

(RE)INDIVIDUALIZING CRIMINAL LAW

ITAY RAVID

INTRODUCTION	257
I. THE RISE OF ALGORITHMIC RISK ASSESSMENTS IN THE CRIMINAL LEGAL SYSTEM	263
<i>A. Algorithm Use in the Criminal Legal System</i>	263
<i>B. Scholarship on Algorithmic Utilization: An Overview</i>	266
II. THE RISE OF SUBJECTIVE LIABILITY IN CRIMINAL LAW	271
<i>A. The Doctrinal Shift and Its Philosophical Foundations</i>	272
<i>B. Manifestation in Practice</i>	274
III. THE NON-SUBJECTIVE ALGORITHMIC LOGIC	282
<i>A. The Algorithmic Logic and the Clash with the Individual</i>	283
<i>B. The Algorithmic Logic and the Clash with Liberal Democratic Values</i>	285
<i>C. The Proliferation of the Non-Subjective Algorithmic Logic</i>	290
1. The Realpolitik of the Criminal Process	290
2. Epistemic Proliferation	291
3. Tangible Proliferation	297
4. The Continuous De-Individualization of Racially Marginalized Communities	299
IV. PRESERVING SUBJECTIVE LIABILITY IN AN ALGORITHMICALLY-DOMINATED CRIMINAL LEGAL SYSTEM	302
<i>A. Legal Framework—Synthesis of Rights</i>	303
<i>B. Implementing the Framework: Re-Individualization Through Contestation</i>	307
CONCLUSION	313

(RE)INDIVIDUALIZING CRIMINAL LAW

ITAY RAVID*

Abstract: Criminal law purports to care about the individual culpability of those committing crimes. Individual mental states or defenses based on one's beliefs are considered foundational to moral responsibility and criminal conviction. The expansion of algorithmic risk assessment tools in the criminal legal system, however, threatens these principles, changing the criminal law subject from an autonomous individual to merely a predicted object. This Article argues that risk assessment tools—now functioning as gatekeepers at both the entry and exit stages of the criminal process, from predictive policing to compassionate release eligibility—have shifted the criminal system away from its foundation in individual culpability.

Scholarship on risk assessment tools has thus far largely analyzed their isolated impact on specific stages of the criminal process while portraying algorithmic tools as being bound by existing laws. This Article bridges these siloed discussions and demonstrates that far from being bound by existing laws, algorithmic tools threaten fundamental criminal law principles and reshape substantive criminal law itself. It thus uncovers the aggregated, overlooked harms these tools inflict on the foundations of modern criminal law. The Article then offers a novel legal framework to tackle these harms. By integrating the right to due process, the right to contest AI, and the right to speak in criminal proceedings, it proposes a system of “re-individualization through contestation.” This framework will not only recalibrate algorithmic tools to accord with fundamental criminal law principles, but also offer an opportunity to leverage mounting criticism of unchecked algorithmic expansion to catalyze solutions to decades-old problems in our criminal legal system.

* Associate Professor of Law & Harold Reuschlein Emerging Scholar, Villanova University, Charles Widger School of Law. JSD '20, JSM '13, Stanford Law School. For insightful thoughts and suggestions, I thank Gilat Bachar, Isabella Banks, Nadia Banteka, Matthew Bruckner, Michal Buchandler-Raphael, Erin Collins, Michelle Dempsey, Brett Frischmann, Maša Galič, Katrina Geddes, Eve Hanan, Paul Heaton, Sarah Lageson, Mark Lemley, Evelyn Malavé, Sandy Mason, Nicola Padfield, Frank Pasquale, Teri Ravenell, David Schwartz, David Sklansky, Christopher Yoo; the participants at the Stanford-UPenn-Northwestern Junior Faculty Forum on Law & STEM; ABA-AALS Criminal Justice Roundtable on Criminal Law and Technology; Andrew Ferguson, Guha Krishnamurthi and Melanie Reid; the Law and Society 2024 Annual Meeting “Philosophical Analyses of Criminal Laws and Punishment” panel; CrimFest 2024; Privacy Law Scholars Conference Europe 2024; The Criminal Jurisprudence and Philosophy Seminar at Cambridge University; and the Villanova Law & Howard Law Faculty Workshops. Paul Cooke, Caleb Kearney and Dave Goulden Naitove provided excellent research assistance. I am also grateful to the *Boston College Law Review* editors, particularly Haley Cole, William Danforth, Nathaniel Kennedy, and Kallie Polgrean, for their thoughtful, careful and diligent editorial work on this Article.

INTRODUCTION

Consider this fictional, but not unrealistic, story:

On a routine day, a police officer stopped John Doe while they were sitting with some of their friends at their neighborhood's central park. Apparently, some equipment was stolen from a nearby store, and John was a suspect. The police had no individual suspicion connecting John to the burglary. John, however, was included on a police "heat list," a product of a risk assessment algorithm that identified certain members of the community as having higher probabilities of offending based on various risk factors. John's inclusion in the heat list was based on their association with "a gang." Although John had no meaningful prior criminal record, their social circle included individuals who were known members of a gang that was allegedly active in that neighborhood. John refused to cooperate with the police, insisting that they had nothing to do with any burglary, but to no avail. John's inclusion on the heat list and presence near the store at around the time the equipment was stolen were sufficient for the police, who arrested John. Based on the same predictive evidence, the police recommended the prosecution charge John with burglary and request the maximum possible punishment in that jurisdiction. The prosecution used their own risk assessment tools to confirm John's predicted dangerousness, charged them, and requested substantial bail as a condition to their release. During the bail hearing, the court utilized another risk assessment tool, leading it to approve the requested bail, well above what John would have been able to post. Thus, John remained in jail awaiting trial. After a few days had passed, the district attorney reached out to John's attorney and offered a plea deal: John would admit their guilt in the burglary and in return receive a reduced sentence. John's attorney was surprised by the harsh deal offer, but the district attorney refused to negotiate and claimed that John's risk profile justified a harsh punishment. Concerned that a trial would end with a worse outcome, John decided to plead guilty and was sentenced to six months in prison.¹

One may notice that something is missing from John's story: at no point in time were the government agents interested in John's individual culpability. Nor were they required to be. John's risk profile dominated the decision-making process and rendered questions about John's individual liability largely irrelevant. This is intriguing because in contemporary criminal law the general assumption is that the prosecution must provide sufficient evidence of individualized guilt to establish a defendant's criminal culpability. John's story, how-

¹ Although John's story is fictional, it is certainly not dystopian. It synthesizes elements of the reality of the criminal process across the United States. See *infra* Parts I and III, for more discussion and illustrations. While reasonable minds might differ on the question of whether we are "there yet" or merely "almost there," the Article's core argument does not turn on that debate. Instead, it raises deeper questions about how the already pervasive use of risk assessment tools in the criminal legal system challenges foundational principles of criminal law.

ever, suggests otherwise. It demonstrates that exploration of an individual's criminal culpability, including one's mental state, belief for purposes of certain defenses, and even traditional evidentiary standards, are de-facto being replaced by the use of technology, particularly risk assessment algorithms in the criminal legal system. This Article offers an in-depth exploration of this troubling phenomenon and its practical and normative implications.

To fully understand why stories like John's raise significant concerns, one must first briefly delve into the history and evolution of criminal law. This historical exploration reveals a significant transition in criminal law from the late eighteenth century, often characterized by criminal law scholars as the "rise of subjective liability,"² or the transition from "character" to "responsibility."³ However one wishes to describe the transition, it reflects similar ideas: the aspiration of modern criminal law to replace broad notions of moral blameworthiness with measurable culpability in the process of guilt determination.⁴

The transition into a criminal law established on the premise of subjective liability was not advanced in isolation. Rather, it was a product of broader Western liberal ideas centering legal systems around autonomous subjects as part of a commitment to "democratic values . . . human dignity and self-determination."⁵ Thus, modern conceptions of subjective liability should be

² See LINDSAY FARMER, MAKING THE MODERN CRIMINAL LAW: CRIMINALIZATION AND CIVIL ORDER 171, 181 (2016) (discussing the evolution of criminal responsibility and introducing subjective liability). Specifically, Farmer ties this concept to the evolution of mens rea. *Id.*

³ See NICOLA LACEY, IN SEARCH OF CRIMINAL RESPONSIBILITY: IDEAS, INTERESTS, AND INSTITUTIONS 112–14 (Andrew Ashworth ed., 2016) (tracking the philosophical and doctrinal evolution of criminal liability).

⁴ FARMER, *supra* note 2, at 171.

⁵ See Sanford H. Kadish, *The Decline of Innocence* (responding to proposals to eliminate the concept of mens rea from criminal offenses and criticizing such proposals for neglecting to consider the connections between mens rea, broader ideas of subjective liability, and their representation of the criminal legal system's commitment to the individual), in *BLAME AND PUNISHMENT: ESSAYS IN THE CRIMINAL LAW* 65, 75–77 (1987); see also *infra* Part II.A note about use of terminology: the idea of subjective liability encompasses the significance of *individual* responsibility to the criminal law. It serves both as an internal clarifying concept that emphasizes the importance of individualizing subjects treated by the system, but also as an avenue to connect criminal law to broader concepts of individual freedom in the modern welfare state. Subjective liability is manifested through different developments in criminal law, both doctrinal and evidentiary. Mens rea is likely the main form through which subjective liability can (and was supposed to) be manifested in criminal law, but other concepts are relevant as well including defenses, accomplice liability, capacity and more. To that end, subjective liability does not only represent the core tool of criminal law to distinguish between the guilty and the innocent, but it is also the main channel through which ideas of individual dignity and autonomy are imported into the criminal legal system. This Article thus urges readers to think about specific criminal law concepts mentioned through an interrelated process: the development of the specific concepts through a subjective lens, and the advancement of individual dignity and autonomy as the broader consequences of their manifestation. The premise of this Article is that algorithmic risk assessment tools de-individualize subjects before the criminal legal system, consequently depriving them of their autonomy and dignity. Indeed, autonomy and dignity are complex concepts with a host

understood as a manifestation of broader notions about the subject's place in the criminal legal system as a whole individual entitled to autonomy in a liberal democracy.⁶ The American Law Institute's influential Model Penal Code (MPC), with its transformative vision of mens rea and emphasis on a subjective approach to defenses, offers a recent illustration of how ideas of subjective liability became the leading force behind substantive modern criminal law.⁷

This Article argues, however, that the use of predictive algorithmic risk assessment tools in the criminal process is in deep tension with ideas of subjective liability and, consequently, the broader, philosophical values standing at the heart of the system, or at least those it aspires to.

Our criminal legal system increasingly utilizes algorithmic risk assessment tools to assist or replace human decision-making in the criminal realm. From predictive policing through prosecution, bail determination, sentencing, post-conviction, and compassionate release,⁸ these tools now serve as gatekeepers at both the entry and exit stages of the criminal process and significantly affect one's treatment by the criminal legal system.⁹ Algorithms advance a certain logic, one that this Article terms the "algorithmic logic." The algorithmic logic is inherently non-subjective. It builds on the premise of predicting one's behavior based on a statistical correlation—not even causation—between an individual's characteristics and those of a statistical sample group. Algorithms do not purport to capture the flesh-and-blood individual's knowledge, intent, or beliefs. Rather, algorithms categorize the subject through broad behavioral patterns. Therefore, individuals assessed by algorithmic risk assessment tools are deprived of traditional notions of autonomy. They become a number, subject to sweeping behavioral generalizations reconstructed based on correlations with large swaths of data.¹⁰

of potential definitions. For the purposes of this Article, they should be understood as broad ideals that the notion of subjective liability aspires to maintain in the criminal legal system.

⁶ See Kadish, *supra* note 5, at 77 ("Much of our commitment to the democratic values, to human dignity and self-determination, to the value of the individual, turns on the pivot of a view of man as a responsible agent entitled to be praised or blamed depending upon his free choice of conduct.").

⁷ See *infra* Part II.

⁸ See *infra* Part I. Another short note on terminology. This Article uses the terminology of "AI" and "algorithmic" tools to refer to computational and statistical tools to resolve certain problems based on big data analytics. Although AI usually refers to the machines designed to resolve such problems, algorithms are the rules that lead to a certain computational outcome. In particular, this Article focuses on predictive risk assessment tools utilized throughout the criminal legal system. These tools are often AI-created algorithms.

⁹ See generally U.S. DEP'T OF JUST., ARTIFICIAL INTELLIGENCE AND CRIMINAL JUSTICE: FINAL REPORT (2024), <https://www.justice.gov/olp/media/1381796/dl> [<https://perma.cc/F7KA-G9NC>], for a comprehensive account of the use of AI, including risk assessment tools, in the criminal legal system.

¹⁰ See *infra* Part III.

Thus, the use of risk assessment tools—regardless of their technical sophistication—undermines the principles of subjective liability.¹¹ In fact, this Article posits that the utilization of these tools is likely one of the most—if not the most—severe liability-eroding mechanism previously introduced into our criminal legal system. This is due to two main reasons. First, the use of risk assessment tools has expanded to almost all parts of the criminal process, pre- and post-guilt determination.¹² That expansion not only introduces new forms of evidence and criteria for determining criminal culpability, but also reshapes the temporality of institutional decision-making, shifting it toward a system-wide *ex ante* adjudication of criminality. Second, the sophistication of these tools and their use of data mining techniques—all cloaked in the purported power of science—make the task of undermining their authority much harder. Relatedly, the continuous scholarly and practical emphasis on technological improvement has led to the acceptance of risk assessment tools by even the most skeptical users, ultimately contributing to their exponential growth.¹³ Technical sophistication, however, cannot reconcile the fundamental tension between the algorithmic logic and subjective liability. Risk assessment tools therefore remain a unique threat to the role of subjective liability in criminal law and, consequently, a catalyst for the decline of autonomy and dignity in the criminal legal system.

This Article identifies two parallel processes through which the algorithmic logic erodes principles of subjective liability, termed *tangible* and *epistemic* proliferation. *Tangible proliferation* demonstrates how the interdependence between different stakeholders and stages in the criminal process strengthens the dominance of the algorithmic logic in determining criminal culpability and case outcomes.¹⁴ Furthermore, it illustrates how the concept of guilt is transformed as actual guilt is conflated with potential guilt as an outcome-determinative factor in criminal cases.¹⁵ Concerns about the effects of risk assessment tools on exploring subjective liability are heightened by the realpolitik of the criminal process, particularly the phenomenon of vanishing trials, where most criminal cases end in a pretrial plea bargain. Therefore, as illustrated in John's case, guilt is *de facto* determined in the early stages of the criminal process where risk assessment tools are increasingly utilized.

¹¹ As such, this Article does not focus on the *methodology* we should use to make predictive judgments of risk but rather investigates whether the use of predictive judgments of risk is appropriate *at all* given the philosophical foundations of modern criminal law.

¹² See *infra* Part I.

¹³ See *infra* Part II.

¹⁴ John Doe's story demonstrates the "domino effect" of risk assessment tools by showing how risk assessment in each stage of the process affects the subsequent stage; how determination of risk—regardless of how problematic it might be—affects the charge, that affects the bail, that affects the plea deal offered, and, ultimately, the guilt determination.

¹⁵ See *infra* Part III.

The parallel *epistemic proliferation* more broadly describes how stakeholders in the criminal process think about the individuals treated under the system. In lieu of treating them as fully-recognized autonomous subjects—a core aspirational tenant of subjective liability—the algorithmic logic normalizes a vision of individuals as predicted objects. That objective vision has expanded well beyond the pre-guilt stage through sentencing, corrections, and even post-conviction releases. I argue that the Supreme Court 6-3 decision in 2024 in *Diaz v. United States*, endorsing general probabilistic evidence over individual mental state assessment, reflects this epistemic process.¹⁶

Erosion of the individual under the algorithmic regime has significant negative effects on those treated by the criminal legal process, particularly from racially marginalized communities. Even before the algorithmic age, these communities were traditionally de-individualized before the system. The adoption of algorithmic tools deepens their desubjectivation and isolation from the system. Consequently, racially marginalized individuals bear the brunt of the exponential growth of the algorithmic logic in the criminal legal system.¹⁷

In recent years, we have been exposed to a slew of scholarship on risk assessment tools in the criminal process. This Article builds on that important work. Most scholarship on risk assessment tools has concentrated on their impact within the stages of the criminal process where they are employed, however, often suggesting that technological advancements and design features can mitigate algorithmic harms.¹⁸ These are deeply critical conversations, but they remain confined to their own universe of exploration. This Article breaks these

¹⁶ In *Diaz*, Justice Thomas, in a majority opinion, allowed the admission of predictive evidence regarding the petitioner's mens rea by determining that such evidence does not violate Rule 704(b) of the Federal Rules of Evidence. 602 U.S. 526, 528 (2024); FED. R. EVID. 704(b). In an unusual dissent, Justices Gorsuch, Kagan, and Sotomayor emphasized how the decision de facto allows the government to use a "powerful new tool," probabilistic evidence, to replace the individual exploration of one's mental state, as subjective liability requires. *Diaz*, at 543 (Gorsuch, J., dissenting). Along the line of this Article's argument, the dissenters emphasized the negative effects of the decision on the "respect for the individual and his liberty in a free society." *Id.* at 543–44.

¹⁷ See *infra* Part III.

¹⁸ See generally Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57 (2018) (examining risk assessment tools in sentencing); Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59 (2017); Lauryn P. Gouldin, *Defining Flight Risk*, 85 U. CHI. L. REV. 677 (2018) (discussing risk assessment tools utilized by the police); Sean Allan Hill II, *Bail Reform and the (False) Racial Promise of Algorithmic Risk Assessment*, 68 UCLA L. REV. 910 (2021) (analyzing pretrial risk assessment instruments); Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490 (2018) (discussing risk assessment tools used during bail proceedings); Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15, 17–18 (2019), <https://nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/> [<https://perma.cc/4BXW-C82H>] (discussing risk assessment tools utilized by the police); Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803 (2014) (discussing risk assessment tools in sentencing). See *infra* Part I.B, for additional discussion.

siloed conversations and brings a new argument to the fore that forces us to think more deeply about the aggregated and overlooked effects of utilizing technology in our criminal legal system. Particularly, this Article explores how the use of technology can, and in fact does, fundamentally challenge the core values of this system, and how it strikes at the very heart of what the criminal legal system is designed to do.

Acknowledging that automated processes are likely here to stay in the criminal legal system, this Article concludes by offering an intervention that could potentially mitigate their effects. Integrating the right to due process, a suspect's right to speak, and the right to contest artificial intelligence (AI), this Article offers a novel legal framework, termed "re-individualization through contestation," that challenges the algorithmic logic by allowing individuals to present evidence about their lived experiences, explanations, thoughts, beliefs, and moral views. The Article moves beyond the normative and offers several preliminary directions for implementing processes to contest automated decision-making. These new processes can help to reintroduce subjective liability into the criminal legal process and restore modern conceptions of autonomy and individuality systemwide.¹⁹ Ironically, it may be that the unrestrained expansion of algorithmic tools in the criminal legal system and the subsequent scholarly criticism of this development marks a turning point. It may allow us to reimagine a subject-centered criminal legal system and advance solutions to decades-long problems.

This Article proceeds in four Parts. Part I describes the rise of algorithmic risk assessment tools in the criminal legal process as the most radical shift thus far from ideas of subjective liability.²⁰ It offers an overview of the current expansion of these tools and the main scholarly and policy arguments praising or criticizing this expansion.²¹ It further discusses the unique scope, sophistication, and risk of algorithmic tools.²² Part II offers a historical overview of the evolution of the concept of subjective liability through modern-day criminal law.²³ It then situates subjective liability within broader themes dominating the development of Western democracies, chief among them the ideas of individual autonomy and dignity.²⁴ Part III delves into the core tension this Article exposes between subjective liability's underpinnings in modern ideas of autonomy and dignity and an algorithmic logic informed by behavioral patterns and rooted in generality.²⁵ Part III also responds to potential critiques, illustrating

¹⁹ See *infra* Part IV.

²⁰ See *infra* Part I.

²¹ See *infra* Part I.

²² See *infra* Part I.

²³ See *infra* Part II.

²⁴ See *infra* Part II.

²⁵ See *infra* Part III.

how algorithmic tools have expanded beyond their designated scope and permeated the guilt phase of the criminal process, particularly in light of the phenomenon of vanishing trials.²⁶ It then explains how such proliferation negatively affects individuals in the criminal legal system, especially those from racially marginalized communities.²⁷ By synthesizing scholarship on suspects' right to speak, due process and the right to contest AI, Part IV offers a legal framework and preliminary derivative mechanisms that can assist in pushing back against the algorithmic logic and re-individualizing subjects in the criminal legal system.²⁸

I. THE RISE OF ALGORITHMIC RISK ASSESSMENTS IN THE CRIMINAL LEGAL SYSTEM

Part I illustrates the evolving use of risk assessment algorithms in the U.S. criminal legal system and discusses the main scholarly and policy arguments around this evolution.²⁹ Section A of this Part details how algorithms are used throughout the criminal legal system, from before a subject has been charged to sentencing (and even in post-sentencing settings).³⁰ Section B of this Part provides a brief review of existing scholarship concerning algorithmic use in the criminal legal system.³¹

A. Algorithm Use in the Criminal Legal System

Algorithms—or “statistical and computer driven decision-making processes”—are no longer a rare phenomenon in the criminal legal process.³² Algorithmic risk assessment tools, often created by AI, share similar ideas regarding the ability to predict one's behavior based on probabilistic formulae extracted from large swaths of data about others' behavior.³³

²⁶ See *infra* Part III.

