration doctrine, prosecutors are also capable of using algorithmic exposure to reshape their offices' pretrial policies. As a group, the new wave of "progressive prosecutors" seem more eager than traditional prosecutors to hire data scientists and other technical staff to collect data on prosecutorial behavior and conduct internal data audits.[416] Larry Krasner's office is collaborating with the University of Pennsylvania to create a public-facing data dashboard to share and analyze prosecutor decisions.[417] Smaller offices that don't command a national spotlight are also taking a data-driven approach to revising internal policy. In 2021, the District Attorney's office for Yolo County, California (located just outside Sacramento) created

[413] Reductions in racial disparities in this example will come at the cost of overall accuracy.

[414] See generally David Freeman Engstrom et al., Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies, SSRN Electron. J. (2020), https://www.ssrn.com/abstract=3551505 (last visited Aug 10, 2021).

[415] Rachel E. Barkow, Prisoners of politics: breaking the cycle of mass incarceration 15 (2019).

[416] Rachael Rollins's office in Boston and Larry Krasner's office in Philadelphia have fulltime staff studying the office's internal data.

[417] Philadelphia DAO, *District Attorney Krasner Announces DAO Transparency & Accountability Research Collaboration*, The Justice Wire (2020), https://medium.com/philadelphia-justice/district-attorney-krasner-announces-dao-transparency-accountability-research-collaboration-27b12ad833db (last visited Aug 9, 2021).

a transparency data portal for the public to track how the office handles cases.[418] This effort has already led to policy change. Data revealed that the District Attorney's office was disproportionately denying people of color the opportunity for their cases to be diverted out of the criminal legal system before trial, largely because having a criminal history could automatically disqualify someone from diversion. The office has eliminated automatic disqualification, and the data portal will continue to track the results.[419] In offices large and small, algorithmic exposure could slip in alongside the data work already being conducted.[420]

Although prosecutor offices are capable of using algorithmic exposure to reevaluate pretrial policy, many offices might be resistant to scrutinizing pretrial incarceration practices in this way. Even among progressive prosecutors who have promised a dramatic change in the government's approach to pretrial incarceration, their approach has been a gentler version of extant pretrial incarceration doctrine, rather than an alternative paradigm. Nonetheless, these offices have been flexible in reevaluating longstanding prosecutorial practice and present the best opportunity for algorithmic exposure to reshape prosecutors' pretrial decision-making. These offices have been amenable to critiques of existing law and practice, and they are experimenting with approaches that diverge from those of their tough-on-crime predecessors.

To date, it has been easy for prosecutors, judges, and the public to overestimate the accuracy of pretrial dangerousness determinations. Prediction seems to point to objective facts in the world: Without some intervention, people will commit crimes, abuse children, defraud the government, and attempt terrorist acts. To prevent this future from happening, government action may be needed. But the predictions that justify this action are hard to parse. The human predictive process is opaque and therefore resistant to analysis. And the outcomes always justify themselves, particularly carceral outcomes.[421] A judge is always proven right for locking someone up. Because the person was incarcerated, they couldn't harm the public. And if a judge lets someone go and that person harms others, that's reason for more government intervention next

---

418 Press Release, Commons Policy Changes, , https://yoloda.org/commons-policy-changes/ (last visited Jun 21, 2021).

419 *Id.*

420 Outside non-profits or other organizations are already positioned to assist prosecutors in these efforts. Collaborations between DA offices sharing data and non-profits analyzing the data can often blur the "inside-outside" line. The Vera Institute's Reshaping Prosecution program is a prime example of an inside-outside venture. Reshaping Prosecution, , Vera Institute of Justice , https://www.vera.org/projects/reshaping-prosecution-program (last visited Jun 21, 2021). Their self-defined goal is to "Help prosecutors implement data-driven policies and practices that reduce incarceration and promote racial equity and equal justice." *Id.* Organizations like these already have the data and staff to conduct this kind of analysis.

