Developing Artificially Intelligent Justice

Richard M. Re & Alicia Solow-Niederman*

22 STAN. TECH. L. REV. 242 (2019)

ABSTRACT

Artificial intelligence, or AI, promises to assist, modify, and replace human decision-making, including in court. AI already supports many aspects of how judges decide cases, and the prospect of "robot judges" suddenly seems plausible—even imminent. This Article argues that AI adjudication will profoundly affect the adjudicatory values held by legal actors as well as the public at large. The impact is likely to be greatest in areas, including criminal justice and appellate decision-making, where "equitable justice," or discretionary moral judgment, is frequently considered paramount. By offering efficiency and at least an appearance of impartiality, AI adjudication will both foster and benefit from a turn toward "codified justice," an adjudicatory paradigm that favors standardization above discretion. Further, AI adjudication will generate a range of concerns relating to its tendency to make the legal system more incomprehensible, data-based, alienating, and disillusioning. And potential responses, such as crafting a division of labor between human and AI adjudicators, each pose their own challenges. The single most promising response is for the government to play a greater role in structuring the emerging market for AI justice, but auspicious reform

* Professor of Law, UCLA School of Law; 2017–2019 PULSE Fellow, UCLA School of Law and 2019-2020 Law Clerk, U.S. District Court for the District of Columbia. Alicia Solow-Niederman completed this work during her tenure as a PULSE Fellow, and the arguments advanced here are made in her personal capacity. The authors, whose names are listed alphabetically, are grateful to participants at the Yale Law School Information Society Project's (Im)perfect Enforcement Conference, the UCLA School of Law AI in Strategic Context Conference, and the UCLA Department of Political Science Media Immersion and Forms of Life Workshop. Thanks also to Dan Bussel, Kiel Brennan-Marquez, Margot Kaminski, Karen Levy, Martha Minow, Ted Parson, Joanna Schwartz, John Tasioulas, and Eugene Volokh for thoughtful comments and helpful conversations, as well as to Cooper Mayne for excellent research assistance. This research was partly funded by a gift from the Open Philanthropy Project to the UCLA School of Law. Finally, many thanks to the editors of the *Stanford Technology Law Review.*

proposals would borrow several interrelated approaches. Similar dynamics will likely extend to other aspects of government, such that choices about how to incorporate AI in the judiciary will inform the future path of AI development more broadly.

TABLE OF CONTENTS

I. INTRODUCTION

In law as in many other fields, artificial intelligence or "AI" promises to assist, modify, and replace human decision-making.[1] Already, human judges

1. The definition of "artificial intelligence" is often contested but for present purposes encompasses automated computer programs that are capable of replacing or supplementing traditional judicial functions. So defined, AI can have a bodied form (e.g., robots) or operate without any special physical presence (e.g., algorithms). *Cf.* STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 2 (3d ed. 2014) (laying out eight definitions of AI, categorized according to whether AI is understood to mean "thinking humanly," "thinking rationally," "acting humanly," or "acting rationally").

increasingly rely on algorithmic analysis when making bail and parole determinations that affect the freedom of many thousands of people every year.[2] Reacting to that trend, Chief Justice Roberts surprised a university audience in 2017 by noting that AI-assisted adjudication is already "here,"[3] and he more recently implored a class of high school graduates to "beware the robots," lest humans themselves become robotic.[4] Meanwhile, academics have begun to speculate about the prospect of AI adjudicators that ascertain guilt, liability, or sanctions.[5] The image of "robot judges" is so prevalent that even *The Daily Show* has run a segment entitled, "Disrupting the Legal System with Robots."[6] Yet it remains unclear how AI adjudication will change courts, the legal profession, and ultimately the law itself.[7]

This Article argues that foreseeable forms of AI will profoundly change more than the mechanics of adjudication: they will also affect the adjudicatory values held by legal actors as well as the public at large.[8] At

2. For an overview of existing use of algorithms in the criminal justice system, *see Algorithms in the Criminal Justice System*, ELECTRONIC PRIVACY INFO. CTR. (EPIC), https://perma.cc/KY2D-NQ2J. *See also Replacing Bail With An Algorithm*, ECONOMIST (Nov. 23, 2017), https://perma.cc/WZW9-3DF2 (discussing states' use of algorithms for bail determinations).