²⁷ See *infra* Part III.

²⁸ See *infra* Part IV.

²⁹ See *infra* Part I.

³⁰ See *infra* Part I.A.

³¹ See *infra* Part I.B.

³² See CHRISTOPHER SLOBOGIN, JUST ALGORITHMS: USING SCIENCE TO REDUCE INCARCERATION AND INFORM A JURISPRUDENCE OF RISK, at vii (2021); U.S. GOV'T ACCOUNTABILITY OFF., GAO-18-142SP, ARTIFICIAL INTELLIGENCE: EMERGING OPPORTUNITIES, CHALLENGES, AND IMPLICATIONS 73–75 (2018), <https://www.gao.gov/assets/700/690910.pdf> [<https://perma.cc/YYYY5-G2QB>] (indicating that algorithms can aid multiple stages of the criminal process). See generally Itay Ravid & Amit Haim, *Progressive Algorithms*, 12 U.C. IRVINE L. REV. 527 (2022) (arguing that the dual trends of progressive prosecution and the growing use of algorithmic decision-making must be reconciled to reform the U.S. criminal justice system).

³³ See SLOBOGIN, *supra* note 32, at viii (discussing what algorithms do and suggesting how they might improve the criminal legal system); U.S. DEP'T OF JUST., *supra* note 9, at 9–10; see *infra* Part III.A.

The use of risk assessment tools in the criminal legal system begins before a subject has been charged with a crime. Indeed, police departments across the United States have adopted predictive policing tools³⁴ intended to direct enforcement efforts in the early stages of the criminal process.³⁵ Predictive policing uses data and analytical techniques to predict different aspects of criminal activity instead of relying on intuitions and heuristics.³⁶ Predictions cover crime incidents, potential perpetrators' identities, and possible victims.³⁷ This is a growing field of technological advancement, with new—and often contested—investigative techniques, such as facial recognition, constantly introduced into the system.³⁸ As others have argued, though some predictive methods are not new, predictive policing is altogether distinct.³⁹

Traditionally, the second stage of the criminal process involves prosecutors assessing the evidence in a given case to decide whether to pursue criminal proceedings. Although a relatively new domain of algorithmic exploration, predictive prosecution or “intelligence-led” prosecution is a growing area of interest for district attorneys and, consequently, developers.⁴⁰ Conceptually,

³⁴ As a general matter, algorithmic tools are often developed by third parties and are offered to the government. See Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1120–28, 1137, 1143–44 (2017) (examining how AI and algorithmic decision-making have been adopted in criminal justice, tracing the evolution of predictive policing, and presenting both practical and theoretical critiques); Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 878–79, 895–901, 929 (2016) (highlighting how predictive policing raises constitutional questions about probable cause standards under the Fourth Amendment); SARAH BRAYNE, *PREDICT AND SURVEIL: DATA, DISCRETION, AND THE FUTURE OF POLICING* 30–36, 56–73 (2020) (documenting the LA police department's multi-year implementation of predictive technologies). Predictive policing and related algorithmic tools draw criticism for potentially amplifying racial bias through their use of historically discriminatory police data. See generally Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2251–54 (2019) (asserting that prediction's fundamental weakness lies in its dependence on historical data, that projects existing disparities forward); *United States v. Curry*, 965 F.3d 313, 320, 330–31 (4th Cir. 2020) (en banc) (discussing predictive policing in jurisprudence and holding that predictive policing alerts alone did not create exigent circumstances sufficient to warrant a suspicion-less search).

³⁵ See WALTER L. PERRY ET AL., RAND CORP., *PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS* 57–60, 80 (2013), http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf [<https://perma.cc/2C6N-PCGR>] (arguing that predictive policing promotes efficiency and efficacy and establishes best practices for its implementation).

³⁶ Ferguson, *supra* note 34, at 1149.

³⁷ PERRY ET AL., *supra* note 35, at 57–60, 80; U.S. DEP'T OF JUST., *supra* note 9 at 42–49.

³⁸ Stephen Caines, *The Many Faces of Facial Recognition*, in RESEARCH HANDBOOK ON BIG DATA LAW 29, 29–30, 32–35, 37, 40–41 (Roland Vogl ed., 2021).

³⁹ See Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 113 (2017) (explaining the idiosyncratic nature of predictive policing methodologies in the current era of big data, where data mining methods are utilized to predict crime); see also Emily Berman, *Without Warrant*, BOS. REV. (Jan. 25, 2023), <https://www.bostonreview.net/articles/without-warrant/> [<https://perma.cc/RMJ6-9G5N>] (reviewing existing methods of police data collection in the digital age).

⁴⁰ See Andrew Guthrie Ferguson, *Predictive Prosecution*, 51 WAKE FOREST L. REV. 705, 705–06 (2016) [hereinafter Ferguson, *Predictive Prosecution*] (describing predictive prosecution as the

these tools exploit the broad discretion enjoyed by prosecutors in choosing who to prosecute and on what grounds. For example, assessments of the risk of reoffending can play a crucial role in deciding whether to prosecute, what charges to bring, what bail conditions to recommend, what plea bargain conditions to pursue, and, eventually, the type of sentence to demand.⁴¹

Although predictive prosecution is still utilized more modestly, the use of actuarial risk assessment in the next stage of the criminal process—decisions on pretrial detention—is widespread. At this stage, algorithmic tools focus primarily on assessing the risk of flight (“failure to appear”) and recidivism (committing a new crime).⁴²

Risk assessment algorithms are also incorporated as a component of the last stage of the criminal process: sentencing decisions.⁴³ Interestingly, the use of “some inchoate notion” of risk assessment at the sentencing stage is far from new, dating back to the late nineteenth and early twentieth century in the United States.⁴⁴ Over the years, however, the use of these tools gradually subsided.⁴⁵ Yet, recently, these tools have reemerged, mostly due to the increased

process of identifying individuals most likely to engage in future criminal behavior to influence decisions on bail, charges, and sentencing); Andrew Guthrie Ferguson, *Big Data Prosecution and Brady*, 67 UCLA L. REV. 180, 217–24, 229, 232–34, 236–38, 242–43 (2020) [hereinafter Ferguson, *Big Data Prosecution*] (highlighting that intelligence-driven prosecution could result in concerning practices regarding exculpatory evidence); Christopher Slobogin, *The Next Steps in Criminal Justice Reform*, JOTWELL (June 10, 2019), <https://crim.jotwell.com/the-next-steps-in-criminal-justice-reform/> [<https://perma.cc/64UX-38E9>] (reviewing MATTHEW W. EPPERSON & CARRIE PETTUS-DAVIS, SMART DECARCERATION: ACHIEVING CRIMINAL JUSTICE TRANSFORMATION IN THE 21ST CENTURY (2017)); ALISSA MARQUE HEYDARI, VAND. PROJECT ON PROSECUTION POL’Y (VPOPP) & POLICING PROJECT AT N.Y.U. UNIV. SCH. OF L., AI & PROSECUTION: MAPPING THE CURRENT AND FUTURE ROLES OF ARTIFICIAL INTELLIGENCE IN PROSECUTION 3–4 (2024), <https://buff.ly/4gob1pw> [<https://perma.cc/3CDN-M9SE>].

⁴¹ Ferguson, *Predictive Prosecution*, *supra* note 40, at 716–27.

⁴² See MONA J.E. DANNER, MARIE VANNOSTRAND & LISA M. SPRUANCE, LUMINOSITY, INC., RISK-BASED PRETRIAL RELEASE RECOMMENDATION AND SUPERVISION GUIDELINES 3–4 (2015), <https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/risk-based-pretrial-release-recommendation-and-supervision-guidelines.pdf> [<https://perma.cc/LY2V-VTY9>] (explaining how risk assessment tools are used in Virginia); Mayson, *supra* note 18, at 509.

⁴³ See Richard Berk & Jordan Hyatt, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, 27 FED. SENT’G REP. 222, 222–26 (2015) (explaining how machine learning algorithms are integrated into the sentencing process); Michael E. Donohue, *A Replacement for Justitia’s Scales?: Machine Learning’s Role in Sentencing*, 32 HARV. J.L. & TECH. 657, 660–67 (2019) (examining the use of machine learning algorithms in sentencing and the challenges they present); Starr, *supra* note 18, at 805, 811 (criticizing the growing use of actuarial recidivism risk prediction tools as a basis for sentencing decisions).

⁴⁴ John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. CLINICAL PSYCH. 489, 490–91 (2016).

⁴⁵ Risk assessment played a key role in U.S. sentencing until the 1970s but was gradually supplanted by other approaches, such as sentencing guidelines and mandatory minimums, that focused more on culpability than risk. See Monahan & Skeem, *supra* note 44, at 491–93 (examining legal theories of sentencing and their perspectives on risk assessment while demonstrating its potential role); Dawinder S. Sidhu, *Moneyball Sentencing*, 56 B.C. L. REV. 671, 687–90 (2015) (discussing the

availability of data and the development of new algorithms.⁴⁶ Lastly, post-conviction decisions—including parole, compassionate release, and other correctional determinations—increasingly rely on algorithms⁴⁷ to assess recidivism risk and evaluate appropriate treatment of prisoners and parolees.⁴⁸

B' Scholarship on Algorithmic Utilization: An Overview

Given the exponential growth in the use of risk assessment algorithms at different stages of the criminal process, it is not surprising to see an equally exponential growth in the literature praising or critiquing these tools. Previous work identified several general frameworks represented in this literature: accuracy, human biases, cost savings, and legitimacy deficits.⁴⁹ A brief summary of these frameworks is offered below.⁵⁰

The argument for *accuracy* contends that algorithms are superior, or at least equal, to human performance in predicting risk to offend or reoffend,⁵¹ and asserts that algorithms represent a more structured upgrade over the inter-

rise of actuarial predictions for criminal offenders' rates of recidivism in sentencing and arguing these are in tension with theories of punishment).

⁴⁶ See Monahan & Skeem, *supra* note 44, at 489–91 (discussing the evolution of risk assessment tools in the sentencing process); Michael O'Hear, *Actuarial Risk Assessment at Sentencing: Potential Consequences for Mass Incarceration and Legitimacy*, 38 BEHAV. SCIS. & L. 193, 193–95 (2020) (providing suggestions to maintain judicial legitimacy in the eyes of defendants while incorporating risk assessment algorithms in sentencing); Collins, *supra* note 18, at 60, 63–72 (documenting the rise of actuarial sentencing).

⁴⁷ See Richard Berk, *An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism*, 13 J. EXPERIMENTAL CRIMINOLOGY 193, 195 (2017) (assessing the role of machine learning forecasting in guiding parole release decisions).

⁴⁸ See Christopher Slobogin, *Introduction to the Special Issue on Implementing Post-Conviction Risk Assessment*, 38 BEHAV. SCIS. & L. 187, 187–89 (2020) (highlighting various applications of predictive tools in post-conviction and correctional environments). Some argue that the use of risk assessment tools in corrections predated the use of risk assessment tools in other areas of the criminal process. Collins, *supra* note 18, at 77–79.

⁴⁹ Ravid & Haim, *supra* note 32, at 539–46.

⁵⁰ See *infra* notes 51–71 and accompanying text.

⁵¹ Support for actuarial prediction dates back to the 1950s. See generally PAUL E. MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION: A THEORETICAL ANALYSIS AND A REVIEW OF THE EVIDENCE (1954); Stefania Egisdóttir et al., *The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction*, 34 COUNSELING PSYCH. 341 (2006). For a more recent review indicating that statistical methods consistently outperform others in predicting various outcomes, see Sharad Goel, Ravi Shroff, Jennifer Skeem & Christopher Slobogin, *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment* (asserting that risk assessment instruments are more accurate than unguided human predictions and criminal justice professionals' predictions for offenders reoffending), in RESEARCH HANDBOOK ON BIG DATA LAW 9, 9–11 (Roland Vogl ed., 2021); J.C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329, 1342 (2011) (highlighting the superiority of actuarial prediction compared to human judgment). See generally Malcolm Feeley & Jonathan Simon, *Actuarial Justice: The Emerging New Criminal Law*, in THE FUTURES OF CRIMINOLOGY 173, 174–75 (David Nelken ed., 1994), for the roots of incorporating actuarial approaches into the criminal legal system.

nal risk assessments already performed by decision-makers.⁵² Some, however, challenge the notion that algorithms are superior to human clinical assessment in predicting risk.⁵³ Some also argue that the evidence to support such a proposition is too limited to responsibly permit its widespread use.⁵⁴ Relatedly, cautious voices have expressed concern with the validation of algorithmic tools⁵⁵ and, more generally, with special considerations in sentencing.⁵⁶

Another leading argument in support of algorithms frames *human bias* as the most fundamental problem of the criminal legal system,⁵⁷ even among the judiciary.⁵⁸ Actuarial methods are more “objective” and can therefore potentially offer unbiased information to counteract human predispositions and inclinations. This argument is framed in the context of the racial biases prevalent in the criminal legal system.⁵⁹

⁵² MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a, at 387–89 (A.L.I., Proposed Final Draft 2017); see Jennifer Skeem, *Risk Technology in Sentencing: Testing the Promises and Perils* (Commentary on Hannah-Moffat, 2011), 30 JUST. Q. 297, 297–98 (2013) (illustrating how judges base their decisions on recidivism evaluations conducted by probation officers).

⁵³ See Starr, *supra* note 18, at 850–55 (holding that studies asserting the superiority of actuarial risk assessment models to clinical assessment offer little compelling evidence to support their claim). See generally Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES, Mar. 30, 2018, at 1, 1 (asserting that commercial risk assessment tools are no more accurate than non-experts in predicting risk).

⁵⁴ Starr, *supra* note 18, at 855.

⁵⁵ Monahan & Skeem, *supra* note 44, at 501–08; see Melissa Hamilton, *Judicial Gatekeeping on Scientific Validity with Risk Assessment Tools*, 38 BEHAV. SCIS. & L. 226, 227 (2020) (contending that the validity of risk assessment tools should not be taken for granted).

⁵⁶ Starr, *supra* note 18, at 870–72; Slobogin, *supra* note 48, at 188.

⁵⁷ Joseph J. Avery & Joel Cooper, *Racial Bias in Post-Arrest and Pretrial Decision Making: The Problem and a Solution*, 29 CORN. J.L. & PUB. POL’Y 257, 260–75 (2019). See generally BIAS IN THE LAW: A DEFINITIVE LOOK AT RACIAL PREJUDICE IN THE U.S. CRIMINAL JUSTICE SYSTEM (Joseph Avery & Joel Cooper eds., 2020) (discussing racial prejudice in the U.S. criminal justice system).

⁵⁸ See MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a, at 387 (A.L.I., Proposed Final Draft, 2017) (“Section 6B.09 takes an attitude of skepticism and restraint concerning the use of high-risk predictions as a basis of elongated prison terms, while advocating the use of low-risk predictions as grounds for diverting otherwise prison-bound offenders to less onerous penalties.”); see also Itay Ravid, *Judging by the Cover: On the Relationship Between Media Coverage on Crime and Harshness in Sentencing*, 93 S. CAL. L. REV. 1121, 1151–77 (2020) (providing empirical evidence and insights to explain the presence of this judicial tendency).

⁵⁹ See 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 FED. PROB. 38, 38 (2016) (“[R]isk assessment tools informed by objective data can help reduce racial bias from its current level. It would be a shame if policymakers mistakenly thought that risk assessment tools were somehow worse than the status quo.”); see also Jennifer L. Skeem & Christopher T. Lowenkamp, *Risk, Race, and Recidivism: Predictive Bias and Disparate Impact*, 54 CRIMINOLOGY 680, 690–705 (2016) [hereinafter Skeem & Lowenkamp, *Risk, Race, and Recidivism*] (identifying minimal evidence of racial bias in the application of risk assessment tools). Human predictions are unclear and difficult to analyze for racial bias, whereas algorithms may be more transparent and easier to evaluate. See generally ORLY LOBEL, *THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER,*

On the other hand, a growing number of scholars have argued that algorithms actually exacerbate biases and racially-oriented predispositions rather than mitigate them.⁶⁰ Scholarship in this domain challenges the façade of neutrality and argues that statistical models, however technically sophisticated, are only as good as the data upon which they rely.⁶¹ Harcourt, one of the fiercest critics of utilizing predictive tools in the criminal legal system, suggests that predictive methods will create a “ratchet effect” wherein minorities continue to suffer unequal enforcement, and the positive feedback loop of biased data and inequitable application is strengthened.⁶² Some argue, however, that this is a design-related argument that can potentially be addressed with the right methodology, such as by debiasing the data through introducing different models into the work of the algorithms,⁶³ or utilizing different, non-government-based sources of data in the design and training process.⁶⁴

MORE INCLUSIVE FUTURE (2022); Jennifer Skeem & Christopher Lowenkamp, *Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism*, 38 BEHAV. SCIS. & L. 259, 261 (2020) [hereinafter Skeem & Lowenkamp, *Trade-Offs*]; Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, J. LEGAL ANALYSIS, 2018, at 113, 114 (arguing that algorithms offer transparency, enabling the detection of discrimination). Some argue, however, that trade-offs are unavoidable in risk prediction no matter who is responsible for the decision. See generally Skeem & Lowenkamp, *Trade-Offs*, *supra*; Monahan & Skeem, *supra* note 44, at 496–97.

⁶⁰ See Goel et al., *supra* note 51, at 6, for an overview of this argument; see also Mayson, *supra* note 34, at 2251–59 (identifying two possible sources of racial disparities in arrest rates and arguing that, in both instances, these disparities would be very difficult to eliminate in algorithmic risk assessment models); Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 AM. CRIM. L. REV. 1553, 1564–77 (2019) [hereinafter Hamilton, *Biased Algorithm*] (documenting bias toward Hispanics in algorithmic risk assessment); Melissa Hamilton, *The Sexist Algorithm*, 37 BEHAV. SCIS. & L. 145, 146–54 (2019) (documenting gender bias in risk assessment tools leading to overclassification of women in high-risk groups).

⁶¹ See Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671, 671–74 (2016) (suggesting that algorithmic methods absorb and reflect biases embedded in the data they utilize); Michael Tonry, *Predictions of Dangerousness in Sentencing: Déjà Vu All Over Again*, 48 CRIME & JUST. 439, 449–50 (2019) (condemning the flawed and biased outcomes of recidivism forecasting); Hamilton, *Biased Algorithm*, *supra* note 60, at 1576–77 (finding that COMPAS, a popular risk assessment tool, is biased against Hispanics); Cecelia Klingele, *Making Sense of Risk*, 38 BEHAV. SCIS. & L. 218, 223–25 (2020) (criticizing risk assessment tools that “disproportionately classify . . . the poor as higher risk”). Starr has contended that actuarial methods using demographic, socioeconomic, or other protected variables raise constitutional and ethical issues, effectively constituting discrimination. Starr, *supra* note 18, at 803, 870–72. On the other hand, some argue that explicitly incorporating race into the model results in better-calibrated outcomes (in terms of false negatives and false positives) compared to race-blind models. Skeem & Lowenkamp, *Trade-Offs*, *supra* note 59, at 274.

⁶² Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT’G REP. 237, 237–38 (2015) [hereinafter Harcourt, *Risk as a Proxy for Race*].

⁶³ Emily Black et al., *Less Discriminatory Algorithms*, 113 GEO. L.J. 53, 53, 61–68, 79–86 (2024); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 678–83 (2017).

⁶⁴ See Ngozi Okidegbe, *Discredited Data*, 107 CORN. L. REV. 2046–50 (2022) (justifying the use of “non-carceral knowledge sources”).

As for *reducing costs*, proponents argue that actuarial methods can decrease incarceration levels without increasing risk or maintain incarceration levels while reducing risk.⁶⁵

Finally, algorithms have institutional-structural effects that may bring about a *legitimacy deficit*. For example, although decisions in the criminal legal system such as sentencing serve multiple purposes and comprise several components, algorithmic risk assessment enhances the relative weight ascribed to utilitarianist justifications for punishment, such as incapacitation and deterrence, in lieu of other retributive justifications, such as moral desert.⁶⁶ Moreover, risk assessment encourages diverse decision-making processes by only highlighting measurable attributes as opposed to more nuanced and complex questions of justice and fairness.⁶⁷ Similarly, some suggest that the “black box” nature of algorithms could erode public trust in the judiciary, regardless of their sophistication or performance.⁶⁸ Ensuring transparency is crucial to prevent these issues and uphold the due process rights of defendants.⁶⁹ Scholars of procedural justice have emphasized the importance of transparency and the ability to voice one’s arguments, including allowing defendants the ability to challenge prediction results in legal proceedings.⁷⁰ Offering a *de facto* path toward protecting such rights is challenging, however, given the technical

⁶⁵ See Skeem & Lowenkamp, *Trade-Offs*, *supra* note 59, at 261; PETER W. GREENWOOD WITH ALLAN ABRAHAMSE, SELECTIVE INCAPACITATION 70–80 (1982) (discussing the effects of predictive models on crime rates); Jennifer Skeem, Lina Montoya & Christopher Lowenkamp, *A Pretrial Release Policy Based on Risk Assessment Would Reduce Unnecessary Incarceration, Increase Racial Fairness, and Save Money* 24–25, 32–33 (draft, July 2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5343174 [<https://perma.cc/5KAG-BCC5>] (finding that the use of the authors’ pretrial release strategy based upon risk assessment algorithms would reduce detention times, increase cost-savings, and decrease racial disparities in the pretrial detention context). See generally Christopher Slobogin, *Introduction to the Special Issue on Implementing Post-Conviction Risk Assessment*, 38 BEHAV. SCIS. & L. 187 (2020) (regarding post-conviction risk assessment).