421 This point was recognized at the dawn of the predictive-carceral era. *See* Tribe, *supra* note 124.

time.[422]

Algorithmic exposure can lay bare how inaccurate predictions of pretrial violence are. Preventively incarcerating people for general predictions of violence requires primarily incapacitating people who do not need to be incapacitated. An algorithmic pretrial tool optimized to produce the most accurate results would not flag for anyone for violence. For any person in a dataset, the most likely outcome is that they will not commit violence on pretrial release. Therefore, to produce predictions of violence and encourage incarceration, accuracy must be sacrificed. This results in substantially more false positives — people who are flagged for violence but do not go on to commit a violent crime — than true positives — people who are flagged for violence and do go on to be arrested for a violent crime. The necessary consequence is jailing the many people labeled "high risk" to prevent the violence of a few of their members.

Accordingly, the effect of algorithmic exposure on prosecutors' pretrial policies ought to be a measure of modesty in dangerousness determinations and pretrial incarceration decisions.[423] Prosecutor offices could narrow the circumstances in which their attorneys are allowed to argue for pretrial incarceration. Permissible incarceration decisions would have to be justified by reasons beyond pure prediction, such as strength of the evidence, the gravity of the present offense, and concerns of the alleged victim(s).

With algorithmic insights in hand, a prosecutor's office would be uniquely positioned to challenge and change pretrial decision-making.[424] In some places, prosecutors have enough power over the pretrial process that a prosecutor office's policy limiting pretrial incarceration would become the *de facto* practice of the court system. Prosecutors' au-

---

[422] This phenomenon has been particularly prevalent within criminal law but has metastasized through American legal systems to domains including civil commitment, child protective services, government benefits fraud, and counter-terrorism policies, among others. *See* Bernard E. Harcourt, Against prediction: profiling, policing, and punishing in an actuarial age (2007).

[423] Algorithmic exposure is a mechanistic process, open to various normative applications. Although this Article explores how algorithms could be used to critique criminal law from a progressive, decarceral perspective, the insights that algorithms reveal about law are not inherently normative, and the reconception of algorithms as a tool for reevaluating law is not bound to a particular political perspective. To put it another way, algorithmic exposure is amenable to varying definition of fairness. Fairness is a capacious, contested idea. Hellman, *supra* note 169 at 814. Because the central claim of this article is a mechanistic one, it can accommodate many different conceptions of predictive fairness.

[424] A prosecutor who concludes that pretrial incarceration on general dangerousness cannot coexist with a commitment to racial justice will need to adopt a new paradigm for pretrial policy. The good news is that this is exactly what they were elected to do. But the academy, civil society and policy think tanks need to seize this moment and think beyond technocratic adjustments to our pretrial laws and policy. At present, the policy options tend to be pretrial incarceration on general dangerousness or bust. There are no off-the-shelf policy alternatives to pretrial incarceration on general dangerousness other than abolition or issuing a moratorium on pretrial incarceration. Both of these are political nonstarters, even in progressive cities like San Francisco or Boston.

thority over pretrial incarceration can vary by jurisdiction. In some places, prosecutors are the gatekeeper for pretrial incarceration: A judge can order a person to be incarcerated pretrial only upon a prosecutor's motion.[425] But this is the exception. In most places, the judges are in charge. Whether by imposing an unaffordable money bond amount or directly ordering someone to be incarcerated pretrial, judges can either respond to a prosecutor's motion for pretrial incarceration or choose to jail people on their own initiative. Even in places where prosecutors cannot singlehandedly limit pretrial incarceration, arguments from the prosecution can influence judges — both with individual decisions and with systemic approaches to pretrial decision-making.[426]

Insights from an algorithmic diagnosis of law could muscle their way into frontline attorneys' arguments in court. State laws afford prosecutors broad discretion over the pretrial incarceration arguments they can make in court. Pretrial incarceration laws tend to be permissive, not mandatory.[427] They don't require the state to incarcerate anyone. They instead permit the incarceration of people who pose a unique danger to the community. Under these laws, prosecutors could argue that the conditions the law requires are rarely met. The law may allow the incarceration of people who pose a serious risk of causing "serious bodily harm" to others, but this a high threshold.[428] Prosecutors could argue that the law requires more than a weak generalized prediction of someone's future behavior to justify pretrial incarceration.