3. Adam Liptak *Sent to Prison by a Software Program's Secret Algorithms*, N.Y. TIMES (May 1, 2017), https://perma.cc/6CF8-7FMK.

4. Deborah Cassens Weiss, *'Beware the Robots,' Chief Justice Tells High School Graduates*, ABA J. (June 8, 2018), https://perma.cc/FBQ6-4PKT.

5. *See generally* Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, 109 J. CRIM. L & CRIMINOLOGY 137 (2019) (mounting "defense of human judgment that focuses on the normative integrity of decision-making" and advocating "role-reversibility" as a requirement for introduction of "robo-judges"); Eugene Volokh, *Chief Justice Robots*, 67 DUKE L.J. 1135 (2019) (arguing that robot judges should generally be used for adjudicatory functions when they write relevant opinions as persuasively as competent human judges); Ian Kerr & Carissa Mathen, Chief Justice John Roberts Is a Robot, Presentation at WeRobot (2014), https://perma.cc/3EBP-3FHC (offering a philosophically-grounded thought experiment that situates jurisprudence against the capabilities of expert systems, AI, and decision-making by mechanized judges); Cass R. Sunstein, *Of Artificial Intelligence and Legal Reasoning* (U. Chicago Pub. L. & Legal Theory, Working Paper No. 18, 2001) (arguing that computers cannot reason by analogy and therefore cannot engage in legal reasoning, while reserving the question of whether such capabilities might develop in the future).

6. Ronny Chieng, *Disrupting the Legal System with Robots*, THE DAILY SHOW (March 7, 2018), https://perma.cc/9DQ6-MH7E.

7. For an early, insightful analysis of how AI might affect judging, particularly by enabling more accurate textualist reasoning, see Betsy Cooper, *Judges in Jeopardy!: Could IBM's Watson Beat Courts at Their Own Game?*, 121 YALE L.J. 87 (2011).

8. Consistent with present and near-term technical capabilities, this discussion assumes some amount of human supervision and involvement as opposed to fully

present, machine learning or "ML"[9] techniques, particularly "deep learning,"[10] drive most current innovations and applications.[11] Deep learning machines are sometimes likened to auto-educating infants, who can absorb knowledge and skills, such as spatial orientation and language, even in the absence of any formal instruction or "coding."[12] As AI capabilities improve, the perceived distinctiveness of human expertise and insight may decline—and human judgment calls may even come into disrepute. By exposing the inefficiencies and biases of human decision-makers, the increasing use of AI may surface more fundamental questions about the nature and purpose of adjudication. It may, in particular, challenge

autonomous "general AI" systems that operate without human input.

9. "Machine learning" refers to a system's ability to learn without ex ante, explicit programming. This definition is attributed to Arthur Samuel, whose 1959 piece asked how machines can solve problems without being explicitly programmed. *See* Arthur Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. RES. & DEV. 211 (1959). Rather than rely on logical inferences or other familiar modes of human reasoning, ML identifies patterns and correlations in very large volumes of data. For example, a machine-learning system that has analyzed a database of photographs might recognize that pictures with two even and relatively dark spots in the top half of a rectangular frame tend to be coded as "portraits"—without requiring programmers to specify traits associated with human anatomy or portraiture. For more detail on machine learning and its salience for legal scholars and practitioners, see generally David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653 (2017); see also Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89-101 (2014).

10. Deep learning is a subset of ML that generally relies on layers of "neural networks" operating in tandem. For an introduction to the topic, see MICHAEL NIELSEN, NEURAL NETWORKS AND DEEP LEARNING (2017), https://perma.cc/8BTB-CE5U. For a more technical account, see IAN GOODFELLOW ET AL., DEEP LEARNING (2016), https://perma.cc/TYW4-NPD8.