⁶⁶ Starr, *supra* note 18, at 807–08; see John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 427–28 (2006) (claiming that valid risk factors, like past and present criminal behavior, indicate moral responsibility). *Contra* Sidhu, *supra* note 45, at 717 (suggesting that risk assessment tools in sentencing are also in tension with utilitarianist theories of punishment).

⁶⁷ See Harcourt, *Risk as a Proxy for Race*, *supra* note 62, at 237 (suggesting that risk assessment primarily centers on past criminal history, overlooking other factors that are more difficult to measure).

⁶⁸ See O’Hear, *supra* note 46, at 196–98, 204–06 (proposing strategies to maintain the legitimacy of judges in the eyes of defendants while utilizing risk assessment algorithms in sentencing).

⁶⁹ See Alyssa M. Carlson, Note, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L. REV. 303, 324–29 (2017) (advocating that private companies be mandated to adhere to public transparency standards within the criminal justice system).

⁷⁰ See Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1957 (2021) (contending that individuals should have the right to contest AI decisions); BRANDON L. GARRETT, DEFENDING DUE PROCESS: WHY FAIRNESS MATTERS IN A POLARIZED WORLD 145–50 (2025) (arguing against traditional “[b]lack-box AI risk assessment systems” in favor of a right to a more transparent “glass-box AI”).

opacity and complexity of algorithms and proprietary secrets.⁷¹ More broadly, algorithms force us to rethink the roles of humans in exercising the ultimate state power—to take one’s liberty. In the age of risk assessment tools, the decisional power is arguably shifted from frontline adjudicators, such as prosecutors, judges, and parole board members, to technical designers of algorithmic functions and processes.⁷² This raises meaningful questions related to the relationship between individuals and the state in the algorithmic era that will be addressed in Part III.⁷³

This Part introduces the widespread use of risk assessment algorithms in the criminal legal system and offers a broad overview of the scholarship in support of or in opposition to their increased utilization.⁷⁴ An analysis of such scholarship identifies several different areas of focus: (A) questions of design and assessment, including government and public participation in the design and assessment processes; (B) questions related to the algorithmic impact on proceedings and decision-makers, particularly in the context of governments and government-related administration, including issues related to reliance on technology; and (C) broader questions about institutional legitimacy in a democratic society, including questions about accountability, transparency, and individual rights.⁷⁵ Questions of AI regulation transcend all of these areas of focus.⁷⁶

⁷¹ O’Hear, *supra* note 46, at 196–98, 204–06; Slobogin, *supra* note 48, at 191.

⁷² It is commonly argued that algorithms do not dictate outcomes or specific decisions but rather guide decision-makers. See *State v. Loomis*, 881 N.W.2d 749, 764–65 (Wis. 2016) (holding that judges maintain the discretion and access to information to reject a risk assessment when deemed appropriate); *Malenchik v. State*, 928 N.E.2d 564, 573 (Ind. 2010) (stating that evidence-based assessment tools serve as supplementary aids in sentencing); David E. Patton, *Guns, Crime Control, and a Systemic Approach to Federal Sentencing*, 32 CARDOZO L. REV. 1427, 1456 (2011) (claiming that actuarial assessments are not “the determining factor in any given sentence”). For a discussion on whether there is a *right* to a human-made decision, see Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 651–53 (2020) (concluding that only the right to a well-calibrated machine decision should be acknowledged). Yet, even if human decision-makers keep their discretion, their roles could still be reshaped. See, e.g., Starr, *supra* note 18, at 862–64 (holding that the use of risk scores will inevitably affect sentencing even if solely informing, rather than replacing, judges). See generally Raja Parasuraman & Dietrich H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUM. FACTORS 381 (2010) (highlighting that decision-makers may become complacent, growing accustomed to placing unquestioning trust in algorithmic decision support systems). This phenomenon is often referred to as *automation bias*, where decision-makers abandon independent judgment and excessively depend on algorithmic recommendations, instead of exercising thoughtful discretion. Linda J. Skitka, Kathleen Mosier & Mark D. Burdick, *Accountability and Automation Bias*, 52 INT’L J. HUM.-COMPUT. STUD. 701, 701–03 (2000). Additionally, traditional arguments about the workloads of various criminal legal system participants raise similar concerns.

⁷³ See *infra* Part III.

⁷⁴ See *supra* Part I.

⁷⁵ For a discussion about scholarship on AI and the Law beyond the realm of criminal law, see Jonathon W. Penney, *AI and Legal Scholarship: Reflections on Evolution and Influences* (discussing trends in AI and the legal scholarship beyond risk assessment tools), in LEADING LEGAL DISRUPTION: ARTIFICIAL INTELLIGENCE AND A TOOLKIT FOR LAWYERS AND THE LAW 57, 57 (Giuseppina

An assessment of the literature, however, also exposes a common tendency among scholars to focus on specific risk assessment tools within the criminal process.⁷⁷ This work is undoubtedly important. Its siloed nature, however, limits us from thinking more holistically about the effects of the exponential growth of risk assessment tools on the criminal legal system as a whole. In particular, how these tools can affect foundational doctrinal principles that stand at the heart of substantive criminal law and are thus in tension with the values such principles aspire to advance.

This Article will advance this line of thinking. It argues that given the non-subjective logic of risk assessment tools, their introduction into the criminal legal system threatens to change core elements of modern criminal law, particularly the concept of subjective liability. The values on which the modern idea of criminal liability is premised—including democratic ideals of autonomy and personhood—are impugned by the introduction of algorithmic risk assessment tools. To fully develop this argument, the next Part will discuss the relevant criminal law doctrine.⁷⁸ In particular, it will offer a deep dive into the evolution of the fundamental concept of subjective liability, its philosophical underpinnings and—relatedly—the values it seeks to inject into the criminal legal system.

II. THE RISE OF SUBJECTIVE LIABILITY IN CRIMINAL LAW

Part II describes how the notion of subjective liability in criminal law has evolved over time, before placing this transformation within a broader conversation around Western liberal, democratic ideas, especially the autonomy and dignity of the individual.⁷⁹ Section A of this Part traces the shift in criminal law from a focus on broad concepts of moral culpability to subjective liability.⁸⁰ Section B of this Part demonstrates how the shift toward subjective liability has influenced modern legal practices and concepts, especially regarding the evolution of mens rea.⁸¹

D’Agostino, Aviv Gaon & Carole Piovesan eds., 2021); Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1326–37 (2019) (discussing the existing body of work on AI and the law); Catalina Goanta, Gijs van Dijk & Gerasimos Spanakis, *Back to the Future: Waves of Legal Scholarship on Artificial Intelligence* (discussing the growth of artificial intelligence as a subject of legal inquiry), in TIME, LAW, AND CHANGE: AN INTERDISCIPLINARY STUDY 327, 327–28 (Sofia Ranchordás & Yaniv Roznai eds., 2020).

⁷⁶ See Yonathan Arbel, Matthew Tokson & Albert Lin, *Systemic Regulation of Artificial Intelligence*, 56 ARIZ. ST. L.J. 545, 583–95 (2024) (advocating for a systemic regulation of AI systems).

⁷⁷ See, e.g., BRAYNE, *supra* note 34, at 30–36 (discussing the Los Angeles Police Department’s use of multiple risk assessment tools, including “TEAMS II” and another, unnamed tool designed by Palantir Technologies); see *supra* note 18 (detailing the use of risk assessment tools by the police, pretrial, in bail proceedings, and in sentencing).

⁷⁸ See *infra* Part II.

⁷⁹ See *infra* Part II.

⁸⁰ See *infra* Part II.A.

⁸¹ See *infra* Part II.B.

A. The Doctrinal Shift and Its Philosophical Foundations

For centuries now, legal and moral philosophers of criminal law, political theorists, and legal historians have been preoccupied with the concept of criminal responsibility, also referred to as liability or culpability.⁸² Their preoccupation is unsurprising given the centrality of this concept to the state's process of differentiating between individuals in deciding whether to limit their freedom.⁸³ This Article focuses on a particular set of questions related to the evolution of the concept in modern-day criminal law, and the philosophical foundations of such evolution.

Although scholars diverge on many questions related to criminal culpability that are beyond the scope of this Article, there seems to be a consensus regarding its evolution from the late eighteenth century through the twentieth century. Generally, and despite some terminological differences, scholars recognize a gradual theoretical and practical transition from a broader notion of moral culpability toward a subjective or individual one.

Farmer tellingly describes this process as the "rise of subjective liability," a transition from broad notions of moral blameworthiness to the exploration of "individual responsibility" as increasingly central to the modern criminal process.⁸⁴ Turner similarly conceptualized modern criminal law as recognizing "subjective fault" and moving away from historical concepts of "absolute liability".⁸⁵ Lacey also identified the transition from imprecise and generalized notions of responsibility to the modern "legal criteria of liability."⁸⁶ In particular, she documents a move from attribution based on bad character to one based on the assessment of conduct accompanied by the exercise of individual capacity.⁸⁷ In sum, existing scholarship indicates that there is little disagree-

⁸² For simplicity, this Article will use the terminology of responsibility, culpability, and liability interchangeably, as different scholars tend to do as well, all referring to a similar idea: the rules that allow "distinguishing guilt from innocence, legitimate from illegitimate conduct." LACEY, *supra* note 3, at 1. For additional work related to criminal responsibility, not necessarily in the context of this Article, see generally H.L.A. HART, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW (1968); MICHAEL MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW (1998); LARRY ALEXANDER & KIMBERLY KESSLER FERZAN WITH STEPHEN J. MORSE, CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW (2009).

⁸³ See LACEY, *supra* note 3, at 1 ("[P]roving that responsibility . . . is crucial to the core modality of criminal justice.").

⁸⁴ FARMER, *supra* note 2, at 181; see also Lindsay Farmer, *Criminal Responsibility and the Proof of Guilt* (arguing that the shift in the use of evidence in criminal law is "consistent with the developing emphasis on individual responsibility in nineteenth-century political and legal thought"), in MODERN HISTORIES OF CRIME AND PUNISHMENT 42, 53 (Markus D. Dubber & Lindsay Farmer eds., 2007).

⁸⁵ J.W.C. Turner, *The Mental Element in Crimes at Common Law*, in THE MODERN APPROACH TO CRIMINAL LAW 195, 195–99 (L. Radzinowicz & J.W.C. Turner eds. 1945); FARMER, *supra* note 2, at 186.

⁸⁶ LACEY, *supra* note 3, at 115.

⁸⁷ *Id.* at 112.

ment regarding the doctrinal shift in modern criminal law toward a search for subjective elements in the establishment of criminal culpability.⁸⁸

The shift toward subjective liability, however, was not confined to internal conceptual clarification within criminal law. Rather, scholars agree that it is tied to broader, modern Western philosophical notions of individual freedom in the welfare state.⁸⁹ Subjective liability, then, is often understood as formal recognition within criminal law of modern notions of “identity, freedom, voluntariness, choice, agency, [and] self-control.”⁹⁰ Indeed, Hart and Williams made these connections between advancing the centrality of subjective concepts such as *mens rea* and the idea of a “civilized legal system.”⁹¹ Furthermore, Hart and later Kadish connect subjective liability concepts and the modern state’s commitment to values of autonomy.⁹² Kadish suggests that society’s devotion to democratic values, personal dignity and self-determination, and recognizing the individual’s value as a whole person, is premised on the view that one is a responsible agent worthy of not only being praised for their conduct, but also being punished for their choices.⁹³ A conception of the individual “merely as alterable, predictable, curable or manipulatable things,” continues Kadish, “is the foundation of a very different social order indeed.”⁹⁴ Such an understanding of the individual does not only represent a preferred doctrinal vision of the criminal legal system. Rather, it also tends to legitimize this system in the eyes of its subjects, distinguishing it “as a system of justice rather than one of sheer force.”⁹⁵ The fairness of the system, according to Lacey, is “fundamentally

⁸⁸ *Id.* at 1; see *Negligence and the General Problem of Criminal Responsibility*, 81 YALE L.J. 949, 952 (1972) (“[I]t is now widely agreed that criminal punishment requires a subjective element”); Anthony Duff, *Searching for Criminal Responsibility: What Are We Looking for?* (recognizing the doctrinal shift toward subjective liability, while raising questions about its manifestation in practice), in *ON CRIME, SOCIETY, AND RESPONSIBILITY IN THE WORK OF NICOLA LACEY* 33, 41 (2010).

⁸⁹ FARMER, *supra* note 2, at 181–82; LACEY, *supra* note 3, at 8; Kadish, *supra* note 5, at 77.

⁹⁰ LACEY, *supra* note 3, at 4.

⁹¹ See FARMER, *supra* note 2, at 186–87 (discussing how both Hart and Williams contend that the advancement of *mens rea* is the “hall-mark of a civilized legal system” (quoting HART, *supra* note 82, at 187 n.23)).

⁹² See Sanford H. Kadish, *The Decline of Innocence*, 26 CAMBRIDGE L.J. 273, 287 (1968) (citing HART, *supra* note 82, at 182) (connecting subjective liability to a modern state’s commitment to values of personal autonomy).

⁹³ See *id.* (citing HART, *supra* note 82, at 182); see also Claire Finkelstein, *The Inefficiency of Mens Rea*, 88 CALIF. L. REV. 895, 895 (2000) (summarizing Kadish’s view on the connections between the doctrine of *mens rea* and the criminal legal system’s commitment to individual autonomy); *Morissette v. United States*, 342 U.S. 246, 250 (1952) (“It is as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil.”).

⁹⁴ Kadish, *supra* note 92, at 287 (quoting HART, *supra* note 82, at 182).

⁹⁵ LACEY, *supra* note 3, at 137.

premised on the idea that” individual capacities “of understanding and self-control were properly engaged at the time of the alleged offence.”⁹⁶

B. *Manifestation in Practice*

The broad, philosophical concept of subjective liability has borne itself out in modern criminal law through various standards and practices. Probably the most dominant manifestation of subjective liability has been the evolution of the foundational concept of mens rea, though additional concepts and practices similarly represent the transition, including defenses, accomplice liability, capacity, and the use of evidence during the criminal process.⁹⁷

Discussing the evolution of mens rea, for example, is illustrative. By the first half of the twentieth century, mens rea was mostly understood in Austinian terms—with intention and recklessness as distinct states of mind, both relevant to liability and the degree of culpability. This development reflected the growing recognition of the importance of individual responsibility to criminal law.⁹⁸ In essence, this modern approach to the mental element reiterated the aspiration to replace the idea of *moral* fault in the common law with a *descriptive-legal* fault.⁹⁹ Centuries of jurisprudential and scientific advancements had now offered more concrete ideas on how to accomplish such a transition. The idea was to focus on certain legally “provable” states of mind as a means of connecting between the presumed and actual psychological capacities of the legal subject.¹⁰⁰ As part of this process, the presumption that individuals intend the natural consequences of their actions was generally abandoned because it contained “a serious threat to any rational theory of intention.”¹⁰¹

The importance of separately proving act and intention gradually garnered more and more support from scholars and courts.¹⁰² For example, debates around the appropriateness of imposing criminal liability for negligence reflected the transition in the concept of mens rea and “the place of individual responsibility within a modern system of criminal law.”¹⁰³ Similarly, legisla-

⁹⁶ *Id.*

⁹⁷ *Id.* at 1; Duff, *supra* note 88, at 46–52.

⁹⁸ FARMER, *supra* note 2, at 181.

⁹⁹ *Id.* at 181–82; *see also* LACEY, *supra* note 3, at 112–15 (discussing how during the eighteenth-century, guilt determination was more “the application of broad community standards of responsibility and guilt” than “legal criteria of liability”).

¹⁰⁰ FARMER, *supra* note 2, at 182.

¹⁰¹ GLANVILLE L. WILLIAMS, CRIMINAL LAW: THE GENERAL PART 77–81 (1953). It should be noted, however, that in the American context, characterized by diverse state-level criminal legislation, the doctrine remains valid in several states.

¹⁰² *See, e.g.,* Woolmington v. DPP [1935] AC 462 (HL) 482 (appeal taken from Eng.) (creating a connection between the proof of mental state and presumption of innocence, and quashing a conviction due to the prosecution’s failure to prove *both* the fact of the killing and the intention of the accused).

¹⁰³ FARMER, *supra* note 2, at 185.

tion recognizing strict or absolute liability offenses was eventually criticized for its inherent tension with the increasingly accepted processes for realizing subjective liability.¹⁰⁴

Fletcher identified similar transitions in the U.S. context, noting a movement away from “manifest” culpability—that focuses analysis on a particular kind of “act”—to “subjective” culpability, hinging on the intent to “do” the act.¹⁰⁵ Fletcher identified such patterns, for example, in the law of theft and attempts.¹⁰⁶ Others have suggested similar transitions in other contexts, including sexual offenses.¹⁰⁷

Probably the most paradigmatic illustration of the transition to subjective liability in the United States was the creation of the American Law Institute’s highly influential Model Penal Code (MPC) in the early 1960s.¹⁰⁸ One of the main contributions of the MPC is its comprehensive and unequivocal endorsement of the centrality of the guilty mind in the process of assessing criminal culpability.¹⁰⁹ It first states that a “person is not guilty of an offense unless his liability is based on conduct that includes a voluntary act or the omission to perform an act of which he is physically capable,” and clearly establishes that one cannot be charged with a crime for accidentally causing an injury due to an involuntary spasm.¹¹⁰ But most importantly for our discussion, the MPC adopts an elemental approach under Section 2.02(1) that requires proof of culpability “with respect to each material element of the offense.”¹¹¹ Robinson

¹⁰⁴ See *id.* at 185–86 (discussing Turner’s view that absolute liability—as an objective standard—was a regressive step in modern-day criminal law, where a subjective standard should prevail).

¹⁰⁵ GEORGE P. FLETCHER, *RETHINKING CRIMINAL LAW* 115–17 (1978).

¹⁰⁶ *Id.* at 115. According to Fletcher, under the pattern of manifest criminality “a neutral third-party observer could recognize the activity as criminal even if he had no special knowledge about the offender’s intention.” *Id.* at 116. By contrast, subjective criminality assumes that the core of criminal conduct is the intention to violate a legally-protected interest. *Id.* at 115–23. Fletcher demonstrates this point through the analysis of property offenses and case law, but also homicide and additional crimes. *State v. Young*, 271 A.2d 569 (N.J. 1970); see also George P. Fletcher, *The Metamorphosis of Larceny*, 89 HARV. L. REV. 469, 490–91 (1976) (providing additional discussion on the distinction between manifest and subjective criminality).

¹⁰⁷ Guyora Binder & Robert Weisberg, *What Is Criminal Law About?*, 114 MICH. L. REV. 1173, 1184–85 (2016); see also Luis Chiesa, *The Model Penal Code, Mass Incarceration, and the Racialization of American Criminal Law*, 25 GEO. MASON L. REV. 605, 612–13 (2018) (discussing this evolution in different contexts).

¹⁰⁸ See Paul H. Robinson & Markus D. Dubber, *The American Model Penal Code: A Brief Overview*, 10 NEW CRIM. L. REV. 319, 320 (2007) (“Promulgated in 1962, the code prompted a wave of state code reforms in the 1960s and 1970s, each influenced by the Model Penal Code.”).

¹⁰⁹ MODEL PENAL CODE § 2.01(1) (A.L.I. 1962).

¹¹⁰ *Id.*; Herbert Wechsler, *On Culpability and Crime: The Treatment of Mens Rea in the Model Penal Code*, 339 ANNALS AM. ACAD. POL. & SOC. SCI. 24, 27–28 (1962); Gerhard O.W. Mueller, *On Common Law Mens Rea*, 42 MINN. L. REV. 1043, 1047–49 (1958).

¹¹¹ Wechsler, *supra* note 110, at 28 (quoting MODEL PENAL CODE § 2.02(1) (Tent. Draft No. 4, A.L.I. 1955); see Paul H. Robinson, *Mens Rea*, in *ENCYCLOPEDIA OF CRIME & JUSTICE* 995, 997 (2d ed., 2002) (discussing the Model Penal Code’s requirement for proof of culpability)).

notes that this elemental approach brought forth for the first time a thorough statement of the requisite culpability for an offense and allowed legislatures to reclaim the power to define what makes one criminally liable from the judiciary.¹¹² Such a vision aligns with processes developed to vindicate autonomy and dignity through the illumination of subjective liability.

Additional features of the shift toward subjective liability were also included in the MPC.¹¹³ For example, the dismantling of strict liability offenses.¹¹⁴ Wechsler notes that around the time of the adaptation of the MPC, there was a “disturbing phenomenon” of strict liability offenses that did not take into account the mental element but only whether the actor committed the act.¹¹⁵ Further, he believes the MPC attempted to “mount a frontal attack on such rejection of culpability as a prerequisite to criminal conviction,” with the idea that strict liability is an injustice and is immoral in penal law.¹¹⁶ The MPC does this by noting that only a violator with a culpable mental state may face legal sanction.¹¹⁷

In addition, the MPC is careful to address the additional mental states of recklessness and negligence, the former being when a person consciously disregards a “substantial risk,” whereas the latter is when one is unaware of a substantial risk of which they should be aware.¹¹⁸ Robinson argues the distinction between negligence and other states of culpability is pivotal to criminal law because other mental states require some form of awareness of the circumstances that make certain conduct criminal, whereas negligence does not.¹¹⁹ Furthermore, the MPC emphasizes the importance of subjectivity when determining recklessness, as evidenced by how the reasonable person standard used in determining recklessness or negligence is adjusted to “take account of the defendant’s ‘situation.’”¹²⁰

The MPC also offers some level of protection to those who are truly mistaken in their subjective belief that the crime they committed either was not a crime or was justifiable under the circumstances.¹²¹ One example is the doctrine of self-defense.¹²² Although the common law of self-defense traditionally requires *reasonable* (objective) *belief* in assessing the danger to one’s bodily integrity, the MPC offers this defense against murder for one who honestly and

¹¹² Robinson, *supra* note 111, at 997.