In an analogous pattern, the Philadelphia District Attorney's office now requires assistant district attorneys to calculate the projected costs to the state for incarcerating someone and to share that information with the court at sentencing.[429] If the assistant district attorney believes the costs of incarceration are justified, the attorney must make a cost-benefit argument at the sentencing hearing. A similar policy could be adopted in pretrial hearings. For every person that the court is considering incarcerating pretrial, frontline attorneys would have to present to the court, first, the statistical likelihood of that person committing a violent crime if released and, second, statistics of the disparate burdens that pretrial incarceration places upon local communities of color. Prosecutor recommendations for pretrial incarceration or judicial orders of pretrial incarceration would have to be made in light or in spite of this data.

---

[425] *E.g.*, N.J. Stat. Ann. § 2a:162-16, 18 (West 2017).

[426] As I've argued with co-authors before, court culture is understudied, underappreciated element of criminal law reform, particularly in circumstances in which judges have broad discretion, like pretrial hearings. *See generally* Mitali Nagrecha, Sharon Brett & Colin Doyle, *Court Culture and Criminal Law Reform*, 69 Duke Law J. Online 84 (2020).

[427] *E.g.*, Mass. Gen. Laws Ann. ch. 276, § 58A (West 2021).

[428] Cal. Const. art. I, § 12.

[429] Chris Palmer, *In latest edict, Philly DA Larry Krasner tells prosecutors to seek lighter sentences, estimate costs of incarceration*, The Philadelphia Inquirer, https://www.inquirer.com/philly/news/crime/philadelphia-district-attorney-larry-krasner-plea-deals-shorter-sentences-cost-of-mass-incarceration-20180315.html (last visited Aug 10, 2021).

Outside of court, prosecutors could use insights from algorithmic analysis of pretrial doctrine to reframe the public debate over crime and preventive incarceration. The prosecutor plays an important role as spokesperson for the government on all matters of criminal law. Prosecutors tend to be in the limelight when high-profile cases attract media attention. But prosecutors are regularly asked to comment on other aspects of our criminal legal systems. In public disagreements with judges or police, prosecutors could use algorithmic analysis of their own policies and compare them to how courts deviate from their recommendations to show what would have resulted if a prosecutorial understanding of fairness had been adopted. Data from algorithms critiquing law can be a countermeasure to the newspaper headlines that inevitably appear when someone on pretrial release commits a notorious crime. Predicting violence will always seem easier than it is, particularly in hindsight. Empirical insights, particularly insights drawn from data about the jurisdiction itself, can counter these tendencies and bring a clear-eyed realism to emotionally charged — but statistically exceptional — circumstances.

Concerns & Opportunities

So far, this Article has developed the concept of algorithmic exposure of law and has demonstrated the concept's practical utility. This Part looks ahead, noting the conditions that might prevent algorithmic exposure from being adopted or might limit its effectiveness. Algorithms can be a powerful tool for revealing a law's potential, but the method cannot answer all questions and can be susceptible to misuse. This part examines some of the prospects and pitfalls that lay ahead.[430]

Biased Data

Criminal law data is biased. And algorithms that are trained on this data will replicate those biases. Garbage in, garbage out: Any machine learning model is only as good as the data fed into it. If it is fed biased data, it will produce biased results. Algorithmic tools currently used in criminal law rely on historical records of arrests, charges, convictions, and sentences to generate predictions about a person's future behavior. These tools implicitly assume that criminal history data are a reliable and neutral measure of underlying criminal activity. But these records reflect not just people's activity but also the activity of courts and police.

Court and police data is dirty data. People of color are treated more harshly than similarly situated white people at each stage of the legal system, which results in serious distortions in the data used to develop legal algorithms.[431] Take arrest records as an ex-

[431] See generally The Sentencing Project, Report of the Sentencing Project to the United Nations Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xeno-

ample. Arrest records are both under- and over-inclusive of the true crime rate. Arrest records are under-inclusive because they only chart law enforcement activity, and many crimes do not result in arrest.[432] Less than half of all reported violent crimes result in an arrest, and less than a quarter of reported property crimes result in an arrest. Arrest records are also over-inclusive because people are wrongly arrested and arrested for minor violations, including those that cannot result in jail time. For decades, communities of color have been arrested at higher rates than their white counterparts, even for crimes that these racial groups engage in at comparable rates.[433]