11. AI developers debate how far ML approaches can take the field. *See, e.g.*, JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT (2018) (arguing that a focus on correlation, rather than causation, is an error that will limit prospects for major AI advances); Adnan Darwiche, *Human-Level Intelligence or Animal-Like Abilities?* (July 8, 2017) (unpublished manuscript), https://perma.cc/Q4LH-CCBX; Yann Lecun, *Power & Limits of Deep Learning*, YOUTUBE (Nov. 7, 2017), https://perma.cc/KZ53-JWTC; Gary Marcus, *Deep Learning: A Critical Appraisal* (Jan. 2, 2018) (unpublished manuscript), https://perma.cc/7QXE-3E2L. However, the principal challenges and concerns discussed here are not unique to ML or deep learning.

12. *See* Matthew Hutson, *How Researchers Are Teaching AI to Learn Like a Child*, SCIENCE (May 24, 2018), https://perma.cc/27ZK-MJQ9. *But see* Alex Beard, *How Babies Learn—And Why Robots Can't Compete*, GUARDIAN (Apr. 3, 2018), https://perma.cc/T4FJ-N2C2 (suggesting that complex social processing distinguishes infant learning from machine learning processes).

prevailing understandings of the legal system's proper structure and of what justice requires.

Most fundamentally, AI adjudication is likely to affect both lay and professional views of the law, particularly in areas—including criminal justice and appellate decision-making—where "equitable justice," or discretionary moral judgment, is frequently considered paramount. By offering efficiency and at least an appearance of impartiality, AI adjudication will foster a turn toward "codified justice," that is, a paradigm of adjudication that favors standardization above discretion. In the past, the law has often placed a premium on explicit reasoning in the service of moral and other forms of discretionary judgment, as expressed in written judicial opinions.[13] But AI introduces what is in essence a new kind of adjudication, whereby machines produce correlations across vast amounts of data without constructing an explanatory or causal model.[14] Because market dynamics as well as the general appeal of efficiency will make AI adjudication attractive, new and potent forces will push toward codified justice—at equitable justice's expense. And that shift in values will in turn facilitate greater use of AI adjudication, creating a self-reinforcing cycle. In other words, AI adjudication's *development path* will affect not just how the technology is used, but also the legal system in which it operates. This emerging pattern is already visible in recent headlines as governments the world over reach for technological means of increasing their courts' efficiency, accessibility, and consistency.[15]

That prospect raises several interrelated concerns. Most immediate is the risk that legal decision-making will become incomprehensible and that

13. Over a century ago, Roscoe Pound critiqued the rise of "mechanical jurisprudence" in which jurists applied common law rules without reference to their effects and advocated "sociological jurisprudence" that embraced "pragmatism as a philosophy of law." Roscoe Pound, *Mechanical Jurisprudence*, 8 COLUM. L. REV. 605, 608-10 (1909).

14. Such correlation-based reasoning is the state of the art in AI, and hence the focus of this analysis. *See supra* note 10. Judea Pearl has recently suggested that advances in AI require the field to seek and embrace causal reasoning. *See generally* PEARL & MACKENZIE, *supra* note 11.

15. *See, e.g.*, Eric Niiler, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, WIRED (Mar. 25, 2019), https://perma.cc/C7LA-GMD6 (describing Estonia's plans to use a "robot judge" to reduce case backlogs at low cost); Peng Shen, *Adoption of AI in Chinese Courts Paves the Way for Greater Efficiencies and Judicial Consistency*, BAKER MACKENZIE (Feb. 28, 2018), https://perma.cc/M7BS-Y84Y (discussing the Chinese judicial system's integration of AI to "facilitate improved efficiencies" and increase access to litigation services).

the law itself will adapt to take advantage of rich data sources at the expense of relatively non-quantifiable values, like mercy. Increasing use of AI will also foster lay and even professional alienation from law as adjudication increasingly moves within the exclusive dominion of technical specialists. At the same time, corporate developers and others who sell AI adjudication will have incentives to encourage public disillusionment with the legal system by accentuating the inconsistency and bias of human judgment—and pitching their own services as superior. There are several possible responses, such as endeavoring to combine human and AI adjudicators within a single adjudicatory process. The most important response is for public actors to play a greater role in the "market for justice," whether by structuring the market or by operating as vendors themselves. However, no one response on its own is likely to reverse the coming turn toward codified justice. To maximize the chances of preserving a role for equitable justice, reformers should adopt a set of interrelated responses, each of which can partially compensate for the others' weaknesses.