¹¹³ MODEL PENAL CODE § 2.02(1); Wechsler, *supra* note 110, at 32.

¹¹⁴ § 2.02(1); Wechsler, *supra* note 110, at 32.

¹¹⁵ Wechsler, *supra* note 110, at 31.

¹¹⁶ *Id.* at 32.

¹¹⁷ § 2.02(1); Wechsler, *supra* note 110, at 32.

¹¹⁸ § 2.02(2)(c)–(d); Robinson, *supra* note 111, at 1000–01.

¹¹⁹ Robinson, *supra* note 111, at 1001.

¹²⁰ *Id.*

¹²¹ MODEL PENAL CODE §§ 3.04, 3.09.

¹²² § 3.04.

mistakenly *believes* they needed to kill in self-defense.¹²³ The defense might not be available against MPC manslaughter or negligent homicide charges, however, if one was reckless or negligent in forming said belief.¹²⁴

Another illustration is the defense of extreme emotional disturbance (EMED), recognized by the MPC, that can reduce purposeful or knowing killings to manslaughter under the MPC.¹²⁵ Although to a certain extent this defense mirrors the common law provocation doctrine that traditionally adopted a “reasonable person” standard, the MPC has introduced a subjective standard, followed by an objective one, strengthening the subjective focus in determining liability in homicide cases.¹²⁶ Other examples that reflect the MPC’s adherence to subjective liability include its attempt liability doctrine that requires a “substantial step,”¹²⁷ and its dispensation with the requirement of actual breaking for the burglary offense.¹²⁸

In sum, historical analysis illuminates significant changes in the doctrinal and practical criminal law borne out through a process identified by scholars as the “rise of subjective liability.”¹²⁹ The rise of subjective liability enjoys internal criminal law justifications and is also closely aligned with broader Western ideas of the law and the connections between the state and its subjects in liberal democracies.

Therefore, subjective liability is not only a core component of modern criminal law, but it also offers an opportunity to assess the different values standing at the heart of the modern criminal legal system. Indeed, as the historical overview suggests, at least until the early 1970s, the preeminence of subjective liability represented a normative aspiration for a criminal legal system that recognized individual autonomy, agency, and dignity.¹³⁰

¹²³ *Id.*

¹²⁴ Wechsler, *supra* note 110, at 31.

¹²⁵ MODEL PENAL CODE § 210.3.

¹²⁶ The issue of how the MPC standard should be interpreted was raised in *People v. Casassa*, 404 N.E.2d 1310, 1312 (N.Y. 1980). The court, recognizing the subjective standpoint of the MPC, concluded:

[T]he determination whether there was reasonable explanation or excuse for a particular emotional disturbance should be made by viewing the subjective, internal situation in which the defendant found himself and the external circumstances as he perceived them at the time, however inaccurate that perception may have been, and assessing from that standpoint whether the explanation or excuse for his emotional disturbance was reasonable, so as to entitle him to a reduction of the crime charged from murder in the second degree to manslaughter in the first degree.

Id. at 1316.

¹²⁷ MODEL PENAL CODE § 5.01; Chiesa, *supra* note 107, at 612.

¹²⁸ FLETCHER, *supra* note 105, at 126–27.

¹²⁹ FARMER, *supra* note 2, at 181.

¹³⁰ Kadish, *supra* note 5, at 77; LACEY, *supra* note 3, at 9. The historical evolution toward subjective liability is clearly established and reflects an aspiration to advance internal and external values as

As is often the case, however, normative and doctrinal aspirations and practical applications do not always go hand in hand. Indeed, this Article does not seek to romanticize the pre-algorithmic criminal legal system. Rather, it acknowledges the disparity between the aspirational vision of subjective liability and its real-world application, mostly through two broad frameworks: the phenomenon of vanishing trials, and the paradigm identified as the “new penology.”¹³¹

A system that purports to function on the premise that the presence or absence of criminal liability is outcome-determinative on the imposition of criminal sanctions should strive to support legal mechanisms that will increase the probability of assessing this liability. Traditionally, the adversarial Anglo-American system has considered the trial, and more precisely the jury trial, the main tool for “true” liability determination.¹³² The reality of the criminal legal system, however, is just the opposite. Just as subjective liability has achieved relatively widespread jurisprudential and doctrinal recognition, as discussed above, the criminal legal system has concurrently normalized a mechanism that is inherently in opposition to the real quest for liability: plea bargaining.

Several scholars have explored the historical evolution of plea bargains in the Anglo-American system, and they agree that this phenomenon is not new to the twentieth century. They identify the mid-nineteenth century as the era when plea bargaining became a more utilized form of disposing with criminal cases.¹³³ The rise of plea bargaining could likely be attributed to a host of rea-

discussed above. *See supra* Part II. At the same time, one should recognize that some doctrines within criminal law still reflect a diversion from principles of individual culpability, particularly in the American context given the legislative variations across states. For example, the Pinkerton doctrine (that was rejected by the MPC) was adopted by several states, or accomplice liability based on the doctrine of natural and probable consequences. This is true even in states that generally adopted MPC provisions and, in doing so, aspired to advance principles of subjective liability.

¹³¹ Malcolm M. Feeley & Jonathan Simon, *The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications*, 30 *CRIMINOLOGY* 449, 449 (1992).

¹³² *See* ISSA KOHLER-HAUSMANN, *MISDEMEANORLAND: CRIMINAL COURTS AND SOCIAL CONTROL IN AN AGE OF BROKEN WINDOWS* POLICING 71–72 (2018) (discussing the adjudicative model of traditional criminal courts that seek “to determine whether the defendant in fact committed the criminal act of which she is accused” by “receiving and evaluating ‘proofs and reasoned arguments’ in the form of evidence about the facts of the case and information about the legal standards that pertain to an accusation of guilt for a specific offense”). To be clear, I do not intend to idealize trials in the American criminal legal system (or any other system for that matter). In fact, as others have argued, I posit that criminal trials suffer from significant issues to an extent that meaningfully challenges their current institutional capacity to truly achieve their stated goals. Regardless, at least conceptually, trials in the adversarial system are considered the core tool to investigate one’s criminal liability, a setting where the presumption of innocence can be investigated.

¹³³ *See* John H. Langbein, *Understanding the Short History of Plea Bargaining*, 13 *LAW & SOC’Y REV.* 261, 261, 269 (1979) (explaining how the American justice system’s preference for jury trials partially led to the emergence of plea bargains); Malcolm M. Feeley, *Legal Complexity and the Transformation of the Criminal Process: The Origins of Plea Bargaining*, 31 *ISR. L. REV.* 183, 187–88 (1997); William Ortman, *When Plea Bargaining Became Normal*, 100 *B.U. L. REV.* 1435, 1441

sons, but in part it was related to the transformation of the criminal legal system from an institution dominated by private prosecution¹³⁴ to one led by the central government through state prosecution.¹³⁵ During the early twentieth century, however, legal scholars and commissioners who discovered the meaningful presence of plea bargaining in the system were “aghast by the phenomenon.”¹³⁶ Between the 1930s and 1960s, Ortman suggests, plea bargaining was present in the criminal legal system, but was by no means accepted by the legal profession (or “normal” in Ortman’s language).¹³⁷ By the late 1960s and early 1970s, however, the use of plea bargaining proliferated and was universally praised by members of the legal profession.¹³⁸ Since then, plea bargains have dominated the American criminal legal system.

As the Supreme Court previously recognized, “criminal justice today is for the most part a system of pleas, not a system of trials.”¹³⁹ Only recently, the 2023 Plea Bargain Task Force Report stated that, “trials have become rare legal artifacts in most U.S. jurisdictions—and even nonexistent in others.”¹⁴⁰ For example, the Task Force reported on research conducted by Pew Research Center showing that ninety percent of criminal cases in the federal system end with a plea agreement.¹⁴¹ In state systems the problem is the same. For instance, states like New York, Pennsylvania, and Texas have trial rates of less

(2020); Malcolm M. Feeley & Rosann Greenspan, *The Long History of Plea Bargaining*, in RESEARCH HANDBOOK ON PLEA BARGAINING AND CRIMINAL JUSTICE 441, 441 (Máximo Langer, Mike McConville & Luke Marsh eds., 2024).

¹³⁴ See ALLEN STEINBERG, *THE TRANSFORMATION OF CRIMINAL JUSTICE: PHILADELPHIA, 1800–1880*, at 1–2 (1989) (discussing the centrality of private prosecution to Philadelphia’s law enforcement and legal culture in the mid-nineteenth century). Steinberg defined private prosecution as “one citizen taking another to court without the intervention of the police.” *Id.* at 1.

¹³⁵ See *id.* at 24–25, 119–21 (detailing the shift from local- to state-dominated prosecution in Philadelphia in the nineteenth century); LAURA F. EDWARDS, *THE PEOPLE AND THEIR PEACE: LEGAL CULTURE AND THE TRANSFORMATION OF INEQUALITY IN THE POST-REVOLUTIONARY SOUTH* 6, 10 (2009) (describing the gradual centralization of the Southern legal system into one dominated by state, rather than local, law).

¹³⁶ Ortman, *supra* note 133, at 1437.

¹³⁷ *Id.* at 1437–38.

¹³⁸ *Id.* at 1437–39; see Albert W. Alschuler, *The Prosecutor’s Role in Plea Bargaining*, 36 U. CHI. L. REV. 50, 51 (1968) (“[T]he legal profession now [in 1968] seems as united in its defense of plea negotiation as it was united in opposition less than a half-century ago.”).

¹³⁹ *Lafler v. Cooper*, 566 U.S. 156, 170 (2012).

¹⁴⁰ THEA JOHNSON, A.B.A. CRIM. JUST. SECTION, *PLEA BARGAIN TASK FORCE REPORT* 14 (2023), https://www.americanbar.org/content/dam/aba/administrative/criminal_justice/reports/plea-bargain-tf-report.pdf [<https://perma.cc/BNM5-RG47>]; see Chesa Boudin & Eric S. Fish, *Towards Pretrial Criminal Adjudication*, 66 B.C. L. REV. 1135, 1137, 1146 (2025) (“Ours is no longer a system of criminal trials.”).

¹⁴¹ JOHNSON, *supra* note 140, at 36 n.2 (citing John Gramlich, *Only 2% of Federal Criminal Defendants Go to Trial, and Most Who Do Are Found Guilty*, PEW RSCH. CTR., (June 11, 2019)).

than three percent, and “[s]ome jurisdictions in the country report not having had a criminal trial in years.”¹⁴²

A full summary of criticism of the institution of plea bargaining is beyond the scope of this Article. One key critique, however, is that plea bargains increase the likelihood that innocent suspects will plead guilty.¹⁴³ Furthermore, plea bargains replace the trial and thereby diminish the Anglo-American ideal of assessing criminal culpability through an adversarial process. Plea bargains, then, are in sharp opposition to modern, subjective ideas of determining criminal liability. Instead, plea bargains either punish the innocent or, at least, strip criminal subjects of their right to an individualized exploration of their criminal liability.

Therefore, when thinking about the idea—or ideal—of criminal liability, we should diverge from our traditional and technical view of culpability as being determined at trial. Instead, we should recognize that criminal culpability is being determined in other, earlier stages of the criminal process.¹⁴⁴

The phenomenon of the “vanishing trial,” illustrated most explicitly through the mechanism of plea bargains, reflects a much deeper change in our penal system, one famously identified by Feeley and Simon more than thirty years ago as the “new penology.”¹⁴⁵ According to Feeley and Simon, this transition, that extends through all parts of the system, reflects its abandonment of and lack of interest in liability concepts such as guilt or innocence.¹⁴⁶

As part of that change, Feeley and Simon, and additional scholars building upon their work, have identified a transition from traditional adjudicative models seeking individual guilt or punishment to group management based on risk, also known as the managerial model.¹⁴⁷ The transition identified by the

¹⁴² *Id.* (citing Jeffrey Q. Smith & Grant R. MacQueen, *Going, Going, but Not Quite Gone: Trials Continue to Decline in the Federal and State Courts. Does It Matter?*, 101 JUDICATURE 4, 32–34 (2017)).

¹⁴³ See, e.g., Oren Gazal-Ayal & Avishalom Tor, *The Innocence Effect*, 62 DUKE L.J. 339, 341 (2012) (claiming that “plea bargains can lead innocent defendants to plead guilty”); Samuel R. Wiseman, *Bail and Mass Incarceration*, 53 GA. L. REV. 235, 252 (2018) (describing how cash bail systems lead to increased rates of pretrial detention among poor defendants, and this in turn can increase rates of wrongful convictions); Daniel S. Medwed, *Emotionally Charged: The Prosecutorial Charging Decision and the Innocence Revolution*, 31 CARDOZO L. REV. 2187, 2188–92 (2010) (describing insufficient limits on prosecutorial charging decisions as a factor in wrongful convictions).

¹⁴⁴ See, e.g., Anna Roberts, *Arrests as Guilt*, 70 ALA. L. REV. 987, 989 (2019) (discussing how “in a wide range of ways, in a wide range of contexts, and in the assumptions of a wide range of people, arrests appear to be fused with guilt”). Similar logic, as I will discuss below, applies to other stages of the criminal process as early as the initial police investigation. See *infra* Part III.

¹⁴⁵ See Feeley & Simon, *supra* note 131, at 449, 452–58 (explaining the elements of the new penology).

¹⁴⁶ *Id.* at 449–51.

¹⁴⁷ *Id.* at 449, 452–54; Kohler-Hausman, *supra* note 132, at 71–72; see Sandra G. Mayson, *Collateral Consequences and the Preventive State*, 91 NOTRE DAME L. REV. 301, 331 (2015) (applying a similar framework in the context of collateral consequences); Alexandra Natapoff, *Misdemeanors*, 11

“new penology” thus reflects not only a decline in the criminal legal system’s interest in the idea of criminal *liability*, but also in the *subject*; the individual before that same system.

The modern American criminal legal system evolved to be more subject-focused, aiming to place the individual at the center of the criminal process. Feeley and Simon agree with this assertion, suggesting that the “old” penology indeed focused on individuals as the unit of analysis. The new penology, however, is concerned with something completely different, namely “techniques to identify, classify, and manage groupings sorted by dangerousness.”¹⁴⁸ Relatedly, the new penology prioritizes concepts like “probability” and “risk,” while “old” concepts such as “retributive judgment” and “clinical diagnosis” are pushed aside.¹⁴⁹

Since Feeley and Simon’s seminal work, many have further observed the presence of the new penology throughout the criminal legal system, from “policing tactics[,] . . . risk assessments, . . . collateral consequences, electronic monitoring[,] . . . arrests,” and criminal courts.¹⁵⁰

Indeed, algorithms can be characterized as the newest frontier in the new penology paradigms, representing a sweeping decline in the ideas underpinning subjective liability. This Article, however, offers a different take on the algorithmic revolution in the criminal legal system. It argues that algorithmic risk assessment tools likely represent the most radical shift from the ideal of subjective liability in the criminal legal system.

Risk assessment tools share some traits with the broader actuarial trends discussed above but also have unique features that deserve special attention when thinking about their impact.¹⁵¹ *First*, sophisticated risk assessment methodologies allow these tools to expand into new domains of the criminal process, such as prosecution and certain new facets of police work. *Second*, historical utilization of statistical tools in different stages of the criminal process did not involve the same level of sophistication or data mining processes featured in current models. This in turn has the potential to increase disparate impact,¹⁵²

ANN. REV. L. & SOC. SCI. 255, 263 (2015) (observing that our system of misdemeanors is “less about establishing guilt under law”); Evelyn Malavé, *Criminal Courteaucracy*, 61 AM. CRIM. L. REV. 1205, 1212–17 (2024).

¹⁴⁸ Feeley & Simon, *supra* note 131, at 451–52.

¹⁴⁹ *Id.* at 450–51.

¹⁵⁰ Malavé, *supra* note 147, at 1214–15.

¹⁵¹ See *supra* Part II.

¹⁵² See Selbst, *supra* note 39, at 115–16 (emphasizing the differences between statistical predictions before the algorithmic age and new big data mining predictions). Among the differences Selbst discusses is the lack of commitment—or at least, less of a commitment—to the theory of predictive risk in the algorithmic age. *Id.* See generally VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* 55 (2013) (discussing how, in the age of big data, data sets are so big that researchers no longer have to limit themselves to a hypothesis-driven approach, but instead can allow computational analysis to identify

minimize transparency, and quell individual and collective opportunities to scrutinize decisions.¹⁵³ *Third*, the growing reliance on algorithms is reshaping the temporal dimension of institutional decision-making. Specifically, the capacity to evaluate potential guilt, sometimes even before any act has taken place, marks a shift from *ex post* to *ex ante* assessments of criminality, along the lines of what Steiker has defined as the “preventive state.”¹⁵⁴ This transformation is fueled by political pressures to eliminate any conceivable risk before it materializes, a tendency that ultimately heightens liberty deprivation.¹⁵⁵ *Fourth*, algorithms exemplify the increasingly attractive allure of scientific objectivity. The sophistication, buzz, terminology, and private money devoted to their expansion have contributed to—often misleading—narratives of accuracy and efficiency.¹⁵⁶ These narratives promote governmental trust in algorithmic outputs and heavier reliance on their recommendations, while simultaneously decreasing scrutiny. As such, thinking about algorithmic risk assessments as a natural continuation of existing trends seems to understate their transformative impact on criminal law—not only technically, but, as this Article argues, conceptually. The next Part will explain how and why.¹⁵⁷

III. THE NON-SUBJECTIVE ALGORITHMIC LOGIC

Part III illustrates the tension between notions of subjective liability core to the Western, liberal ideas of a democratic criminal legal system and the de-individualizing nature of algorithmic risk assessments grounded in general categorization.¹⁵⁸ Section A of this Part discusses the conflict between increasingly prevalent risk assessment algorithms and the traditional emphasis on subject-

the best hypothesis); Rob Kitchin, *Big Data, New Epistemologies and Paradigm Shifts*, BIG DATA & SOC’Y, Apr.–June 2014, at 1, 1–2 (analyzing how big data is challenging epistemologies across different disciplines, including “the sciences, social sciences and humanities”).

¹⁵³ LEADERSHIP CONF. ON CIV. & HUM. RTS. ET AL., PREDICTIVE POLICING TODAY: A SHARED STATEMENT OF CIVIL RIGHTS CONCERNS (Aug. 31, 2016), <https://assets.aclu.org/live/uploads/document/JointStatementPredictivePolicing17Orgs.pdf> [<https://perma.cc/SMT6-4VW9>].

¹⁵⁴ Carol S. Steiker, *The Limits of the Preventive State*, 88 J. CRIM. L. & CRIMINOLOGY 771, 774 (1998); see also Hideyuki Matsumi & Daniel J. Solove, *The Prediction Society: AI and the Problems of Forecasting the Future*, 2025 U. ILL. L. REV. 1, 11 (“[A]lgorithms have taken prediction to a new level of systemization, data usage, and prevalence The use of modern algorithms . . . has exacerbated these problems [problems related to prediction discussed in their article] to an alarming degree.”).

¹⁵⁵ See Eric S. Janus, *The Preventive State, Terrorists and Sexual Predators: Countering the Threat of a New Outsider Jurisprudence*, 40 CRIM. L. BULL. 576, 576–80 (2004) (discussing the concept of “radical prevention” and illustrating it in the context of terrorism prevention). I argue that risk assessment algorithms offer new opportunities for those wishing to enhance this approach.

¹⁵⁶ See Erin Collins, *Abolishing the Evidence-Based Paradigm*, 48 BYU L. REV. 403, 407–08 (2022) (challenging the “appearance of objectivity” often used to justify the expansion of risk assessment tools in the criminal legal system); Eaglin, *supra* note 18, at 64–65.

¹⁵⁷ See *infra* Part III.

¹⁵⁸ See *infra* Part III.

tive liability in criminal law.¹⁵⁹ Section B of this Part highlights the tension between the use of risk assessment algorithms and liberal democratic values, especially regarding respect for the autonomy and dignity of the individual.¹⁶⁰ Section C of this Part reveals how risk assessment algorithms have fundamentally altered many different aspects of the criminal legal system.¹⁶¹

A. The Algorithmic Logic and the Clash with the Individual

What are algorithms? As discussed earlier, despite their sophisticated terminology, we should think about risk assessment tools in simple terms as “statistical and computer driven decision-making processes.”¹⁶² The idea behind such data-driven and actuarial risk assessment processes is to predict the risk of recidivism posed by individuals based upon their correlation with indicative factors in existing data.¹⁶³ The heart of the algorithmic work is thus to *predict* how one will behave based on how *others* have *behaved* in the past. Although these tools aspire to measure an individual’s risk of recidivism, they do not predict that individual’s propensity for risk, but rather proffer predictions based on an obscure individual with characteristics that can resemble those of the individual under exploration without any “reference to actual culpability.”¹⁶⁴

The focus of the algorithmic investigation, by definition, is not to “faithfully portray the underlying flesh-and-blood individual . . . but to facilitate their classification . . . through the identification of high-level behavioral patterns.”¹⁶⁵ Therefore, subjects assessed by algorithmic tools are deprived of

¹⁵⁹ See *infra* Part III.A.