Biased data is an intractable problem that raises ethical concerns for using court and police data to justify decisions to punish, incarcerate, or otherwise restrict people's liberty. But for exposing law with algorithms, the ethical concerns are not quite as clear. Biased data can present ethical concerns about using algorithms to analyze laws, but these concerns are dependent on what analysis is being done and toward what ends. There should be no doubt: the same fundamental problem of biased results persists. Algorithms that are used to study law that have been trained on biased data will produce biased results. These biases should be taken into account when making any claims

---

phobia, and Related Intolerance Regarding Racial Disparities in the United States Criminal Justice System (2018); Lynn Langton & Matthew Durose, U.S. Dep't of Justice, Police Behavior During Traffic and Street Stops, 2011 (2013); Stephen Demuth & Darrell Steffensmeier, The Impact of Gender and Race-Ethnicity in the Pretrial Release Process, 51 Soc. Probs. 222 (2004); Jessica Eaglin & Danyelle Solomon, Brennan Center for Justice, Reducing Racial and Ethnic Disparities in Jails: Recommendations for Local Practice (2015); Sonja B. Starr & M. Marit Rehavi, Racial Disparity in Federal Criminal Sentences, J. Pol. Econ. 1320 (2014); Marc Mauer, Justice for All? Challenging Racial Disparities in the Criminal Justice System (2010).

[432] FBI, 2017 Crime in the United States: Clearances, https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topicpages/clearances (last visited June 28, 2019).

[433] Megan Stevenson & Sandra G. Mayson, The Scale of Misdemeanor Justice, 98 B.U. L. Rev. 731, 769-770 (2018). This comprehensive national review of misdemeanor arrest data has shown systemic and persistent racial disparities for most misdemeanor offences. The study shows that "black arrest rate is at least twice as high as the white arrest rate for disorderly conduct, drug possession, simple assault, theft, vagrancy, and vandalism." Id. at 759. This study shows that "many misdemeanor offenses criminalize activities that are not universally considered wrongful, and are often symptoms of poverty, mental illness, or addiction." Id. at 766. For example, Black people are 83% more likely to be arrested for marijuana compared to whites at age 22 and 235% more likely to be arrested at age 27, in spite of similar marijuana usage rates across racial groups. "[R]acial disparity in drug arrests between black and whites cannot be explained by race differences in the extent of drug offending, nor the nature of drug offending." Ojmarrh Mitchell & Michael S. Caudy, Examining Racial Disparities in Drug Arrests, Just. Q., Jan. 2013, at 22. Similarly, Black drivers are three times as likely as whites to be searched during routine traffic stops, even though police officers generally have a lower "hit rate" for contraband when they search drivers of color. Ending Racial Profiling in America: Hearing Before the Subcomm. on the Constitution, Civil Rights and Human Rights of the Comm. on the Judiciary, 112th Cong. 8 (2012) (statement of David A. Harris).

based on information produced by these algorithms. But dirty data doesn't necessarily contaminate or discount an entire project.

Another aphorism worth remembering is that all models are wrong, but some are useful.[434] All models are imperfect representations of the world. Some useful claims can be made even when the data and results are biased. Consider this Article's proposal to use algorithms to understand the distribution of outcomes that necessarily result from a pretrial incarceration law. This algorithmic model would be trained upon biased criminal law data. And yet, the bias would not reduce the usefulness of the insight the model provides because the model would not be making a claim about the behavior of people accused of crimes. To make that claim would require interpreting biased data as though it were an unbiased report of ground truth. Instead, the algorithm reports on how a legal system would operate. The distribution of legal outcomes is surely distorted by biased data. But claims about the distribution are not.

Biased data introduces uncertainty into any model that depends on that data to represent truth. Biased data creates a rift between reality and the statistical model that is being built. Consider claims about accuracy in pretrial dangerousness predictions. Unlike the claim about distribution of outcomes, this claim is much more susceptible to being distorted by biased data. Any claims of predictive accuracy that are based on biased data will be biased, because the claims are based on accepting the data as a true representation of how people behave. Because the data that guides the model diverges from the real world, the model is likely to overestimate the potential accuracy of predictions of violence. Therefore, any claims about accuracy must be qualified whenever dirty data is an issue. The less reliable the underlying data set is, the less reliable the claims about accuracy can be. But this doesn't mean that a claim about accuracy can never be made if data is biased. All data and all models diverge from reality. Claims about accuracy must be qualified and informed by knowledge about the domain. In some circumstances biased data may so overwhelm a dataset that one can't reliably make certain claims. But this claim would not just be about the deficiencies of an algorithmic model that is trying to understand accuracy. This would be a claim about the ability to make these kinds of predictions at all. Under these circumstances, given that the information about the world is so unreliable, neither humans nor algorithms would have sufficient knowledge to make accurate predictions.