The issues posed by AI adjudication, moreover, will likely have analogs in other branches of government, as well as in areas outside of government. The present investigation into judicial decision-making accordingly sheds light on governance issues that are likely to emerge more subtly or slowly elsewhere.

II. MODELING AI DEVELOPMENT

This Part argues that AI adjudication is likely to generate a shift in attitudes and practices that will alter the values underlying the judicial system. In particular, AI adjudication will tend to strengthen codified justice at the expense of equitable justice, and an increasing emphasis on codified justice will in turn influence how AI adjudication is developed and used. To many proponents of equitable justice, the coming push toward codified justice will seem intrinsically objectionable. But that is not our focus here. Instead, this Part exposes AI adjudication's self-legitimating power: over time, increasing use of AI adjudication will foster changes in values that are conducive to even greater use of AI adjudication, thereby creating a self-reinforcing cycle.

A. Legal Change

How can a technology like AI catalyze changes in the law? As we will see, AI adjudication's development path will not only foreseeably alter legal rules, but also cause the values underlying those rules to evolve. This Subpart sets the stage for that claim by contrasting two models of legal change: Rule Updating and Value Updating.

1. Rule Updating

The first model is the simpler and more familiar one: legal values prompt the adoption of settled rules, which are applied to technology through judicial application of precedent in particular cases.[16] Over time, innovation generates a new technology that calls for a change within the legal system. The required adjustment of the rules could take the form of a wholly new law or a clarification in how an old law applies in new circumstances. This process repeats as innovation prompts ever-newer technologies. Transitional unsettlement may occur as old rules interact with both old and new technologies. Over time, a new equilibrium may emerge until innovation again intervenes and the cycle begins once more. The point here is not the precise means of change, but rather the idea that, within this model, technology evolves and rules change without altering the underlying values that prompted adoption of the rules in the first place. The values underlying and motivating legal rules remain fixed, even as technologies affect the legal means of achieving those values. Technology's dynamic potential is thus checked, as fixed values continue to guide policymakers when they decide whether and how to adopt new technologies.

16. *Cf.* Gregory N. Mandel, *Legal Evolution in Response to Technological Change*, *in* THE OXFORD HANDBOOK OF LAW, REGULATION, AND TECHNOLOGY 225 (Brownsword et al. eds., 2017) (discussing the traditional legal impulse to extend old legal categories to new technologies); Urs Gasser, Commentary, *Recoding Privacy Law: Reflections on the Future Relationship Among Law, Technology, and Privacy*, 130 HARV. L. REV. F. 61, 64 (2017) (describing three ways law responds to technological change: subsumption within existing rules; gradual innovation and adjustment; and more radical paradigm shifts). This Article assumes that the described pattern occurs within a legal system that, like the United States's, features a mix of rules and standards, with legal precedent, statutory and regulatory texts, and constitutional principles, and with judges exercising considerable discretion in at least some significant range of cases.