¹⁶⁰ See *infra* Part III.B.

¹⁶¹ See *infra* Part III.C.

¹⁶² SLOBOGIN, *supra* note 32, at vii.

¹⁶³ Collins, *supra* note 18, at 57. Actuarial judgment is the type of judgment predicated on statistical inference from data, as opposed to clinical judgment, that is exercised by humans applying intuition and inductive reasoning. See Robyn M. Dawes, David Faust & Paul E. Meehl, *Clinical Versus Actuarial Judgment*, 243 SCIENCE 1668, 1668–69 (1989) (distinguishing the two methods of judgment). Actuarial risk assessment thus refers to the use of data and statistics to predict risk from offenders (or even potential offenders). See Richard F. Lowden, Note, *Risk Assessment Algorithms: The Answer to an Inequitable Bail System?*, 19 N.C. J.L. & TECH. 221, 230–31 (2018) (describing the recent surge in the use of risk assessment algorithms for pretrial decisions); Collins, *supra* note 18, at 63–64; CHRISTOPHER SLOBOGIN, PROVING THE UNPROVABLE: THE ROLE OF LAW, SCIENCE, AND SPECULATION IN ADJUDICATING CULPABILITY AND DANGEROUSNESS 101 (2007).

¹⁶⁴ Collins *supra* note 18, at 64–65; Renée Jorgensen, *Algorithms and the Individual in Criminal Law*, 52 CANADIAN J. PHIL. 61, 63 (2022).

¹⁶⁵ See Katrina Geddes, *The Death of the Legal Subject*, 25 VAND. J. ENT. & TECH. L. 1, 3–5 (2023) (offering an account of the harms stemming from the use of algorithmic tools); see also Jorgensen, *supra* note 164, at 63–64 (“[S]everal legal theorists, philosophers, and judges have objected that inferences of this form functionally make it a ‘crime to belong to a reference class,’ violating the right to be ‘treated as an individual.’” (quoting Mark Colyvan, Helen M. Regan & Scott Ferson, *Is It a Crime to Belong to a Reference Class?*, 9 J. POL. PHIL. 168, 168 (2001))). In her intriguing work, Jorgensen further offers a particular interpretation of that right. Jorgensen, *supra* note 164, at 64.

their traditional place as an individual before the system and instead transformed into a predictable object “constructed from correlations identified in population-level data.”¹⁶⁶ This is what this Article refers to as the “algorithmic logic.” Although this logic can be traced back to earlier periods, broadly recognized as the “actuarial age,”¹⁶⁷ the recent expansion of algorithmic risk assessments pose particular challenges to the subject within the criminal legal process.¹⁶⁸ Indeed, this is the same legal subject that, as we saw in Part II, became the heart of substantive criminal liability during the twentieth century.¹⁶⁹ The transition to subjective liability is now well-embedded in various parts of substantive criminal law, such as through mens rea, property offenses, the law of attempts, sexual offenses, homicide, self-defense, and more.¹⁷⁰ The generality of the algorithmic logic is thus strikingly at odds with the substantive processes of modern criminal law developed out of philosophical notions of dignity and autonomy.

Although the rise of subjective liability was rooted in an aspiration to close “the gap between the presumed capacities of the legal subject and their actual psychological capacities,”¹⁷¹ algorithmic logic necessarily fails to investigate an individual’s subjective intention. Instead, it suggests that subjective intention can be constructed statistically from the history of the subject’s statistical equivalences.¹⁷² Predictive tools do not focus on unique individuals. In fact, they cannot, as their statistical power builds on the use of large numbers of individuals and their relationships.¹⁷³ Consequently, the algorithmic logic no longer requires, nor has any interest, in the individual’s lived experiences, thoughts,

¹⁶⁶ Geddes, *supra* note 165, at 16; see also JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 68–70 (2019) (discussing the use of population-level data in the context of financial markets as “markets for the outputs of data refineries”).

¹⁶⁷ See generally Jonathan Simon, *The Ideological Effects of Actuarial Practices*, 22 LAW & SOC’Y REV. 771 (1988) (discussing the growth of actuarial practices over the twentieth century and the dangerous ideological effects of this growth, particularly among historically disempowered classes like women).

¹⁶⁸ See Mireille Hildebrandt, *Proactive Forensic Profiling: Proactive Criminalization?* (discussing how profiling technologies “unveil a heteronomous subject rather than the autonomous agent we like to think we are”), in THE BOUNDARIES OF THE CRIMINAL LAW 113, 125 (R.A. Duff et al. eds., 2010); see Geddes, *supra* note 165, at 26–27 (describing how the rise of algorithmic risk assessments have caused the “death of the legal subject” by silencing subjects through this algorithmic logic).

¹⁶⁹ See *supra* Part II.

¹⁷⁰ Chiesa, *supra* note 107, at 607, 610–15.

¹⁷¹ FARMER, *supra* note 2, at 182.

¹⁷² Antoinette Rouvroy & Thomas Berns, *Algorithmic Governmentality and Prospects of Emancipation: Disparateness as a Precondition for Individuation Through Relationships?*, 177 RÉSEAUX 163, 169–71 (Liz Carey-Libbrecht trans., 2013); Geddes, *supra* note 165, at 5. See generally Michal Buchhandler-Raphael, *Overmedicalization of Domestic Violence in the Noncarceral State*, 94 TEMP. L. REV. 589, 638–39 (2022) (discussing entrenched biases in risk assessment algorithms in criminal and non-criminal contexts).

¹⁷³ See Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 579–80 (2021) (offering a theory of data as social relations to inform data-governance law).

beliefs, moral views, or explanations. Instead, algorithmic risk assessment tools look for patterns of behavior based on external, objective, and institutional data, working under the assumption that the available data offers a better understanding of the subject than the subject themselves.¹⁷⁴ In any event, much of that unique, individual data, such as lived experiences, past traumas, beliefs, and feelings, are not—and probably cannot—be measured by risk assessment tools. Therefore, the algorithmic logic presumes that qualitative, individualized data about a “subject” is negligible and necessarily replaces theories of autonomy and individuality with—purported—objectivity and generalization.

B. The Algorithmic Logic and the Clash with Liberal Democratic Values

The algorithmic conception of the individual in the criminal process is not only in deep tension with processes for determining subjective liability, but also pushes against the broader values that inspired the criminal legal system to transition toward subjective liability. As discussed in Part II, modern, democratic ideas of autonomy and respect for the individual as a capable and responsible subject were closely linked with the substantive transition in criminal law toward a subjective liability-centered system.¹⁷⁵ Predictive decisions made using statistical instruments, however, such as algorithmic risk assessment tools, are inherently in tension with individual autonomy.¹⁷⁶

First, the algorithmically-constructed subject in the criminal legal process reflects an unknown individual that is a statistical representation of others, but not necessarily of itself—that is, the actual criminal defendant. As such, the individual is perceived as a predictable thing rather than an autonomous human being.¹⁷⁷

Second, when the core of the criminal investigation shifts from an individual’s capacities, interests, views, and beliefs to an allegedly more reliable objective data analysis, that individual’s autonomy before the system is irrepa-

¹⁷⁴ Laurence Barry, *The Rationality of the Digital Governmentality*, 23 J. FOR CULTURAL RSCH. 365, 369 (2019). *Contra* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 22 (2007) (suggesting that the actuarial “impulse” is premised on a “desire to individualize” (emphasis omitted)). Whether the premise on which actuarial methods were advanced originated in some interest in the individual (a contested idea by itself), this Article contends that, de facto, algorithmic risk assessment tools diminish individual subjectivity. *See, e.g.*, Barbara D. Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L.J. 1408, 1414 (1979) (contending that the use of predictive statistics infringes upon respect for individual autonomy); *supra* Part II.B. (providing additional discussion).

¹⁷⁵ *See supra* Part II. As Kadish describes it, modern values of liberal democracies adopt a view of the subject as a “responsible agent entitled to be praised or blamed depending upon . . . free choice of conduct.” Kadish, *supra* note 92, at 287.

¹⁷⁶ Underwood, *supra* note 174, at 1414.

¹⁷⁷ *Id.*

rably compromised. The less control an individual has over the data used for prediction, the greater the decline in their autonomy.¹⁷⁸ As such, use of unchangeable or less flexible factors undermines one's autonomy.¹⁷⁹ Changeable factors reflect a subject's unique qualities, including their individuality, emotions, lived experiences, and beliefs. These qualitative factors cannot be reproduced by an algorithmic statistical model, illustrating why the tension between subjective liability and algorithmic logic is heightened when predicting risk for suspects and defendants of color. Many of the factors assessed by algorithms reflect broad social structures that criminalize communities of color based on these same "objective" factors, leaving no space for the amplification of individual voices.¹⁸⁰ Further, because the algorithmic logic treats persons as if they have the characteristics of the group they match, they may end up being "normalized into" this profile, a particular concern for marginalized communities already suffering from socially normalized negative responses.¹⁸¹ This issue will be addressed below.¹⁸²

Third, predictive data is inherently deterministic, suggesting that how individuals behave does not depend on their freely made choices, but instead on how certain objective categories are distributed in their algorithmic-induced self. Therefore, the algorithmic logic suggests that individuals are powerless to alter their statistically predefined path, and this takes away their capacity to be outliers.¹⁸³ In a more recent work more directly related to the use of data-driven algorithmic tools, Cen and Raghavan express similar concerns and call for recognizing the right to be an exception to a data-driven rule.¹⁸⁴ This vision of individuals under the algorithmic model conflicts with the autonomous vi-

¹⁷⁸ *Id.* Again, this is not to suggest that the pre-algorithmic system did particularly impressive work in exploring these components, but the aspirations behind notions of subjective liability at least preserved their relevance to the guilt determination process.

¹⁷⁹ *Id.* at 1436, 1442. It should be noted that traditionally, more risk assessment algorithms established risk based on unchangeable factors, such as race, gender, and ethnicity, but neglected to include changeable—hence more in control—factors. This is changing, however, in part due to academic work on the topic, with race, for example, being systematically excluded as a factor. Note, however, that some other categories that are still being used in assessments serve as a proxy for race. See Sidhu, *supra* note 45, at 718–22 (discussing factors used in sentencing risk assessment tools); *infra* Part III.C.4.

¹⁸⁰ Harcourt, *Risk as a Proxy for Race*, *supra* note 62, at 237–38.

¹⁸¹ Hildebrandt, *supra* note 168, at 133; see *infra* Part III.C.4.

¹⁸² See *infra* Part III.C.4. (discussing the algorithmic logic's effect on racially marginalized communities).

¹⁸³ Geddes, *supra* note 165, at 30; David T. Wasserman, *The Morality of Statistical Proof and the Risk of Mistaken Liability*, 13 CARDOZO L. REV. 935, 943 (1991); see also Collins, *supra* note 18, at 61 (arguing that the use of algorithmic risk assessment was intended to "guide decisions about how to administer punishment, not about how much punishment is due," thus illustrating the deep tension with individual autonomy).

¹⁸⁴ See Sarah H. Cen & Manish Raghavan, *The Right to Be an Exception to a Data-Driven Rule*, ARXIV, Dec. 28, 2022, at 1, 5–8, 12, 15, <https://arxiv.org/pdf/2212.13995v1> [<https://perma.cc/XVG9-DYYZ>] (tying the recognition of the right to be an exception to ideas of dignity).

sion of the individual under the law in liberal democracies.¹⁸⁵ To be clear, predictive tools take away one's agency regardless of how they decide to behave. By predetermining a subject's propensity to offend, risk assessment tools also remove the long-standing criminal requirement that a subject make a decision to act dangerously.¹⁸⁶ Regardless of whether they should be punished for such behavior, the algorithmic logic strips their ability to act based on their free will and imposes punishment before criminal action has been taken.¹⁸⁷

Fourth, it is difficult for individuals to contest algorithmic decision-making due to its opacity and sophistication, and, above all, big-data's authority. As for the latter point, there is no doubt that in the modern data-enhanced world, individuals have limited capacity and/or opportunity to contest data-driven evidence. Data-driven assessments are considered objective as a product of scientific exploration, while individual experiences, beliefs, and explanations are considered inferior, anecdotal evidence. In an algorithmic system, such evidence—that distinguishes one individual from another—is heavily discounted because it cannot be factored into the algorithm's sweeping prescriptions.

Opportunities to contest algorithmic decision-making are slim because there are very few built-in mechanisms for individuals to challenge those algorithmic decision-making processes.¹⁸⁸ Individuals do not have an institutional option to offer or explain the reasons behind their behavior or provide context that would warrant exceptions to machine rules normally applied to their circumstances.¹⁸⁹ Even if such opportunities existed,¹⁹⁰ challengers would face meaningful hurdles as studies have shown that stakeholders in the criminal legal system tend to rely heavily on algorithmic decisions, making challenges

¹⁸⁵ Kadish, *supra* note 5, at 65, 77. Algorithmic risk assessments deny human unpredictability. See Hildebrandt, *supra* note 168, at 132–33 (contending that actuarial techniques overdetermine human behavior and accordingly leave “no room for indeterminate spontaneous discontinuities”). Hannah Arendt referred to the individual “right” to remain unpredictable as natality. HANNAH ARENDT, *THE HUMAN CONDITION* 9 (1958). Judith Butler referred to this as the constitutive opacity of the self. JUDITH BUTLER, *GIVING AN ACCOUNT OF ONESELF* 19–20 (2005); see also Matsumi & Solove, *supra* note 154, at 30–32 (critiquing algorithmic predictions as creations that make, rather than predict, the future and, in doing so, inhibit individual autonomy).

¹⁸⁶ Sidhu, *supra* note 45, at 713–15.

¹⁸⁷ Hildebrandt, *supra* note 168, at 134–35.

¹⁸⁸ Ngozi Okidegbe, *To Democratize Algorithms*, 69 UCLA L. REV. 1688, 1704 (2023). There is not only difficulty in contesting the decisions themselves but also the initial decision to introduce algorithmic tools into a system, and later, the particular design. See *id.* at 1705–10 (discussing the different ways algorithms harm democratic participation); Jessica M. Eaglin, *When Critical Race Theory Enters the Law and Technology Frame*, 26 MICH. J. RACE & L. 151, 161–63 (2021) (discussing the social exclusion caused by algorithm use and development).

¹⁸⁹ See Kaminski & Urban, *supra* note 70, at 1957 (making the case for recognizing the right to contest AI); *infra* Part IV (providing additional discussion).

¹⁹⁰ See *infra* Part IV (suggesting some potential venues to allow these opportunities).

to algorithmic authority difficult, if not futile.¹⁹¹ Furthermore, excessive trust in data-based algorithmic tools can potentially lead agencies in the criminal legal system to move away from individualized elements that contradict the statistical assessment. This could happen even if such opportunities—like sending letters of support for defendants at hearings—do exist, or at least existed, in the past.¹⁹² These limitations exclude the subject from the algorithmic decision-making process, thereby compromising their individuality and depriving them of autonomy.¹⁹³ The legal framework this Article proposes in Part IV is an attempt to address these concerns.¹⁹⁴

So far, this Article has discussed two distinct evolutions of modern criminal law and the criminal legal system.¹⁹⁵ On the one hand, it examines the rise of subjective liability as a substantive feature of modern criminal law that has manifested through several practical developments.¹⁹⁶ On the other hand, it analyzes the expansion of algorithmic risk assessment tools that are in inherent tension with subjective liability to almost all stages of the criminal legal process.¹⁹⁷ This Article contends that although some evidence of this transition can be found in the pre-algorithmic age, the current expansion of risk assess-

¹⁹¹ Kathleen L. Mosier & Linda J. Skitka, *Human Decision Makers and Automated Decision Aids: Made for Each Other?*, in *AUTOMATION AND HUMAN PERFORMANCE: THEORY AND APPLICATIONS* 201, 203–06 (Raja Parasuraman & Mustapha Mouloua eds., 1996). Further, the institutional incentives to create mechanisms that will allow such challenges, particularly for tools that advance penal goals such as successful prosecution, are low. Andrea Roth, *Trial by Machine*, 104 *GEO. L.J.* 1245, 1250 (2016).

¹⁹² In fact, while the Federal Bureau of Prisons policy not to allow prison staff to submit letters of support on behalf of prisoners in compassionate release proceedings seems to be in place since the 1990s. See Program Statement, Kathleen Hawk Sawyer, U.S. Dep’t of Just.: Fed. Bureau of Prisons, Staff Correspondence About Inmates (Sep. 1, 1999) (on file with author) (outlining “procedures for staff recommendations and correspondence to agencies and persons about Bureau inmates”). There is some evidence that such letters were in fact submitted in the past, for example, in 2019, the same year a new risk assessment tool for compassionate release cases was introduced (known as *PATTERN*). *PATTERN Risk Assessment*, FED. BUREAU PRISONS, <https://www.bop.gov/inmates/fsa/pattern.jsp> [perma.cc/FG4N-86PW]; Press Release, U.S. Dep’t of Just., Department of Justice Announces the Release of 3,100 Inmates Under First Step Act, Publishes Risk and Needs Assessment System (July 19, 2019), <https://www.justice.gov/archives/opa/pr/departement-justice-announces-release-3100-inmates-under-first-step-act-publishes-risk-and-needs-assessment-system> (text=The%20Risk%20and%20Needs%20Assessment, to %20good%20conduct%20time%20effective [perma.cc/AVW7-NESC] (introducing the *PATTERN* risk assessment tool); see Amanda K. Rogers, *Importance of Counsel in Compassionate Release Cases*, 86 *OHIO STATE L.J.* 443, 465 (2025) (discussing a “compassionate release” case from 2019 in which an attorney was able to obtain letters from BOP staff about the attorney’s client. These letters, and additional materials, were introduced before the court deciding on the motion). Indeed, lawyers working on compassionate release cases have reported on the current challenges in introducing letters of support from prison staff. Rogers, *supra* at 473.

¹⁹³ Kaminski & Urban, *supra* note 70, at 2000–01. *Contra* Huq, *supra* note 72, at 656–61 (questioning the foundation for any prospective right to contest AI).

¹⁹⁴ See *infra* Part IV.

¹⁹⁵ See *supra* Parts II & III.

¹⁹⁶ See *supra* Part II.

¹⁹⁷ See *supra* Part III.

ment tools is the most extreme version of this evolution and has some unique idiosyncrasies.

As the scholarship on risk assessment algorithms demonstrates, there is an overlooked tension between algorithmic risk assessments and the modern subjective theory of criminal liability that deeply affects criminal law jurisprudence and redefines the values on which the modern criminal legal system has evolved. Such tension is unique because, as discussed in Part I, although scholarship on AI and the law recognizes that algorithmic tools can transform discrete legal processes across the criminal process, it overlooks their broad conceptual effects on the legal doctrine in which they function—in our case, criminal law.¹⁹⁸ This Article bridges this scholarship gap to show that the epistemic dominance of risk assessment algorithms in the criminal legal system—regardless of their design or performance—challenges fundamental principles of criminal law.

This might not be an intuitive argument. In fact, I anticipate some criticism arguing that subjective liability and algorithmic risk assessment do not contradict each other. How is that? As discussed in Part I, algorithmic risk assessment tools have been introduced to the criminal legal system at different stages, from predictive policing through prosecution, bail decisions, sentencing, and, later, in post-conviction settings such as parole and other administrative decisions including compassionate release.¹⁹⁹ To this point, one stage in the criminal process has remained “free” of algorithmic logic: the guilt phase. This is the stage where one’s culpability is determined, *de facto*, by a jury or a judge. This is also the stage where the *mens rea*, the foundational component of subjective liability, is investigated. It is thus tempting to argue that the de-subjectivizing influences of the algorithmic logic present no risk to the concepts of subjective liability and individual autonomy under the modern criminal legal system. Indeed, this is also reflected in the existing scholarship on risk assessment tools that focuses on specific stages where risk assessment tools operate.²⁰⁰

This compelling, albeit technical argument, misses the mark by ignoring the *realpolitik* of the U.S. criminal process and the interdependence between the different actors and stages of the criminal legal process. As the next Section discusses, the reality of the criminal process dramatically shifts the determination of guilt to other, earlier stages, where risk assessment algorithms are increasingly utilized.²⁰¹ Further, the interdependence between different actors and stages in the criminal process leads the non-subjective algorithmic logic to

¹⁹⁸ See *supra* Part I.

¹⁹⁹ See *supra* Part I.

²⁰⁰ See *supra* note 18 (detailing the use of risk assessment tools at various stages of the criminal legal process).

²⁰¹ See *infra* Part III.C.

pervade all stages of the criminal process, including what is left of the determination of guilt during trials.

C. The Proliferation of the Non-Subjective Algorithmic Logic

1. The Realpolitik of the Criminal Process

Understanding the vast impact of the algorithmic logic on the determination of subjective liability requires some exploration of the current reality of criminal processes in the American penal system.

Algorithmic risk assessment tools serve as gatekeepers at both ends of the criminal process (and beyond). They are being utilized at the “entry gate” of the criminal process—police investigations—and its “exit gate”—the sentencing stage.²⁰² In fact, algorithmic risk assessment tools oversee even post-conviction stages as they are utilized to guide decisions related to parole and additional administrative programs, including compassionate release.²⁰³ In between, risk assessment tools are used in decisions such as charging and bail.²⁰⁴ Technically speaking, risk assessment tools—at least as of now—are not being utilized during the guilt phase, the stage where criminal culpability is determined by jury or a judge in a criminal trial.