Knowledge and Values

Algorithmic exposure has the potential produce new empirical knowledge about that world gained through algorithmic processing of data. It's the type of knowledge that would be expected from this process: knowledge about the world acquired through observation.[435] Here, knowledge has been produced through observing the process

434 Insert from chapter

435 Within the article, existing empirical knowledge filled in some insights that algorithms

and results of prediction. The process of constructing algorithms is a means by which one can view and understand how a predictive legal process operates. The claim of this Article is that algorithms can help to reveal information about legal prediction that was previously unknown.

At the same time, these algorithms may produce knowledge that is already known to some groups and not others based upon their social identities.[436] "Situated knowledge" is a term introduced by Prof. Donna Haraway "as a means of understanding that all knowledge comes from positional perspectives."[437] One's social position always determines what it is possible to know about a given object of study.[438] Too often what passes for "objectivity" is the view from the position of a white, male researcher. In criminal law, this position more closely aligns with people administering the criminal legal system than those affected by it. Some of the knowledge produced by using algorithms as a diagnostic tool for criminal law would be new information for those in

---

would reveal about the law. If this works, why do we need the process of using algorithms to expose the law at all? Why not just rely upon social science research?

In most contexts, this kind of robust social science research doesn't conveniently exist to fill in the gaps. This case study was chosen, in part, because it's an area that has been subject to extensive study by social scientists, and this information overlaps with the information algorithms could reveal about pretrial law. Most areas of law will not have this kind empirical research to fall back on. Even when this research does exist, it's often less useful for local advocacy than insights into how local laws affect the local population. Empirical research tends to be about one particular jurisdiction, either at the state or county level. Advocates must make a series of inferential steps to connect research about Kentucky to a District Attorney campaign in San Francisco. These inferences require trust in the out-of-state study design and trust that research conducted in other places at other times applies to the current time and place.

Social science research and machine learning tools are complementary, not exclusive. In practice, machine learning is an engineering tool more than it is a tool of social science research. There are significant overlaps between what can be uncovered in this process and what can be uncovered by the empirical methods used in economics and other disciplines. Indeed, some academic research explicitly designs and tests machine learning models. Kleinberg et al., *supra* note 87. As it is generally deployed, machine learning lacks some of the rigor of social science empirical methods. Machine learning does not generate causal insights, and its findings are not held to the same exacting standards of empirical methods. Machine learning is a practical and flexible engineering tool deployed in business and government to decipher troves data and derive workable insights. Across many domains, the steady march of social science and the nimble dance of machine learning can work together.

[436] Rua M. Williams & Juan E. Gilbert, *Cyborg Perspectives on Computing Research Reform*, *in* Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems 1–11, 1 (2019), https://dl.acm.org/doi/10.1145/3290607.3310421 (last visited Jul 7, 2021). "Situated knowledge" is a term introduced by Prof. Donna Haraway. *Id.*

[437] *Id.* at 2.

[438] *Id.* at 2.

that privileged social position. But much of that knowledge would not be new to people from communities who are heavily policed, prosecuted, and punished.

Arguments about the systemic inequities of criminal law are not new.[439] Part III's focus on pretrial incarceration is a good example. Community groups across the country, particularly community bail funds, have spent years protesting and advocating against contemporary pretrial incarceration laws and practices. A cursory look at their advocacy would reveal their knowledge that prosecutors — including the new flock of progressive prosecutors — are incarcerating primarily poor Black men on weak predictions of future wrongdoing. An algorithmic diagnosis of pretrial incarceration laws may confirm that knowledge, but it would not be revelatory to those groups in the way it might be revelatory to people from positions of privilege and remove from the criminal legal system.

Although much of the knowledge that algorithmic exposure produces may not be new to the communities most affected by the criminal legal system, these algorithmic insights can still support community-led efforts both as an empirical foundation and as a lightning rod for attention. These algorithmic insights can make advocates' arguments more difficult to dispute, more broadly appealing, and more likely to garner attention.