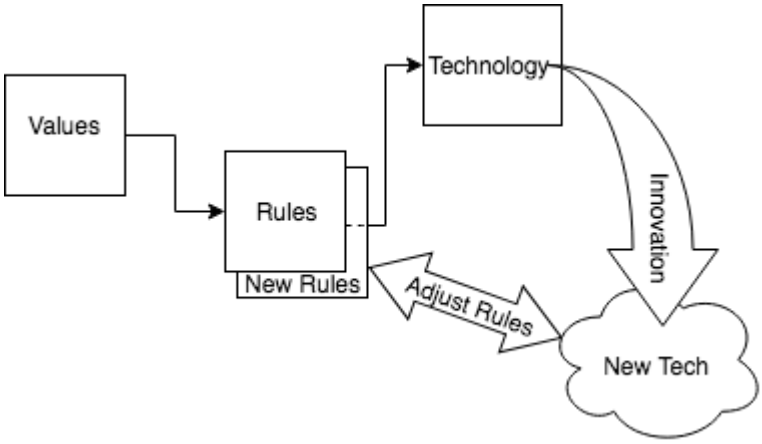

FIGURE 1: RULE UPDATING

2. Value Updating

But technology can have more profound effects. As a social force, new technology can create new capacities, knowledge, social practices, and risks that call into question regulators' preexisting goals. These fundamental changes in the inputs to legal rules can both prompt evolution from old to new rules and, more fundamentally, influence the values that prompt adoption of legal rules.[17] Those possibilities point toward a second model:

17. *See, e.g.*, BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY (2018) (arguing that new technological applications change people's habits and tolerances with respect to values such as privacy); Jack M. Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIR. 45, 49 (2015) ("What we call the effects of technology are not so much features of *things* as they are features of *social relations* that employ those things. These social relations include relationships of power and authority, forms of human organization and production, and features of human sexuality, association, and family life."); Ethan Katsh, *Law in a Digital World: Computer Networks and Cyberspace*, 38 VILL. L. REV. 403, 406 (1993); *infra* notes 25-28 (collecting sources). The values motivating legal change can be viewed as the "legal culture." *See* Lawrence M. Friedman, *Is There a Modern Legal Culture?*, 7 RATIO JURIS 117, 118 (1994) (proposing the concept of "legal culture" as "the ideas, values, attitudes, and opinions people in some society hold, with regard to law and the legal system," and explaining that legal culture is an "*intervening variable*" mediating between "social forces" and "the law").

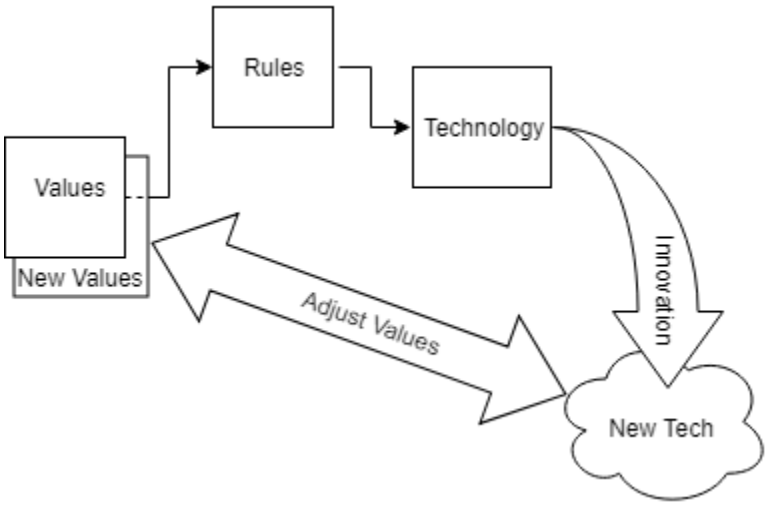

FIGURE 2: VALUE UPDATING

In the model outlined above, values still produce rules, and new technology still results in updated rules. But the nature of this adjustment is different. Even without top-down or conscious reform, the emergence of a new technology can change regulators' objectives.[18] New technology can thereby act as more than an object of regulation. Operating instead as a social force, new technology can yield a discursive interaction between technology, rules, and values.[19] New values destabilize old rules and require rule revision; these new rules then apply to old and new technology alike. Through those changes in values, new technology affects how people view old technology, as well as old rules.[20]

18. *Cf.* Gasser, *Recoding Privacy Law*, *supra* note 16, at 64-65 (arguing for a "recoding" of privacy law by "reframing technology, broadly defined, no longer (only) as a threat to privacy, but as part of the *solution space*").