Criminal trials, however, despite their prominence in the public imagination,²⁰⁵ are few and far between.²⁰⁶ Therefore, the assumption that subjective liability is determined primarily at trial is erroneous. Instead, it should be acknowledged that criminal culpability is frequently determined in other, earlier stages of the criminal process where risk assessment tools dominate the landscape.²⁰⁷

Furthermore, a criminal investigation of culpability, implicating several different stages and actors, does not occur in a siloed fashion. There is a meaningful interdependence between the different stages of the process and the different actors, including police, prosecutors, and parole officers. Such interdependence is heightened by the reality of vanishing trials. Indeed, the interplay of information gathered at each stage of the process and its utilization by vari-

²⁰² See Ferguson, *supra* note 34, at 1143–44 (discussing algorithmic risk assessment tools in police investigations); Ferguson, *Predictive Prosecution*, *supra* note 40, at 705–06 (discussing algorithmic risk assessment tools in the work of prosecutors); Collins, *supra* note 18, at 60, 63–72 (documenting the rise of actuarial sentencing).

²⁰³ See Berk, *supra* note 47, at 195 (discussing the role of algorithmic risk assessment tools in determining parole release decisions).

²⁰⁴ See Mayson, *supra* note 18, at 490–96 (discussing risk assessment tools used during bail proceedings).

²⁰⁵ This is a result of heavily saturated media coverage of criminal cases and popular culture.

²⁰⁶ JOHNSON, *supra* note 140, at 14.

²⁰⁷ See Roberts, *supra* note 144, at 989 (discussing how “in a wide range of ways, in a wide range of contexts, and in the assumptions of a wide range of people, arrests appear to be fused with guilt”).

ous actors leads to what I entitle the proliferation of the algorithmic logic. This Article identifies two types of proliferation: first, *epistemic proliferation*, focuses broadly on the type of subject that enters—and leaves—the criminal process in the modern algorithmic age; second, *tangible proliferation*, focuses on practical connections across different stages of the criminal process to show how risk evaluation in each stage is tightly connected with the subsequent stage, causing the determination of culpability to shift away from subjective liability in favor of algorithmic logic. As argued earlier, the expansion of algorithmic risk assessment tools across several stages of the criminal process, paired with advanced technical capacities and scientific appeal, presents a unique threat to the institution of subjective liability that goes beyond threats previously identified in the aftermath of the new penology scholarship.

2. Epistemic Proliferation

Epistemic proliferation captures the process through which the perception of the subject in the eyes of different stakeholders in the criminal process evolves. The purportedly “objective” algorithmic logic discussed in Parts I and II dominates the categorization of subjects from the earliest stages of the criminal process, helping to construct an image of these subjects that aligns with the algorithmic logic.²⁰⁸ Such a construction in turn affects the thinking of different stakeholders along the criminal process. Furthermore, it normalizes a certain line of thinking about subjects treated in the system. Imagine the effect of predictive policing tools when utilized at the early stages of the process, the entry gate. As others have previously discussed, predictive policing is a generic term that captures a host of methodologies.²⁰⁹ Among these strategies are so-called person-based predictions that target individuals who have been deemed high-risk to commit crimes within a community. For instance, a person-based prediction model which was utilized by the Chicago Police Department until 2019 aimed to “identify the most likely offenders of violent crime” before they commit such crimes.²¹⁰ While some programs, including those operated in Chicago or Los Angeles, were shut down, the person-based methodology is still in use—though currently more limited—across the U.S.²¹¹

²⁰⁸ See *supra* Parts I & II.

²⁰⁹ See Ferguson, *supra* note 34, at 1123–44 (offering a detailed history of the evolution of predictive policing and the two most dominant methods encompassed by this terminology: place-based predictions and individual based predictions).

²¹⁰ See Ferguson, *Predictive Prosecution*, *supra* note 40, at 714 (citing Jeremy Goner, *Chicago Police Use ‘Heat List’ to Prevent Violence*, CHI. TRIB. (Aug. 23, 2013), <https://www.police1.com/chiefs-sheriffs/articles/chicago-police-use-heat-list-to-prevent-violence-aYQ4dZmhaqIvEB2g/> [<https://perma.cc/M4QL-JARN>]). Chicago’s controversial Strategic Subject List program (SSL or “Heat List”) was halted in 2019, and Los Angeles’s equivalent program (LASER) was halted in 2020.

²¹¹ U.S. DEP’T OF JUST., *supra* note 9 at 42–49; Maria Lungu, *Predictive Policing AI Is on the Rise—Making It Accountable to the Public Could Curb Its Harmful Effects*, THE CONVERSATION,

In essence, person-based prediction resembles any other algorithmic risk assessment model and thus carries with it the same algorithmic vision of the individual discussed earlier. The use of algorithmic-based prediction methods to identify individuals by the police illustrates how, from the earliest stages of the criminal process, stakeholders neglect to determine criminal culpability through an individualized investigation of subjective guilt.

When applying algorithmic logic, stakeholders do not consider the actual, measurable capacity of a certain individual to commit the crime or the likelihood that an individual will recidivate, but rather the likelihood that a group of individuals with similar characteristics will do so.²¹² For these reasons, the individual entering the criminal legal system, whose culpability is supposed to be assessed under principles of subjective liability, is stripped of their autonomy and individuality from the very beginning of the process. Instead, they are treated as a predictable object and funneled through the system based on algorithmic logic.

Although police officers are just the entry gate, their decision to arrest a subject based on other, statistically similar individuals will frame and epistemically affect both the thinking and behavior of additional stakeholders in the process, including after the guilt determination stage in sentencing, corrections, and post-conviction release. Though each of these processes has different focuses and goals, and consequently risk assessment should be deployed differently,²¹³ the algorithmic logic is one and alike: “training” actors to think about the individual in front of them in similar, non-subjective terms. As such, the expansion of uncontested risk assessment tools will most certainly cultivate the non-subject vision among stakeholders that rely on algorithmic risk assessment tools, such as prosecutors, magistrate judges in bail hearings and judges in sentencing. It will also affect the very few juries or judges sitting during the guilt determination stage. At this point in the criminal process, all stakeholders—including prosecutors and even defense lawyers—are accustomed to envisioning the individual in front of them as the algorithmically created object, a product of statistical prediction.²¹⁴ They see the individual as someone whose propensity to offend is not predicated on their own views, beliefs, and moral capabilities, but rather on the correlation between some of

<https://theconversation.com/predictive-policing-ai-is-on-the-rise-making-it-accountable-to-the-public-could-curb-its-harmful-effects-254185#:~:text=Predictive%20policing%20AI%20is%20on,could%20curb%20its%20harmful%20effects> [https://perma.cc/AD33-N7UP]; NAT'L ACADS. OF SCIS., ENG'G, MED., LAW ENFORCEMENT USE OF PREDICTIVE POLICING APPROACHES: PROCEEDINGS OF A WORKSHOP 21–23 (2025).

²¹² Collins, *supra* note 18, at 64–65.

²¹³ *Id.*

²¹⁴ My decision to discuss the effects of epistemic proliferation on defense lawyers and prosecutors alike suggests a potential shared cognitive impact of algorithmic logic also on their views of the individual within the system (even if, for some, the effect is likely implicit).

their characteristics with a group of others algorithmically categorized as “dangerous.”

Although substantive criminal law expects the jury, judges, prosecutors, and defense lawyers to consider the individual and not their predicted analogue, it is naïve to assume that, at this stage, they can still epistemologically delineate between the predictable object and the individual subject under trial. Insisting that they can simply because no algorithmic risk assessment tools are utilized at this stage²¹⁵ is technical and ignores human nature, framing effects, and the perceived power of the actuarial vision as a scientific and objective approach.²¹⁶

Furthermore, turning to criminal trials as a forum for the investigation of liability ignores the reality of vanishing trials. As previously mentioned, the vast majority of criminal cases simply do not get to trial, instead finishing at the plea bargain stage.²¹⁷ Therefore, it is a rare case indeed when the algorithmically constructed (non) subject, identified from the early stages of the process based on algorithmic logic, is challenged or modified to conform with modern ideas of mens rea and subjective liability.

I argue that the 2024 Supreme Court decision in *Diaz v. United States* offers an alarming demonstration of the potential harm stemming from the epistemic proliferation of the algorithmic logic.²¹⁸

Diaz was not a risk assessment-related case.²¹⁹ It focused on an allegedly narrow question: whether an expert’s testimony regarding the mental state of “most people” in certain situations is permitted under Federal Rule of Evidence 704(b), that states: “In a criminal case, an expert witness must not state an opinion about whether the defendant did or did not have a mental state or condition that constitutes an element of the crime charged or of a defense.”²²⁰

Diaz was arrested while entering the United States from Mexico.²²¹ Her car was searched and more than fifty-four grams of methamphetamine were found in it.²²² *Diaz* was charged with importing methamphetamine in violation of 21 U.S.C. §§ 952, 960.²²³ The mental state required to establish culpability in such offenses is one of “knowingly.”²²⁴ *Diaz* argued she did not have the requisite mental state because she did not know that drugs were hidden in her

²¹⁵ This is at least the situation as of now.

²¹⁶ See *supra* Part III.

²¹⁷ Boudin & Fish, *supra* note 140, at 1137, 1146.

²¹⁸ *Diaz v. United States*, 602 U.S. 526, 538 (2024).

²¹⁹ *Id.*

²²⁰ *Id.*; FED. R. EVID. 704(b).

²²¹ *Diaz*, at 528.

²²² *Id.* at 529.

²²³ *Id.*

²²⁴ *Id.*

car.²²⁵ The Government did not have any positive evidence to establish the required mens rea.²²⁶ Therefore, it decided to call a Homeland Security Investigations Special Agent as an expert to testify “about the common practices of Mexican drug-trafficking organizations.”²²⁷ In particular, the Agent planned to explain that drug traffickers “generally do not entrust large quantities of drugs to people who are unaware they are transporting them.”²²⁸

Diaz objected to the testimony based on Rule 704(b), arguing that the Agent’s testimony was equivalent to testimony about her mental state.²²⁹ The District Court granted Diaz’s motion in part, agreeing with her that a testimony in absolute terms, such as “drug traffickers never use unknowing couriers,”²³⁰ would be a violation of Rule 704(b).²³¹ The court, however, did not see any issue admitting a testimony described in more relative terms, such as “most couriers know they are transporting drugs.”²³² At trial, the Agent’s testimony aligned with the latter format and was therefore admitted.²³³ Diaz was found guilty by the jury and sentenced to eighty-four months imprisonment.²³⁴ The Court of Appeals agreed with the District Court and determined that the testimony did not violate Rule 704(b).²³⁵ The Supreme Court granted certiorari.²³⁶ On June 20, 2024, in a 6-3 decision, the Supreme Court affirmed the Court of Appeals decision.²³⁷ Justice Thomas delivered the majority opinion, premised on the view that the expert did not express views about the *defendant*, but rather about the knowledge of *most* drug carriers.²³⁸ As such, the testimony left the jury with discretion to decide whether Diaz was like most carriers or not, as intended by Rule 704(b).²³⁹ In her concurrence, Justice Jackson emphasized the ability of both the prosecution and the defense to bring experts to testify on probabilistic evidence, so long as the ultimate decision remains in the hands of the jury.²⁴⁰

Justice Gorsuch, joined by Justices Kagan and Sotomayor, dissented.²⁴¹ In his dissent, Justice Gorsuch recognized the novelty of the tool used by the

²²⁵ *Id.* at 528.

²²⁶ *Id.* at 530.

²²⁷ *Id.*

²²⁸ *Id.* (quoting United States’ Notice, ECF Doc. 30, at 7).

²²⁹ *Id.*

²³⁰ *Id.* (emphasis omitted).

²³¹ *Id.*

²³² *Id.*

²³³ *Id.* at 530–31.

²³⁴ *Id.* at 531.

²³⁵ *Id.*

²³⁶ *Id.*

²³⁷ *Id.* at 528, 535–38.

²³⁸ *Id.*

²³⁹ *Id.* at 535–38.

²⁴⁰ *Id.* at 538–39 (Jackson, J., concurring).

²⁴¹ *Id.* at 543 (Gorsuch, J., dissenting).

Government in this case: allowing experts to offer probabilistic evidence about one's mens rea.²⁴² That is, to establish one's mens rea not based on a defendant's individual exploration of mental state during trial, but rather based on a correlation with the mental state of others that seemingly share characteristics with the defendant.²⁴³ If this sounds familiar to you, it is not surprising—this is a pure manifestation of the algorithmic logic. And indeed, in his dissent, Justice Gorsuch closely ties the justifications for Rule 704(b) with the broader vision and values of the concept of subjective liability discussed in this Article: “Why does our law generally insist not just on a bad act but also a culpable state of mind? A significant part of it has to do with respect for the individual and his liberty in a free society.”²⁴⁴

Justice Gorsuch thus sees no significance in distinguishing between deterministic and probabilistic evidence, as both are generally *about* the defendant, exactly what Rule 704(b) hoped to avoid.²⁴⁵ Therefore, Justice Gorsuch would have found the expert testimony in *Diaz*'s case to be in violation of Rule 704(b).²⁴⁶ The dissent's concluding paragraph summarizes these concerns well:

Allowing into our proceedings speculative guesswork about a defendant's state of mind diminishes the seriousness due them. It risks the reliability of the outcomes they produce (just ask those convicted in cases where government experts opined that “all” couriers know). It undermines our historic commitment that *mens rea* is a necessary component of every serious crime by turning the inquiry into a defendant's mental state from an exacting one guided by hard facts and reasonable inferences into a competing game of “I say so.” It diminishes our respect for the presumptively free person, his free will and individuality, by encouraging the lazy assumption that he thinks like “most.” And it reduces the vital role juries are meant to play in criminal trials.²⁴⁷

I argue that *Diaz* is a product of the epistemic proliferation of the algorithmic logic, both by the government and the courts. Although *Diaz* is not a case about risk assessment tools, it borrows the algorithmic logic and utilizes it in a different domain: proving mens rea.²⁴⁸

In *Diaz*, the Government experimented with a new tactic: bringing probabilistic evidence (through an expert) as opposed to deterministic evidence in an

²⁴² *Id.* at 543, 548–49.

²⁴³ *Id.* at 549–50.

²⁴⁴ *Id.* at 544–45.

²⁴⁵ *Id.* at 549–50.

²⁴⁶ *Id.* at 549–50, 552, 555.

²⁴⁷ *Id.* at 552.

²⁴⁸ *Id.* at 535–38, 543.

attempt to prove the mental state of a defendant.²⁴⁹ There are a few potential explanations as to why such a tactic was not used before. First, the Government may honestly have believed that no path existed to establish the mental state of a suspect based on their statistical resemblance to those of others in similar situations, the leading idea behind the algorithmic logic.²⁵⁰ In this case, *Diaz* may open the floodgates for similar unsavory tactics by the Government going forward.²⁵¹ Alternatively, the Government may have believed this path was open, but was concerned that the jury would be unresponsive to this generalized logic and consequently reject the expert testimony as insufficient to establish the mental state, resulting in acquittal.²⁵² The prevalence of risk assessment tools may have eased these concerns and caused the government to test the strategy in *Diaz*. As we have seen, probabilistic evidence is now becoming more commonplace across the criminal legal system. This type of evidence also often serves, de facto, as a determinant of guilt. Therefore, it seems safe to say that the algorithmic logic has become well accepted within our society. In other words, we have become accustomed to making inferences regarding an individual based on a “scientific” assessment of their characteristics as correlated with those of others. We are willing to let go of individual assessment in favor of statistical evidence suggesting how that individual might act or think. Indeed, *Diaz* suggests that the Government was not wrong in its calculus that generalized expert testimony is enough to convince a jury of a subject’s guilt.²⁵³

The Government was also not wrong to believe that the algorithmic logic has taken over the judiciary itself.²⁵⁴ As the case suggests, courts were generally sympathetic to the Government’s experiment in offering probabilistic evidence to establish mens rea, despite the novelty of the tactic and its potential clash with foundational principles of criminal law like autonomy and free will.²⁵⁵ The algorithmic logic has prevailed; it was normalized. With it, the vision of the subject as a predicted object has also taken hold of the criminal legal system—and the process is ongoing.

²⁴⁹ *Id.* at 536, 543.

²⁵⁰ *Id.* at 535–38.

²⁵¹ *Id.*

²⁵² *Id.* Indeed, we can find evidence that, in the past, courts—in different contexts—were more willing to protect individuals from being disadvantaged by the use of statistical group-based classifications. See, e.g., *G.M.M. ex rel. Hernandez-Adams v. Kimpson*, 116 F. Supp. 3d 126, 135–40, 157–59 (E.D.N.Y. 2015) (describing various group-based classifications and ultimately holding that such classifications are prohibited). In *G.M.M.*, the defendant argued that damages should be reduced because the plaintiff, based on his ethnicity, was less likely to obtain higher education. *Id.* at 129. The court rejected this argument and instead assessed damages based on the plaintiff’s individual likelihood of obtaining higher education, regardless of his ethnicity. *Id.* at 157–59.

²⁵³ *Diaz*, at 535–38.

²⁵⁴ *Id.* at 528–31.

²⁵⁵ *Id.*

3. Tangible Proliferation

The effects of the algorithmic logic on the individual go beyond the broad idea of epistemic proliferation. They can also be seen through a direct, process-focused proliferation resulting from the interdependence between different actors and stages in the criminal process. The relationship between predictive policing and predictive prosecution illustrates this point.²⁵⁶ As previously mentioned, one of predictive policing's leading approaches is person-based prediction, that is, identifying individuals with a propensity to engage in violence before they do so based on several factors including past criminal behavior, relations to gang activity, behavior of associates, and other attributes related to the person's criminal record.²⁵⁷

In recent years, prosecutors have also embraced data-driven approaches under the assumption that they can "improve the prosecutorial function."²⁵⁸ Among these approaches is the identification of individuals (or places) that are likely to increase crime rates in the community and then subsequently use prosecutorial power to intervene before they do so. Ferguson identifies two main models of predictive prosecution: "Enforcer Model" and "Investigator Model."²⁵⁹ Under the "Enforcer Model," any re-arrest of an individual who is predicted to be involved in criminal activity" leads to an immediate police recommendation that the prosecutor charge the highest possible charges and advocate against bond release.²⁶⁰ This model illustrates how predictive policing strategies have immediate effects on prosecutorial decisions. The "Investigator Model" also uses a data-driven approach to target individuals—also known as "priority offenders"²⁶¹—who are *likely to be involved with crime*.²⁶² The data utilized in these processes—not necessarily based on algorithmic judgments—is aimed at assessing the risk posed by these individuals and includes first identifications of geographic areas of concern, and then identifying relevant individuals based on history of criminal involvement (including connection to crimes they were not charged with), past police investigations, information from community members, and more.²⁶³

²⁵⁶ John Doe's story is a vivid illustration of this process. See *supra* INTRODUCTION and accompanying text.

²⁵⁷ Ferguson, *Predictive Prosecution*, *supra* note 40, at 714–15.

²⁵⁸ *Id.* at 716; see also Ferguson, *Big Data Prosecution*, *supra* note 40, at 193–200 (detailing the rise of intelligence-driven prosecution initiatives).

²⁵⁹ Ferguson, *Predictive Prosecution*, *supra* note 40, at 716.

²⁶⁰ See *id.* at 716–17 (quoting CHI. POLICE DEP'T, CUSTOM NOTIFICATIONS IN CHICAGO, SPECIAL ORDER S10-05, at V.D. (Oct. 6, 2015) (citing to the Special Order adopted by the Chicago PD regarding the process of custom notification for individuals that were part of the Heat List)).

²⁶¹ Ferguson, *Big Data Prosecution*, *supra* note 40, at 200–01.

²⁶² Ferguson, *Predictive Prosecution*, *supra* note 40, at 716–17.

²⁶³ Ferguson, *Big Data Prosecution*, *supra* note 40, at 201–02.

In both models, mere placement on a priority list based upon algorithmic prediction meaningfully affects prosecutorial discretion in several ways. It can lead prosecutors to ask for harsher bail conditions as part of pretrial release conditions,²⁶⁴ seek maximum possible charges, offer harsher deals with minimal flexibility, seek harsher sentences, and even intervene post-conviction in parole decisions and requirements for release.²⁶⁵

Such a domino effect of predictive models on future stages in the criminal process is not only relevant for priority offenders. Rather, it is relevant for all suspects in the criminal legal system. For example, algorithmic risk assessments utilized during the bail determination stage will likely have an effect on the plea bargain terms offered by the prosecution. If an individual is predicted to be high risk, the prosecution will likely push for a harsher plea deal, again with limited negotiation bandwidth.

Thus, the interdependence of stages and stakeholders in the criminal process underscores the dramatic effect of an early-stage prediction regarding the potential risk of an individual on their treatment at future stages of the criminal process. Even before—and regardless of—the determination of guilt, decisions on how an individual should be treated from arrest through bail, plea bargaining, sentencing, and post-conviction conditions are becoming increasingly pre-determined by early risk assessments.

In principle, the guilt determination stage remains free of predictions and supposedly allows a true exploration of the individual's subjective intent. As mentioned, however, the reality is that most cases end in plea bargains and do not involve an actual guilt determination stage.²⁶⁶ This reality requires a paradigm shift in our conception about when subjective liability is determined. De facto, it is no longer the guilt determination stage where liability is established, but rather earlier stages of the criminal process. These same stages are also those where algorithmic risk assessment tools are increasingly used.