Quantifying and confirming grassroots knowledge can support grassroots efforts. It can validate the experience of members of those communities to a broader public and galvanize support for change. Empirical knowledge can affirm information that is already known by some groups of people and broaden the audience of people who will acknowledge that information. This empirical knowledge can be helpful by revealing how the law works step-by-step and pinpointing where to direct efforts at reform. Data-driven insights can make concrete much of what can be uncertain or speculative within arguments rooted in personal experience.[440] And the opposite is true: stories of people's experiences with criminal legal systems can enrich the data that algorithms glean about pretrial laws by contextualizing this information and framing arguments within a moral reasoning.

Technology can be a lightning rod for attention. As Rediet Abebe has highlighted, technological insights can act "as synecdoche when [they make] long-standing social problems newly salient in the public eye."[441] For better or worse, empirical information has political cache. There are constituencies who will discount community voices and

---

[439] *See e.g.*,Steinberg, *supra* note 128; Tribe, *supra* note 124; Hegreness, *supra* note 128; Baradaran, *supra* note 128; Thaler, *supra* note 128; Foote, *supra* note 124; Alschuler, *supra* note 128.

[440] *See e.g.*, Tribe, *supra* note 124; Foote, *supra* note 124; Alschuler, *supra* note 128; Baradaran, *supra* note 128; Thaler, *supra* note 128.

[441] Rediet Abebe, Designing Algorithms for Social Good 11–12 (2019). Depending on how progressive prosecutors respond to these attempts at accountability, algorithmic insights can also support broader critiques of progressive prosecution as a means of transformative change to criminal law.

advocacy but who will listen to empirical data that confirms that community's experiences.

This terrain is fraught. The process of using algorithms to expose problems in criminal law ought be approached with humility, grace, and accountability to the people affected by these laws. Claims of "new" information should not be overstated when this information confirms what many have been experiencing, protesting, and contesting for years. Researchers and advocates must be careful in their work to use this information as a support for grassroots efforts and not use their privileged positions to supplant or ignore that situated knowledge. We do not need algorithms to discover that our criminal legal system is inequitable, but algorithmic insights have the potential to reinvigorate, expand, or reinforce existing lines of critique.

Algorithms for All

The biggest obstacle to exposing law with algorithms is access to clean, high-quality datasets.[442] Machine learning depends upon high quality labeled datasets. The quality of government data sets can vary significantly by place, particularly with state and local governments. In some places, court records are still mostly paper records and not digitized. Other places have advanced to using an entirely digitized process. Many court systems do not have common databases or do not collect data in different courts and offices in a way that can work together. In many places, data is spread across different institutions. A court system may use a software that differs from probation that differs from local jails that differs from prisons that differs municipal record keeping. Each of these institutions may produce data that is labeled differently or can't interact well with another institution's data or does not track important people and actions. Most of the data that legal systems collect has not been collected with later statistical analysis in mind. It has been collected simply as a matter of record keeping. Accordingly, the quality of data can often depend on the quality needed to maintain administrative files. In many cases, this results in very messy data.

A lack of quality, labeled datasets is not unique to law but is *the* classic data science problem.[443] This is the problem that most data science projects face in the real world. It's not easily fixable, it but is something that ought to be gradually overcome with time. Courts and governments may be slow to adapt the digital world, but progress is being made as our world becomes ever more digital. For the short term, it may be enough of an impediment that algorithms cannot be built to study certain laws.

---

[442] Not too many years ago, computing power may have been an obstacle. That's no longer the case. Home and office computers are more than capable of data science projects well beyond what's contemplated here. No expensive hardware or software is required.

[443] The reason we have such robust algorithms for image detection is in no small part due to the open source release of ImageNet.

Even where usable datasets exist, dataset disclosure may be another obstacle. In most places, court data is not publicly available and is also not available by request. Most states' Freedom of Information Act laws include barriers to acquiring this kind of data en masse. Sometimes, the cost of making the FOIA requests and paying fees to the government to acquire the data can be prohibitive. In most cases, getting datasets requires government cooperation. For projects that are critical of the government, this is a challenge.