19. *See* Bert-Jaap Koops, *Criteria for Normative Technology: The Acceptability of 'Code as Law' in Light of Democratic and Constitutional Values*, *in* REGULATING TECHNOLOGIES: LEGAL FUTURES, REGULATORY FRAMES AND TECHNOLOGICAL FIXES 172 (Roger Brownsword & Karen Yeung eds., 2008) ("Normative technology should be added to the list of developments that trigger a reconsideration of what it means to live in a democratic constitutional state."); Ronald Leenes et al., *Regulatory Challenges of Robotics: Some Guidelines for Addressing Legal and Ethical Issues*, 9 L. INNOVATION & TECH. 1, 25 (2017) ("[S]ocial norms affect the technology that is the object of the regulatory action but are also at the same time influenced by the technology: meanings and values change because of changing affordances.").

20. *Cf.* Orin S. Kerr, *The Problem of Perspective in Internet Law*, 91 GEO. L.J. 357, 373,

When can new technology act on values? That is a profound question that we cannot resolve definitively here. For present purposes, however, there are two general and often overlapping answers. First, new technology can alter individual and social *capabilities* in ways that disrupt established practices, catalyzing new practices and related ways of thinking.[21] For example, in rendering personal travel and relocation far less costly, mass public access to automobiles is thought to have enhanced Americans' sense of personal autonomy, as well as its importance.[22] Values changes have likewise been attributed to the invention of oral contraceptives, which facilitated a broader movement for women's equality by allowing greater bodily control and workforce participation.[23] And, most relevant here, Bernard Harcourt has argued that "what triggered the shift in our conception of just punishment from notions of reform and rehabilitation to notions of risk assessment in the late twentieth century is *the production of technical knowledge*," particularly "progress in techniques of predicting criminality."[24]

Second, new technology could facilitate the spread of *information* that disrupts once-established understandings and opinions. Books offer an example, as the printing press has been credited with putting religious and technical texts in the hands of the public in a way that, respectively, fomented the rise of Protestantism and modern science.[25] This dynamic can have negative effects, too; for instance, the internet is sometimes thought to spread extremism by connecting geographically disparate individuals to otherwise inaccessible, radical ideas.[26] In addition to new media formats,

380 n.124 (2003) (invoking the image of old wine in new bottles to discuss technological change and the law).

21. *Cf.* Lydia Bennett Moses, *Regulating in the Face of Sociotechnical Change*, *in* THE OXFORD HANDBOOK OF LAW, REGULATION, AND TECHNOLOGY 573 (Brownsword et al. eds., 2017) (arguing that socio-technological changes due to rapid technological innovation, not new technology per se, produces legal and regulatory challenges).

22. *See* Sarah Seo, *The New Public*, 125 YALE L.J. 1616, 1653-55 (2016).

23. *See* Planned Parenthood v. Casey, 505 U.S. 833, 856 (1992) ("The ability of women to participate equally in the economic and social life of the Nation has been facilitated by their ability to control their reproductive lives." (citing ROSALIND PETCHESKY, ABORTION AND WOMAN'S CHOICE 109, 133 & n.7 (rev. ed. 1990)).

24. BERNARD HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 32 (2007).

25. *See generally* ELIZABETH L. EISENSTEIN, THE PRINTING PRESS AS AN AGENT OF CHANGE (1979).

26. *See generally* CASS SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL

technology can propagate information by allowing private actors to spread information that overlaps with their own marketing objectives. Think of hybrid car manufacturers who raise consciousness of global warming.[27] These sorts of information effects could be stronger for, or limited to, experts or individuals who are especially focused on a particular issue, as opposed to applying with equal force to all lay or non-expert persons.

As we will see, AI adjudication will implicate both of the above ways that technology can act on values. That is, AI adjudication will not only create new, disruptive capabilities but also foster the spread of new information. And these AI-driven developments will affect the ways that humans interact with and relate to the judiciary itself.