The idiosyncrasy of algorithmically-based evidence discussed above exacerbates the concerns in these scenarios. The strength of algorithmic evidence, particularly its veneer of scientific objectivity, leads overworked and understaffed stakeholders to over-rely on it, imperiling or eliminating their engagement with additional lines of investigation and further exploration of the

²⁶⁴ Ferguson, *Predictive Prosecution*, *supra* note 40, at 723. The Crime Strategies Unit in Manhattan has done this, drafting bail applications for priority targets before their arrest asking for stricter conditions. James C. McKinley, Jr., *In Unusual Collaboration, Police and Prosecutors Team Up to Reduce Crime*, N.Y. TIMES (June 4, 2014), <https://www.nytimes.com/2014/06/05/nyregion/in-unusually-close-partnership-police-dept-and-district-attorney-team-up-to-reduce-crime.html> [<https://perma.cc/B74Y-7SRU>]. See generally K. Babe Howell, *Fear Itself: The Impact of Allegations of Gang Affiliation on Pre-Trial Detention*, 23 ST. THOMAS L. REV. 620 (2011) (providing more detail on heat list inclusion based on gang association).

²⁶⁵ Ferguson, *Predictive Prosecution*, *supra* note 40, at 723; Collins, *supra* note 18, at 64–66.

²⁶⁶ Roberts, *supra* note 144, at 1014.

suspect's individual culpability. As such, risk assessment tools—imbued with the purported power of science—have become a fundamental channel where criminal liability is determined. Accordingly, liability is being determined based on the non-subjective algorithmic logic, standing in sharp contradiction to modern criminal processes of determining subjective liability.

4. The Continuous De-Individualization of Racially Marginalized Communities

Concerns about how the algorithmic logic reshapes substantive criminal law are applicable to the general population but heightened in the context of racially marginalized communities.

These communities suffer from discriminatory treatment under the criminal legal system, a product of decades of systemic explicit and implicit racism.²⁶⁷ That is not news. Such unequal treatment has clear invidious ramifications for these communities and yet, despite consistent conversations in recent years, these discriminatory patterns persist.²⁶⁸ As discussed in Part I, at some point in the debate, the introduction of algorithmic tools was promoted as a mechanism to overcome such discriminatory treatment under the assumption that such treatment is primarily a product of human bias.²⁶⁹ Therefore, the argument goes, replacing clinical decision-making with actuarial “scientific” decision-making will eliminate human bias.²⁷⁰

There is a vast literature criticizing this approach, however, arguing that, in fact, the introduction of actuarial tools has deepened racial disparities and exacerbated the existing racial imbalance in the carceral state.²⁷¹ Actuarial tools, particularly in their early years, have used race and other categories associated with race, such as education, socio-economic status, and more as a “predictor of dangerousness,”²⁷² thus preserving discriminatory patterns in enforcement.²⁷³ Furthermore, by receiving a natural-scientific tag of objectivity,²⁷⁴ algorithmic tools neutralize the utilization of different factors despite the

²⁶⁷ See Mayson, *supra* note 34, at 2251–54 (asserting that prediction's use of racially disparate historical crime data projects these disparities into the future).

²⁶⁸ See generally KHALIL GIBRAN MUHAMMAD, *THE CONDEMNATION OF BLACKNESS: RACE, CRIME, AND THE MAKING OF MODERN URBAN AMERICA* (Harvard Univ. Press ed. 2011) (discussing the development of the social construct of “black criminality” in modern urban America).

²⁶⁹ See *supra* Part I.

²⁷⁰ Flores, Bechtel & Lowenkamp, *supra* note 59, at 38; Skeem & Lowenkamp, *Risk, Race, and Recidivism*, *supra* note 59, at 680; see *supra* Part I.

²⁷¹ See HARCOURT, *supra* note 174, at 2–3 (surveying the origins and proliferation of actuarial methods in law enforcement, and pushing back against what he calls “the new actuarial paradigm”); see also Selbst, *supra* note 39, at 120–21 (discussing algorithm use with discriminatory outcomes).

²⁷² Harcourt, *supra* note 62, at 238.

²⁷³ *Id.*; Sidhu, *supra* note 45, at 695–99.

²⁷⁴ Jessica M. Eaglin, *Technologically Distorted Conceptions of Punishment*, 97 WASH. U. L. REV. 483, 487 (2019); see Kaminski & Urban, *supra* note 70, at 1969 (“But AI is not always accurate

deep “historical association [between these factors and] political and social oppression.”²⁷⁵

To eliminate issues of racial bias, early critiques of algorithmic tools first raised questions of design, such as eliminating race from the analyzed data. Scholars argue, however, that modern algorithms continue to produce racial disparities even after the elimination of race as a category in many risk assessment tools.²⁷⁶ This is because the factors utilized by risk assessment tools traditionally correlated with an increased propensity to engage in criminal activity “are the very conditions that Black and other politically oppressed communities are disproportionately forced to live in.”²⁷⁷ These factors include, for example, poverty, work instability, housing, and more.²⁷⁸ Moreover, even if algorithms are designed to not include certain discriminatory factors, data mining processes, the bread and butter of algorithmic tools, likely exacerbate their disparate impact on marginalized communities by relying on design choices and data that still reproduce the prejudices of past decisions or even general societal bias.²⁷⁹ Concerns about the discriminatory effects of algorithmic outputs are therefore well discussed and documented.

The drift away from subjective liability creates *additional* pressure on marginalized communities already oppressed by the criminal legal system. Recognizing the tension between algorithmic logic and the notion of the individual in the criminal process exposes yet another, deeper, epistemological layer of how generalization has long impacted marginalized communities. Beyond particular outputs produced by algorithmic tools, the unequal treatment these marginalized communities receive from the criminal legal system is in part a product of years where members of these communities, particularly

and does not eliminate human bias. In some cases, it instead obfuscates bias with layers of ostensibly objective mathematical authority.”).

²⁷⁵ SLOBOGIN, *supra* note 32, at 87; *see also* VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* 168 (2018) (contrasting algorithmic risk assessments as “the discretion of the few” and human discretion as “the discretion of the many”).

²⁷⁶ Eaglin, *supra* note 18, at 94–99; Ngozi Okidegbe, *Beyond More Accurate Algorithms: Takeaways from McCleskey Revisited*, 121 MICH. L. REV. 1109, 1115 (2023). This is a deep and ongoing debate that is beyond the scope of this Article. *See, e.g.*, Skeem & Lowenkamp, *Risk, Race, and Recidivism*, *supra* note 59, at 680 (contending that the use of risk assessment instruments in corrections and sentencing could help “unwind mass incarceration”); Black et al., *supra* note 63, at 79–85 (discussing less discriminatory alternatives).

²⁷⁷ Okidegbe, *supra* note 276, at 1116; *see also* Allegra McLeod, *An Abolitionist Critique of Violence*, 89 U. CHI. L. REV. 525, 541–45 (2022) (discussing how “entrenched inequality is . . . sustained by contemporary economic laws and policies” in the context of racialized inequality and economic development in Chicago); *see also* Harcourt, *Risk as a Proxy for Race*, *supra* note 62, at 238 (arguing that today, risk has become a proxy for race).

²⁷⁸ Okidegbe, *supra* note 276, at 1116.

²⁷⁹ Selbst, *supra* note 39, at 121–23; Barocas & Selbst, *supra* note 61, at 674.

Black people, were not perceived as individuals, but rather as a “collective.”²⁸⁰ This collective was socially stereotyped, racially discriminated against, and constructed to be correlated with crime.²⁸¹ Such a socio-pathological view reflects a social failure to perceive Black people as individuals, subjects of empathy, and “vulnerable human beings who sustain pain and love and hatreds and fears and joy and sorrows and degradations and triumphs.”²⁸² Indeed, these communities have long been de-individualized and stripped of their autonomous self and dignity upon entry through the gates of the criminal legal system. This treatment began well before the widespread use of algorithmic tools.

The introduction of algorithmic tools exacerbated such conditions for these communities, epistemologically and practically. Treating individuals based on racial or other classifications already suggests that the system does not really know how these individuals would behave if treated otherwise.²⁸³ The algorithmic logic adds another layer of separation between members of these communities and the criminal legal system’s purported pursuit of subjective liability, this time with the support of scientific objectivism. In doing so, algorithmic utilization normalizes social resistance to seeing the non-white “others” as complete and unique individuals with different moral views, hopes, fears, and beliefs. Individuals that make good and bad choices, for example, and deserve to be treated with compassion and dignity. Similarly to Foucault’s arguments on the normalizing powers of disciplinary practices,²⁸⁴ by treating certain members of a marginalized group as if they de facto have the characteristics of their group—as the algorithmic logic does—these members might be normalized into that profile,²⁸⁵ with no ability to refute it.

Therefore, the expansion of the algorithmic logic into the criminal legal system and its inherent denial of the individual can be perceived as another mechanism that reinforces the de-humanization of racially marginalized communities. Consequently, it aggravates pre-existing infringements on autonomy in

²⁸⁰ See generally MUHAMMAD, *supra* note 268 (discussing the development of the social construct of “black criminality” in modern urban America and how Black people were collectively associated with crime); Dorothy E. Roberts, *Race, Vagueness, and the Social Meaning of Order-Maintenance Policing*, 89 J. CRIM. L. & CRIMINOLOGY 775, 799–800 (1999) (concluding that “[e]ven my relatively privileged son had become acculturated to one of the salient social norms of contemporary America: Black children, as well as adults, are presumed to be lawless and that status is enforced by the police”).

²⁸¹ MUHAMMAD, *supra* note 268.

²⁸² Jewell Handy Gresham, *White Patriarchal Supremacy: The Politics of Family in America*, NATION, July 1989, at 116, 120.

²⁸³ Hildebrandt, *supra* note 168, at 134.

²⁸⁴ See generally MICHEL FOUCAULT, DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON (Alan Sheridan trans., Vintage Books 2d ed. 1979) (1975) (chronicling the rise of penal institutions and the failures of modern prisons).

²⁸⁵ Hildebrandt, *supra* note 168, at 133; see also LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 154 (1999) (“[P]rofiles will begin to normalize the population from which the norm is drawn The system . . . fits you into a pattern; the pattern is then fed back to you . . .”).

these communities by further abandoning the processes of determining subjective liability. Relatedly, the use of algorithmic tools contributes to a continuous devaluation of the rights of marginalized communities before the system, creating a deeper drift between them and the values the modern criminal legal system aspires to advance, including liberty, fairness, and self-determination.²⁸⁶

IV. PRESERVING SUBJECTIVE LIABILITY IN AN ALGORITHMICALLY-DOMINATED CRIMINAL LEGAL SYSTEM

So far, this Article has raised serious questions about the utilization of algorithmic tools in the criminal process. Not only are risk assessment tools in tension with criminal law concepts of subjective liability and individuality, but their proliferation has narrowed opportunities for re-emergence of the subject by creating a criminal legal system dominated by algorithmic assessment and its logic. As we have discussed, the use of these tools has not only reshaped how guilt is evaluated, debated, and understood—relying on new forms of evidence and criteria—but has also altered the temporal framework of institutional decision-making, shifting it toward an *ex ante* adjudication of criminality. Therefore, as discussed, and due to their unique characteristics, algorithmic risk assessment tools likely represent the most radical shift from modern ideas of dignity and autonomy, one that goes beyond existing scholarly conversations around the new penology.

Given that the algorithmic logic has become embedded in the criminal process, improvement of algorithmic accuracy, sophistication, or any other design-related elements, will likely not resolve the deep tensions and harms discussed in this Article. One may argue that it is impossible to reconcile these tensions while preserving the reign of algorithms. The ultimate goal of this Article, however, is not to call for the abolition of algorithmic tools in the criminal legal system. First, because abolition is not a realistic goal. Algorithms are here to stay and will likely expand with time to influence additional parts of the criminal process. Second, because abolition is not necessarily the desirable outcome given the complexity of assessing the advantages (and potential advantages) of algorithmic tools *vis-à-vis* the current human-made decision-making processes in the criminal context. The impressive amount of scholarship aiming to decipher exactly this point illustrates the intricacies of these questions, that this Article does not intend to resolve.

²⁸⁶ Finkelstein, *supra* note 93, at 895. From a broader socio-technological perspective, the expansion of algorithmic risk assessment can thus be understood as another tool to reflect certain societal values of the dominant class. See, e.g., COHEN, *supra* note 166, at 3–4 (discussing how, rather than representing an unstoppable “technology of freedom,” the internet instead reflects the values of its developers). This discussion that gets to the heart of the political economy of algorithmic risk assessment utilization as a form of social control deserves more attention in a separate setting.

Instead, this Article recognizes that with the expansion of algorithmic utilization, a growing body of academic and policy work criticizing such utilization has developed. This expansive body of work illuminates some social undercurrents that offer an opportunity to rethink settings where subjects can be re-introduced as individuals into the criminal legal system. These settings are critical opportunities to counteract the feedback loop of generalization that has thrived in the criminal legal system in the face of unchecked automation. Ironically, it might be the over-expansion of algorithms that allows us to advance deep structural changes in our penal system by introducing additional avenues for contestation and attacking, head on, inequities that have long plagued it.

This Part begins scratching the surface by offering a preliminary legal framework and derived mechanisms that can assist in envisioning these opportunities.²⁸⁷ Section A of this Part offers a synthesis of rights giving rise to this framework.²⁸⁸ Section B of this Part offers an illustration of how the re-individualization of the criminal legal system can occur through contestation of the algorithmic logic, while recognizing potential limitations and challenges to the application of this framework.²⁸⁹

A. Legal Framework—Synthesis of Rights

The first step in creating a legal framework to re-individualize the criminal legal system requires synthesizing three rights: subjects' rights to express themselves, the right to due process, and the right to contest algorithmic tools.²⁹⁰ This synthesis opens the legal door to potential paths for individuals to engage with algorithmic decisions early in the criminal process. These paths, in turn, can guarantee substantive due process rights while offering individuals the opportunity to reclaim the autonomy so negatively affected by the proliferation of the algorithmic logic.

The idea of creating procedural protections to preserve the opportunity to defend oneself against AI-dominated decision-making processes is not new. It has been discussed in different settings, mostly by drawing parallels to the

²⁸⁷ See *infra* Part IV.

²⁸⁸ See *infra* Part IV.A.

²⁸⁹ See *infra* Part IV.B.

²⁹⁰ Kaminski & Urban, *supra* note 70, at 1989–90. Although Kaminski and Urban offer compelling arguments for its introduction into the American legal system, the right to contest AI is not yet fully recognized in the United States. *Id.* The right to contest AI, however, is recognized in the European Union under the General Data Protection Regulation. Council Regulation 2016/679, 2016 O.J. (L 119) 46; see Alexandra Natapoff, *Speechless: The Silencing of Criminal Defendants*, 80 N.Y.U. L. REV. 1449, 1479–80 (2005) (detailing how suspects' right to speak includes the right to make a statement on their behalf before sentencing, the right to testify, the right to be self-represented, and more); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 18–22 (2014) (discussing due process considerations and emphasizing data transparency).

right to due process.²⁹¹ The right to due process in the United States is protected under the Fifth and Fourteenth Amendments,²⁹² and in essence requires “notice and an opportunity to be heard.”²⁹³ This Article focuses on the opportunity to be heard, or the right to speak, due to its strong ties to the ideas of personhood and individual autonomy. Scholars have identified three main justifications for the right to due process: accuracy, supporting rule of law values, and liberal theory that is strongly tied to ideas of the individual in Western societies.²⁹⁴

Accuracy reflects the legal system’s aspiration to avoid wrong decisions.²⁹⁵ It is recognized, however, more “as a goal or value to be balanced against other” values than as providing support for the individual right to due process.²⁹⁶ Rule of law encompasses, *inter alia*, the idea that decisions should not be “unfair, inconsistent, arbitrary, unpredictable, or irrational.”²⁹⁷ Similar to accuracy, however, it reflects an institutional argument more than a justification for an individual right to due process, which is the focus of this Article’s premise. The third justification for the right to due process—its grounding in liberal theory—seems to best align with the aspiration to re-introduce the individual into a system dominated by non-subjective algorithmic logic. Three main categories of liberal theory are recognized in the literature: Bentham’s utilitarian theory, Locke’s natural rights theory, and Kant’s personhood theory.²⁹⁸ Of the three, Kant’s personhood/dignitary theory is most closely aligned with the idea of subjective liability in criminal law and the values it aspires to advance. Kant’s justification submits that due process rights are required to respect the individual personhood,²⁹⁹ the same personhood taken from the individual assessed by algorithmic tools. This is particularly true when funda-

²⁹¹ See Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1249–50 (2008) (introducing the technological due process framework as a way to address challenges to due process rights in the automated state); Keats & Pasquale, *supra* note 290, at 8 (offering a “technological due process” to “introduce human values and oversight” into scoring systems); Kaminski & Urban, *supra* note 70, at 1989–90.

²⁹² U.S. CONST., amends. V & XIV.

²⁹³ Kaminski & Urban, *supra* note 70, at 1990. See generally GARRETT, *supra* note 70, at 1 (defending due process as a right “to fair treatment from the government” while acknowledging the “human tendency to want to deny it to people”).

²⁹⁴ Kaminski & Urban *supra* note 70, at 1989–90.

²⁹⁵ *Greenholtz v. Inmates of Neb. Penal & Corr. Complex*, 442 U.S. 1, 13 (1979); Huq, *supra* note 72, at 653.

²⁹⁶ Kaminski & Urban, *supra* note 70, at 1990.

²⁹⁷ *Id.* at 1991.

²⁹⁸ *Id.* at 1992 (citing Jerry L. Mashaw, *Administrative Due Process: The Quest for a Dignitary Theory*, 61 B.U. L. REV. 885, 907 (1981)).

²⁹⁹ *Id.* at 1993–94. It should be noted that many scholars in the algorithmic and privacy space reject Kantian dignitary views in the U.S. context, a view that Kaminski and Urban criticize, arguing that “there is a long dignitary tradition in the United States” in the context of process theory. *Id.* at 1993.

mental rights are at stake.³⁰⁰ The right to liberty, that is most at risk in the criminal process, provides strong support for due process as a basis for contesting the use of risk assessment tools. Contesting algorithmic decisions during the criminal process thus offers an opportunity to reassert the right to due process in the face of declining individuality and autonomy.³⁰¹

Kaminski and Urban have offered compelling arguments to build on the theoretical underpinnings of the right to due process to establish an individual right to contest AI.³⁰² Particularly relevant is their dignitary-based argument.³⁰³ As discussed in Part III, statistical categorization of individuals strips people of their individuality.³⁰⁴ As such, a right to contest AI decisions, as an extension of due process, contributes to the recovery of the subject as a whole individual before the criminal legal system.³⁰⁵ Cen and Raghavan's discussion of the right to be an exception to a data-driven rule, that they find closely related to the concept of dignity, likewise emphasizes the importance of allowing a right to

³⁰⁰ *Id.* at 1994.

³⁰¹ *Id.*

³⁰² *Id.* Kaminski and Urban discuss and compare this right to a recognized right in the European Union. *Id.* at 1980.

³⁰³ *Id.* at 2000–01. This is an argument Kaminski and Urban admit is contested in the U.S. context.

³⁰⁴ *Id.*; see *supra* Part III.

³⁰⁵ Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1541 (2019). In this context, several decisions rendered by courts in the United States should be mentioned. See *State v. Headley*, 926 N.W.2d 545, 547–48 (Iowa 2019) (upholding a defendant's sentence by rejecting their argument that the use of risk assessment tools violated his due process rights); *State v. Gordon*, 921 N.W.2d 19, 22–24 (Iowa 2018) (rejecting a defendant's due process challenge to the use of risk assessment tools at sentencing); *State v. Loomis*, 881 N.W.2d 749, 772 (Wis. 2016) (holding that a lower court's use of risk assessment tools did not violate the defendant's due process rights, but explaining that risk assessment tools must be used with caution). In each of the above-cited cases, defense counsel presented similar due process arguments to contest the use of risk assessment at the sentencing phase. As is reflected in their respective holdings, however, although courts refused to ban the mere use of algorithmic risk assessment in sentencing based on the due process argument, they were willing to recognize the potential due process concerns associated with the use of risk assessment tools. See, e.g., *Loomis*, 881 N.W.2d at 757 (“Although we ultimately conclude that a COMPAS risk assessment can be used at sentencing, we do so by circumscribing its use. Importantly, we address how it can be used and what limitations and cautions a circuit court must observe in order to avoid potential due process violations.”). The approach in *Loomis* was closely related to the de-individualization of the algorithmic logic this Article discusses:

However the due process implications compel us to caution circuit courts that because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders—not a particular high-risk individual. Accordingly, a circuit court is expected to consider this caution as it weighs all of the factors that are relevant to sentencing an individual defendant.

Id. at 765. The framework suggested in this Article, including establishing a contestation framework based on due process, is thus aligned with the court's approach.

contest algorithmic decision-making.³⁰⁶ By demonstrating how exceptions to data-driven decision-making are practically unavoidable, these authors reveal the ties between individualization, uncertainty, and harm while elucidating why it is justified to create platforms of contestation as a form of dignity restoration. These arguments are particularly relevant in the context of individuals treated by the criminal legal system, where the risk of harm stemming from this decision is exceptionally high.