But transparency in local criminal government is becoming more common, and there's hope that transparency will increasingly become the norm. In many places, apolitical non-profits have formed with the explicit goal of procuring and sharing criminal case data.[444] Progressive prosecutors' election campaigns have emphasized transparency and accountability. Some prosecutor offices have released unprecedented amounts of data, while others have established data research units within their offices.[445]

Algorithmic exposure is a tool for more than just the government and well-heeled non-profits.[446] The work can also be completed by smaller community advocacy groups, like bail funds and court-watch programs.[447] In recent years, these community groups have quietly become more sophisticated authors of empirical and data-driven projects. Community bail funds operate across the country to challenge state and county money bail systems.[448] They are well known for posting money bond for people who would otherwise remain in jail pretrial on bond amounts that they cannot af-

---

444 *E.g.*, Texas Justice Initiative | Home Page, , https://texasjusticeinitiative.org/ (last visited Aug 7, 2021); Measures for Justice, , Measures for Justice , https://measuresforjustice.org/ (last visited Aug 9, 2021).

445 Lucy Lang & Erica Bond, *Prosecutors and the 'Moral Imperative' for Transparency*, The Crime Report (2021), https://thecrimereport.org/2021/03/19/prosecutors-and-the-moral-imperative-for-transparency/ (last visited Aug 9, 2021). Kim Foxx, the prosecutor for Cook County, Illinois (which includes Chicago), famously released a dataset that detailed what happened in every felony case her office processed over a six-year period. Matt Daniels, *The Kim Foxx Effect: How Prosecutions Have Changed in Cook County*, The Marshall Project (2019), https://www.themarshallproject.org/2019/10/24/the-kim-foxx-effect-how-prosecutions-have-changed-in-cook-county (last visited Aug 9, 2021).

446 Admittedly, larger non-profits are already well-positioned to conduct the data-intensive work of algorithmic exposure. Organizations like the Vera Institute and the ACLU have data research teams and projects. Two of the largest foundations funding criminal law and pretrial reform, Arnold Ventures and the MacArthur Foundation, have heavily funded data-driven research and reforms.

447 This case study looks outside the boundaries of traditional expertise and data-driven authority for two reasons: 1) to show the breadth of opportunities available to use algorithms as a tool for critiquing law, and 2) to bring attention to community groups' underappreciated data analysis capabilities.

448 *See* National Bail Fund Network, , Community Justice Exchange , https://www.communityjusticeexchange.org/en/nbfn-directory (last visited Aug 10, 2021).

ford.[449] By posting bond without considering a person's potential dangerousness or flight risk, these groups challenge the efficacy and wisdom of the money bail system and the judges who operate it.[450] Less well known are their broad advocacy efforts, including detailed, data-filled reports on local criminal legal systems' functioning.[451]

Many grassroots organizations are undertaking increasingly sophisticated data projects.[452] They're not shying away from numbers, and they're using methods one would typically associated with larger government organizations or non-profits.[453] With comprehensive, freely available data science education and software available on the internet, machine learning is an increasingly common skill.[454] Large scale data science projects are now possible outside traditional institutions.[455] Community bail funds have become increasingly sophisticated collectors and disseminators of qualitative and quantitative data, including their own alternative datasets that track the functioning of criminal legal systems. Many bail funds have started or partnered with court-watch programs in which volunteers observe court proceedings and collect qualitative

---

449 *See* Johna Engel Bromwich, *How a Minnesota Bail Fund Raised $20 Million to Help Jailed Protestors*, N.Y. Times, June 1, 2020, https://www.nytimes.com/2020/06/01/style/minnesota-freedom-fund-bail-george-floyd-protests.html (last visited Aug 10, 2021).

450 *See* The Bail Project, https://bailproject.org/ (last visited Aug 10, 2021).

451 *E.g.*, Monitoring Cook County's Central Bond Court: A Community Courtwatching Initiative, (2018), https://chicagobond.org/wp-content/uploads/2018/10/courtwatching-report_coalition-to-end-money-bond_final_2-25-18.pdf (last visited Aug 10, 2021).

452 The case study also bridges a divide in the literature between scholarship that advocates for an expert, empirical path to criminal law reform and scholarship that champions a democratic, community-driven approach. *See* Erin Collins, *supra* note 19 (collecting sources); *see also* John Rappaport, *Some Doubts About "Democratizing" Criminal Justice*, 712 Univ. Chic. Law Rev. 711, 715–17 (2019) (tracing this tension in the literature). This case study opens the question of whether the empirical and the democratic paths are as diametrically opposed as the literature may seem to suggest. *See* Benjamin Levin, *Criminal Justice Expertise* ___ Fordham L. Rev. __ (forthcoming 2022) (analyzing different conceptions of expertise as reflective of underlying political and ideological values). For decades, data analysis was the exclusive purview of professional experts within large institutions. Their capabilities still dwarf those of bail funds and court-watch programs, but they no longer have a monopoly. With limited funding and labor, grassroots organizations need to be selective with the data projects they undertake, but these projects are increasingly part of their portfolio.

453 Philadelphia Bail Fund, Rhetoric vs. Reality: The Unacceptable Use of Cash Bail by the Philadelphia District Attorney's Office During the COVID-19 Pandemic 5 (2020) (using random sampling to make inferences about the District Attorney's office's broader practices).

454 David Venturi, *I ranked every Intro to Data Science course on the internet, based on thousands of data points*, freeCodeCamp.org (2019), https://www.freecodecamp.org/news/i-ranked-all-the-best-data-science-intro-courses-based-on-thousands-of-data-points-db5dc7e3eb8e/ (last visited Aug 10, 2021).

455 [Insert reference to the data for good movement and volunteer opportunities for data scientists].

and quantitative information.[456] Court-watch programs have used this information to track judges' adherence to pretrial laws,[457] to share data on pretrial incarceration decisions, and to share stories of the daily injustices that occur within our criminal courts.[458]

The Tyranny of Metrics

Data science is a sociotechnical process, and researchers play an active role in constructing meaning from data.

Using algorithms to critique law embodies the "studying up" research that co-authors and I have encouraged data scientists and the broader algorithmic fairness research community to pursue.[459] In prior work, we introduced the algorithmic fairness community to the anthropological concept of "studying up,"[460] by drawing upon Prof. Laura Nader's call for her fellow anthropologists to move beyond the study of people and cultures at the peripheries of Western society — who comprised the bulk of the anthropological canon — in favor of studying how power operates through elite institutions and positions of authority.[461] This article answers that call to "study up" by repurposing legal algorithms to study the laws that algorithms are typically intended to optimize. Instead of studying and predicting the behavior of marginalized groups, algorithms have an alternate potential of studying and exposing how the law uses the language and logic of prediction to justify legal judgments against these groups.

Demystify metrics.

Demystify prediction.

Surfacing normative values in algorithmic and legal discourse.

---

456 Jia Tolentino, *Where Bail Funds Go From Here*, The New Yorker, June 23, 2020, https://www.newyorker.com/news/annals-of-activism/where-bail-funds-go-from-here (last visited Aug 10, 2021).

457 *E.g.*, Monitoring Cook County's Central Bond Court: A Community Courtwatching Initiative, *supra* note 201 at 35–37.

458 *E.g.*, Court Watch NYC Spring Newsletter, 3 (2018), https://static1.squarespace.com/static/5a21b2c1b1ffb67b3f4b2d16/t/5b0ee79cf950b7742628e513/1527703453246/Spring+2018+CWNYC+Newsletter.pdf (last visited Aug 10, 2021).

459 Barabas et al., *supra* note 17 at 1.

460 *Id.* at 1.

461 Laura Nader, *Up the Anthropologist: Perspectives Gained from Studying Up* 29 (1972).

Conclusion

The future of criminal law and the prospects for racial justice are tied up with the future of machine learning and artificial intelligence. As machine learning and artificial intelligence reshape law and society, activists and scholars have identified how this new technology poses the threat of generating new inequalities and "cloak[ing] and amplifying existing ones."[462] This concern is legitimate and important. When limited to optimizing current practices, legal algorithms entrench and perpetuate current inequality. But the law is not neutral backdrop in this process, and algorithms can do more than optimize how a law is applied. Algorithms have an overlooked potential to expose unjust laws and provide empirical support for transformative changes. In campaigns to reshape our criminal legal systems, algorithms should be recruited as a welcome — if unexpected — ally.

[462] Mimi Onuoha, Notes on Algorithmic Violence (2018), https://github.com/MimiOnuoha/On-Algorithmic-Violence (last visited Aug 10, 2021).