Indeed, and at least *de jure*, individuals who go through the criminal process are supposed to enjoy several rights that preserve the concept of “self” by allowing them to express themselves. Based on these protections, then, the right to contest AI in the criminal context should include the opportunity to offer individualized evidence to challenge certain conclusions made by the criminal legal system’s representatives with help from risk assessment algorithms.³⁰⁷ *De facto*, however, real opportunities to speak in the criminal process are few and far between, at least in ways that fully advance an individual’s dignity and autonomy.³⁰⁸ Similarly to Kaminski and Urban, Natapoff emphasizes the dignitary deficit caused by this inability to speak. There is a deeply sad irony in this situation, where the fewest opportunities for individuals to speak are offered in a system where they face the most severe harm. Here, again, the phenomenon of vanishing trials has strong negative effects on the

³⁰⁶ Cen & Raghavan, *supra* note 184, at 5–6. Cen and Raghavan’s unique proposition justifies the right from the perspective of the individuals affected by it, while emphasizing the position of the decision-maker offering it, requiring “that the decision-maker not necessarily presume that a data-driven rule is suitable for a decision-subject, particularly when the risk of harm is high.” *Id.* at 12. By doing so, their proposition shifts the burden to the decision-maker to offer an *ex ante* justification that a data-driven decision has considered the right’s main components. *Id.* at 16. Thus, the decision-maker “must presume that the data-driven rule is not fit, only applying it when sufficiently convinced that the rule’s levels of individualization and certainty justify the level of harm it may inflict.” *Id.* They also recognize, however, the importance of *ex post* contestation by the individuals affected by the decision. *Id.* at 16–17.

³⁰⁷ See Natapoff, *supra* note 290, at 1449–51, 1476 (describing the lack of speaking opportunities for defendants in the modern criminal justice system); Akhil Reed Amar & Renée B. Lettow, *Fifth Amendment First Principles: The Self-Incrimination Clause*, 93 MICH. L. REV. 857, 925 (1995) (recognizing the potential of defendants’ utterances to reveal their inner self in the context of criminal interrogation, but arguing against their compelled introduction by the government). For similar ideas about the need to allow individuals to share their own narratives in a different context, see generally Frank Pasquale & Mathieu Kiriakos, *Contesting the Inevitability of Scoring: The Value(s) of Narrative in Consumer Credit Allocation* (discussing the right to contest credit scoring in the context of consumer credit allocation), in *ALGORITHMIC TRANSFORMATIONS OF POWER: BETWEEN TRUST, CONFLICT, AND UNCERTAINTY* (C. Burchard & I. Spiecker eds., forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4947680 [<https://perma.cc/F8ZU-W4KE>].

³⁰⁸ See Natapoff, *supra* note 290, at 1450, 1475–76 (“Fifth and Sixth Amendment doctrines contain significant theoretical commitments to defendant expressive choice and dignity, but in operation they inevitably succumb to the instrumental pressure in favor of silence.”); see also M. Eve Hanan, *Talking Back in Court*, 96 WASH. L. REV. 493, 521–23 (2021) (arguing that although defendants are given formal opportunities to speak in court, their speech usually conforms to the court’s expectations). This type of speech does not fulfill the autonomy and dignitary values I discuss in this Article.

ability to speak, as it eliminates a meaningful process where suspects are allegedly entitled to express themselves. The introduction of algorithmic risk assessment tools exacerbates such dignitary concerns. Introducing opportunities to contest algorithmic decision-making at multiple stages of the criminal process can thus potentially tighten the gap between the de jure right to speak and its de facto implementation.

Thinking conjointly about the suspect's right to due process, right to contest AI, and right to speak in criminal proceedings, as well as the underlying values of these rights, introduces a promising legal path to tackle the desubjection concerns raised in this Article. This path emphasizes the rights at stake, the (at least de jure) commitments to individual liberty, and the unique harms stemming from the intervention of algorithmic outputs. All of these rights share a common interest in protecting the subject and they emphasize the clash between liberal democratic values of personhood and the inability to speak or contest. As such, this Article submits that creating an opportunity to respond, by itself, carries a personhood-regaining value.

B. Implementing the Framework: Re-Individualization Through Contestation

With the understanding that risk assessment tools will continue to dominate the criminal legal system, the legal framework discussed above prioritizes engagement and contestation across multiple stages of the criminal process as promising avenues to counteract the decline of autonomy that has accompanied the rise of algorithmic decision-making.³⁰⁹ The reality of the criminal legal system, particularly the phenomenon of vanishing trials, underscores the importance of providing opportunities to engage and contest at even early stages of the criminal process, given that preliminary algorithmic outputs increasingly exert a substantial effect on final outcomes.³¹⁰

Implementing this framework could yield a host of mechanisms and procedures. These include, for example, the ability to respond to an individual's inclusion in a list indicating a propensity to offend by the police, or to determinations of risk established by prosecutors when deciding on charges, bail, or plea bargains. Similar opportunities to contest risk assessment decisions should be given during the sentencing and post-conviction stages. How each of these mechanisms will look is a complicated and contested issue that should be left

³⁰⁹ See *supra* Part IV.A.

³¹⁰ See *supra* notes 139–140 (discussing how the U.S. criminal legal system is dominated by plea agreements, not trials); Kaminski & Urban, *supra* note 70, at 1957 (advocating for the right to contest AI decisions in the criminal legal system).

to future studies.³¹¹ This Section, however, will kickstart the conversation by exploring several key issues that should be considered in any attempt to implement the proposed framework.³¹²

Stages in the Criminal Process. In which stages of the criminal process should these procedures be adopted? As discussed above, the de facto criminal legal process inflates the importance of determinations, including automated determinations, made in the early stages of a case. Thus, it will be important to construct a framework that properly addresses the use of AI across multiple stages of the criminal process. Different stages will implicate different opportunities for contestation and pose unique challenges for both the individual and the government. Contesting a decision to be included in a list of individuals with propensity to commit crimes, for example, is likely to occur before any individualized suspicion has materialized.³¹³ As such, challenging AI at this stage would require both a public platform to share the information—raising privacy concerns—and mechanisms for interested parties to contest it—posing constitutional concerns under the Fourth Amendment.³¹⁴ By contrast, a challenge to the use of risk assessment tools during a sentencing hearing, where the individual has already been charged and found guilty, is more likely to elicit a due process argument from the defense.³¹⁵

³¹¹ Kaminski & Urban, *supra* note 70, at 2003–04. The main challenges revolve around what Kaminski and Urban define as questions of speed and scale, the same questions AI was initially designed to resolve. *Id.* at 2004.

³¹² In their article, Kaminski and Urban offer a general, and useful, framework—entitled contestation archetypes—for the establishment of a right for contestation. *Id.* at 2004–05. While practically and normatively related, this Section does not intend to contest or replace that framework. Instead, it offers a set of considerations, uniquely focused on the criminal process, that should be considered in adopting contestation rights in the criminal process. As I suggest, different stages of the process might yield different contestation rights under the Kaminski and Urban archetypes (an option they recognize as well). *Id.* at 2009.

³¹³ See generally Ferguson, *supra* note 34, at 1137–43 (discussing the process through which a subject on a “heat list” is notified of their status). Ferguson notes that individuals included on “heat lists” are informed of their status and penalties for subsequent criminal activity. *Id.* at 1139–40, 1141. An individual included on such a list is subjected to increased police surveillance and intervention and may face heightened punishment in the event of criminal activity. *Id.* at 1141–43. Ferguson’s observation, then, suggests that the first opportunity to contest the use of AI in the criminal legal process may occur *prior* to the commission of any particular crime.

³¹⁴ Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 329–32 (2015); Selbst, *supra* note 39, at 116–17; see also Ángel Díaz, *Data-Driven Policing’s Threat to Our Constitutional Rights*, BROOKINGS INST. (Sep. 13, 2021), <https://www.brookings.edu/articles/data-driven-policings-threat-to-our-constitutional-rights/> [<https://perma.cc/D9AZ-NEHJ>] (discussing the propensity for predictive policing to conflict with constitutional rights and recognizing political resistance as a viable, non-legal means of contesting such practices). Díaz argues that predictive policing programs discount constitutional rights to privacy, reasonable suspicion, and probable cause to justify “ongoing monitoring” of high-risk individuals. Díaz, *supra*.

³¹⁵ See *supra* note 305 (finding that courts have refused to ban the use of algorithmic risk assessments in sentencing on due process grounds while acknowledging the potential for due process violations associated with using these tools).

The Parties Participating. Questions related to parties involved in these processes are similarly challenging. The subject of the decision should clearly be a part of the process, but what about on the government side? Representatives from the decision-making entity (police, prosecution, et cetera) should naturally be included, but what about representatives of the companies designing the risk assessment tools? In response to the rise of algorithmic decision-making in the criminal process, scholars have suggested that the companies behind AI models should be held more accountable and assume a more substantial role in their implementation.³¹⁶ Undoubtedly, including algorithm creators in the contestation process would enhance the parties' respective understandings of an algorithm and the factors that led to a contested decision. Such collaboration would facilitate a more effective contestation process by addressing asymmetries in sophistication and information frequently present in criminal trials.³¹⁷ Nevertheless, collaboration and disclosure face difficult questions. For one, creators have been historically reluctant to divulge information about proprietary models and have had success claiming trade secret privilege to avoid doing so.³¹⁸ Furthermore, private parties are not privy to criminal proceedings, and bringing them into the process may raise procedural and conceptual questions.³¹⁹ Given the effectiveness that collaboration and disclosure would usher in, this is a substantively desirable route. These practical issues, however, must be addressed and resolved to foster the necessary breadth of change.

³¹⁶ See Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1835–36, 1844–46 (2019) (advocating for holding creators accountable for the reliability of their algorithm by providing subjects with information about its design and application, and offering solutions to potential push-back from creators). Deeks recommends increased transparency in AI decision-making by providing defendants with “model-centric” information about the process by which a model was created and operates. *Id.* at 1835, 1841, 1846; see also Kroll et al., *supra* note 63, at 703–05 (discussing “special masters” and partial disclosures as “politically and practically” feasible mechanisms for holding creators responsible through increased oversight). Kroll et al. break with Deeks on the issue of transparency by arguing that effective oversight may require only partial disclosures through various computer science techniques. Kroll et al., *supra* note 63, at 703–05. Despite their different approaches, all of these authors advocate for increased accountability for algorithm creators through the introduction of enhanced oversight mechanisms to ensure the fairness and reliability of AI models.

³¹⁷ See Deeks, *supra* note 316, at 1835–36 (discussing the disclosures implicated by “model-centric” and “subject-centric” xAI). Deeks notes that model-centric xAI would provide subjects with information about the design choices of an AI model and how it operates, generally. *Id.* at 1835. Also relevant in the context of contestation is subject-centric xAI, that would provide parties to a challenge with information about the factors that proved most influential to an algorithmic decision. *Id.* at 1836. Both types of disclosures would foster a more effective contestation process by reducing information asymmetries inherent in many criminal cases. *Id.* at 1846.

³¹⁸ *Id.*

³¹⁹ *Id.* at 1840–41. Indeed, recent scholarship discussing frameworks for contesting AI decisions has largely focused on the application of such frameworks in privatized settings. See Black et al., *supra* note 63, at 58 (excluding criminal proceedings from their analysis); Kaminski & Urban, *supra* note 70, at 2003–05 (alluding to the relevance of their framework to criminal settings, but offering a more concrete discussion and analysis in non-criminal settings).

Access to Data. The ability of individuals to access the data on which they were assessed raises similar questions: How should such access be provided? Should challengers be given an opportunity to receive full access to the information used against them in criminal proceedings? Furthermore, how extensive should disclosures be? A meaningful contestation seems to call for deep engagement not only with the data collected, but also with the models applied in the decision-making process and the design choices made in creating the algorithm.³²⁰ Indeed, scholars have suggested that access to such data not only forms a pillar of responsible AI use, but also provides a practical solution to the otherwise asymmetrical contestation landscape.³²¹ Crucially, transparency at contestation, particularly in regard to what Deeks would term “subject-centric” disclosures, serves as a viable means of re-introducing the individual into the criminal process at points where they are otherwise reduced to a

³²⁰ The model adopted by the GDPR seems to support this approach, requiring exposure to “meaningful information about the logic involved” and the right to access the data. Council Regulation 2016/679, 2016 O.J. (L 119) 41; *see also* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 204, 209–13 (2019) (discussing the importance of offering meaningful data related to the decision in order to allow a meaningful contestation process); GARRETT, *supra* note 70, at 145–49 (arguing for a right to more transparent “glass-box” AI); Brandon L. Garrett, *Artificial Intelligence and Procedural Due Process*, 27 U. PA. J. CONST. L. 933, 940 (2025) (advocating for interpretable “glass box” AI systems to allow compatibility with the right to due process). *See* OREN BAR-GILL & CASS R. SUNSTEIN, ALGORITHMIC HARM: PROTECTING PEOPLE IN THE AGE OF ARTIFICIAL INTELLIGENCE 96–100 (2025), for more on the importance of transparency in addressing algorithmic harm in other contexts.

³²¹ *See* Kars Alfrink, Ianus Keller, Gerd Kortuem & Neelke Doorn, *Contestable AI by Design: Towards a Framework*, 33 MINDS & MACHS.: J. FOR A.I., PHIL., & COGNITION 613, 621 (2022) (advocating for providing decision subjects with general information about a model and an explanation of how it was applied to their case). The authors argue that models should be designed to allow contesting parties to access the “organizational, social and legal context of [a] decision.” *Id.* The enhanced access discussed by Alfrink et al., then, would not signal an important step toward transparency and autonomy, but manifest as a practical tool for parties to verify the procedural regularity and normative fairness of an AI model. *Id.*; *see also* Deeks, *supra* note 316, at 1842–46 (discussing the use of xAI in the criminal space as a means of facilitating fairness in contestation). Deeks notes that the enhanced transparency offered by xAI may be particularly pertinent in the criminal space, where decision subjects are generally unsophisticated actors with limited knowledge of algorithmic decision-making. Deeks, *supra* note 316, at 1846; *see, e.g.*, Angela Jin & Niloufar Salehi, (*Beyond*) *Reasonable Doubt: Challenges That Public Defenders Face in Scrutinizing AI in Court*, ARXIV, Mar. 13, 2024, at 1, 20–21, <https://arxiv.org/abs/2403.13004> [<https://perma.cc/H5NV-UDNC>] (proposing expanded disclosure procedures as a solution to the challenges faced by public defenders challenging AI decisions in court). The authors canvas an array of difficulties faced by public defenders challenging AI decisions including concealment by prosecutors and developers, lack of access to existing information due to monetary constraints, and early-stage plea deals derailing contestation efforts. Jin & Salehi, *supra*, at 6–9. They advocate for the adoption of independent verification procedures prior to the implementation of an AI model and bilateral deliberation of reliable performance outputs to ease the burden on public defenders. *Id.* at 12–14. Both of these proposed avenues for contestation would necessarily entail disclosure of information about model design choices, thereby holding creators more accountable and restoring autonomy to decision subjects.

generalization.³²² Alternatively, could the rationale employed by the Supreme Court in its landmark 1963 holding in *Brady v. Maryland* be applied in this setting?³²³ As discussed, risk assessment tools are not used for determining guilt, which was the center of Brady's due process violation. On the other hand, decisions made by risk assessment tools proliferate into the guilt phase and in some instances seem to replace it altogether, strengthening the ties between the *Brady* rationale and the access to data utilized by the algorithm.³²⁴

Contesting the Decision. Another important question is how defendants might mount a cognizable defense to AI recommendations. Given that broad constitutional arguments aimed at exclusion have been largely unsuccessful, a narrower approach may prove more fruitful.³²⁵ Indeed, subjects of AI decisions may be better served accepting its use at face-value and focusing instead on mounting *substantive* pushback to its application in their particular case. In my view, questions of data access aside, this contestation process should bring to light any relevant individualizing—and contextualizing—data about an individual's thoughts, beliefs, and personal development. Self-testimonies, letters of support, and additional documentation to explain past transgressions may be relevant to building a strong challenge. Exploring this information presents a valuable opportunity to distinguish subject from predicted object and facilitate the reintroduction of autonomy into the criminal process. The National Association of Criminal Defense Lawyers (NACDL) has endorsed this approach, recommending that attorneys contesting risk assessment scores in the pretrial stage prepare individualized explanations to counteract generalized predictions.³²⁶ For example, the NACDL suggests a defense attorney may be well-served to present documents, such as a doctor's note, to explain a defendant's court date absence.³²⁷ Other documents, however, can be similarly effective. A vocational school certificate or a letter of recommendation from a community

³²² See Megan T. Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 WASH. U. L. REV. 681, 682 (2018) (discussing transparency in the use of risk assessment algorithms as a means of preserving defendant individuality); Deeks, *supra* note 316, at 1846.

³²³ *Brady v. Maryland*, 373 U.S. 83, 87–91 (1963); see also Ferguson, *Predictive Prosecution*, *supra* note 40, at 183–84, 243–51 (discussing the *Brady* problem in the context of intelligence-driven prosecution in the age of big data).

³²⁴ See, e.g., Deborah Won, *The Missing Algorithm: Safeguarding Brady Against the Rise of Trade Secrecy in Policing*, 120 MICH. L. REV. 157, 157–60 (2021) (describing the *Brady* problem).

³²⁵ See *supra* note 305 and accompanying text (detailing how many courts have upheld the legality of risk assessment algorithms in the criminal justice context).

³²⁶ See Brandon Buskey & Andrea Woods, *Making Sense of Pretrial Risk Assessments*, NAT'L ASS'N CRIM. DEF. LAWS. (June 2018), <https://www.nacdl.org/Article/June2018-MakingSenseofPretrialRiskAsses> [<https://perma.cc/2Y99-Y2BE>] (discussing strategies to contest risk assessment algorithm's non-individualized pretrial detention recommendations).

³²⁷ *Id.*

leader may also persuasively rebut an algorithmic assumption that a defendant is a high risk to recidivate.

Remedies. Finally, what remedies should be offered for those who are able to contest the decision? The answer will likely depend on the relevant stage in the criminal process where the decision was contested. For example, contesting a decision to be included in a list of individuals with propensity to offend will likely lead to an individual's removal from the list and will require the police to provide direct and specific evidence related to criminal activity associated with an individual. Successful challenges at this early stage may fuel broadscale political scrutiny of such lists by highlighting the biases inherent therein and the tenuous constitutional ground they occupy. Alternatively, a successful contestation of an algorithmic risk assessment determination during a sentencing or bail hearing will likely depend on the nature of the contestation, leading to either the consideration of a different score or a new decision based on the rest of the existing evidence.

No doubt, these illustrative questions are just scratching the surface and offer merely preliminary directions. Additional questions, and answers, to the issues addressed deserve additional exploration that is beyond the scope of this Article.³²⁸ For instance, the proposed framework must take seriously the structural difficulties of contesting decisions made by police and, even more so, prosecutors. Closely related are the persistent barriers to securing adequate legal representation across the various stages of the criminal process—challenges that disproportionately affect marginalized and vulnerable communities and are reinforced by the very legal frameworks intended to regulate such processes. Beyond these specific concerns, the framework discussed above should also grapple with the broader normative question of how to ensure equitable treatment and meaningful participation of all individuals before the criminal legal system.³²⁹

These considerations, among others, point to the inherent complexities of establishing a robust contestation framework. At the same time, the proposed framework opens the possibility of imagining legal avenues where individuals subjected to the autonomy-eroding process of de-subjectivization through automation might reclaim their sense of self. Their personhood. Their dignity. Such an outcome is not just warranted from the perspective of the individual, it is in fact required. It is what substantive criminal law considers its most foundational component: the concept of subjective culpability.

³²⁸ See generally Kaminski & Urban, *supra* note 70, at 2001–03 (discussing the right to contest AI and the many challenging questions that remain once this right is established); Black et al., *supra* note 63, at 53 (advocating for the use of model multiplicity to promote less discriminatory algorithms in the non-criminal context).

³²⁹ See *supra* Part IV.B.

CONCLUSION

Although there is a divergence of views regarding the costs and benefits of utilizing artificial intelligence and machine learning tools in different legal systems, there seems to be an implicit agreement in existing scholarship: algorithmic tools—many of which are created by AI—are confined by the laws under which they function. This Article challenges this proposition. By discussing the fundamental criminal law idea of subjective liability, it argues that the algorithmic logic can, in fact, conceptually alter legal doctrine. As such, it contributes more broadly to scholarship on AI and the law by introducing a novel perspective to rethink the potential effects of AI on legal systems.

Within the domain of criminal law and procedure this Article further places the process of adopting algorithmic risk assessment tools in broader contexts, suggesting that it represents a particularly concerning erosion of individual determination of guilt in the modern American criminal legal system. By doing so, this Article also advances new conversations about our penal system. In particular, it demonstrates the sad irony of how a system already so liberty-eroding can introduce new mechanisms to worsen infringement upon the liberty of its subjects through the erosion of a subjective, individualized guilt determination. It further posits that the decay of the individual under the criminal legal system in the algorithmic age is not equally distributed. Rather, it exacerbates existing discriminatory patterns against racially marginalized communities under the system.

To conclude, this Article raises difficult questions about the disparities between the principles and values of criminal law in liberal democracies and the reality of the criminal legal system in the ever-expanding technological landscape. It aspires to offer new directions for conversations about the place of the individual before the carceral state. Last, it offers a legal and practical framework that, though only scratching the surface, forces us to think deeply about how we can restore subjective liability through a more participatory and democratic criminal process. The first step is confronting how technology strips individuals of the very autonomy criminal law purports to respect.