B. Adjudicatory Justice

This Subpart briefly discusses two stylized models of adjudicatory justice: equitable justice and codified justice. Of course, these models simplify complex jurisprudential questions and processes, and they both can and do overlap in practice. But delineating and contrasting these two views of adjudication illuminates how AI will challenge current ways of understanding the judiciary and its work. While AI adjudication may seem startlingly new, it relates to longstanding debates about the nature and purposes of law—particularly the desirability of legal discretion.[28]

1. Equitable Justice

Equitable justice entails both reflection on the values set in place by the legal system and the reasoned application of those values, in context. Equitable justice is most visible in discrete judicial rulings that are governed by standards and applied to facts ascertained through individualized proceedings. But even widely applicable decisions governed by positive law

---

MEDIA (2017).

27. *See* Christine Woodside, *Insights on Hybrid Vehicles, Making Global Warming Local*, YALE CLIMATE CONNECTIONS (Oct. 10, 2013), https://perma.cc/GP9G-BZZ5 (discussing meeting of car manufacturers and climate change researchers); *see generally About Us*, HYBRIDCARS, https://perma.cc/35LE-28SG (archived Apr. 24, 2019) ("[T]he Internet's premier Web site dedicated to hybrid gas-electric vehicles and the full range of consumer information and tools about cars, energy, and the environment.").

28. *See generally* KENNETH CULP DAVIS, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY 50-51 (1969).

rules, such as cases involving statutory interpretation, often pose important opportunities for discretionary judgment.[29] Unlike policy development or rulemaking within administrative or legislative contexts, equitable justice aspires to apply consistent principles and is prepared to set aside general patterns in favor of unique circumstances.[30] That discretionary power requires legitimation and comes with constraints. In particular, equitable justice typically carries an obligation to provide particularized, case-specific explanation that connects legal principles, as applied through a lawful process, to the particular facts at hand. Because of its discretionary, contextual, and dynamic nature, equitable justice may seem flatly incompatible with automated algorithmic processes. For instance, can a prearranged decision procedure really incorporate an idea like mercy or develop fact-sensitive balancing of mitigation factors in a criminal case? As we will see, however, the answer is more complicated than appearances may suggest.[31]

2. Codified Justice

Codified justice refers to the routinized application of standardized procedures to a set of facts. Over time, judges can apply these standardized procedures, which constitute a rule set—or a non-computerized "legal algorithm"[32]—to a large number of cases. Codified justice thus predates AI

29. For instance, statutory rulings sometimes recognize implicit exceptions for extraordinary cases, and even determining what contextual factors are (not) relevant in interpreting a text involves discretionary judgments.

30. This understanding of equity accords with modern scholarship that characterizes equity "as a model of decision[-]making that emphasizes case-specific judgment, moral reasoning, discretion, or anti-opportunism." Samuel L. Bray, *The System of Equitable Remedies*, 63 UCLA L. REV. 530, 536 (2016) (citing Jack B. Jacobs, *The Uneasy Truce Between Law and Equity in Modern Business Enterprise Jurisprudence*, 8 DEL. L. REV. 1, 4 (2005)); John Tasioulas, *The Paradox of Equity*, 55 CAMBRIDGE L.J. 456, 457 (1996); Martha C. Nussbaum, *Equity and Mercy*, 22 PHIL. & PUB. AFF. 83, 85 (1993); Henry E. Smith, *Why Fiduciary Law Is Equitable*, *in* PHILOSOPHICAL FOUNDATIONS OF FIDUCIARY LAW 261, 261 (Andrew S. Gold & Paul B. Miller eds., 2014); CHRISTOPHER ST. GERMAIN, DOCTOR AND STUDENT 26-27 (T.F.T. Plucknett & J.L. Barton eds., Selden Soc'y reprint 1974)). In British and U.S. law, equity traditionally allows for the fair vindication of rights when other sources of law afford inadequate relief.

31. *See infra* text accompanying notes 53-59.

32. *See* Tarleton Gillespie, *Algorithm [draft] [#digitalkeywords]*, CULTURE DIGITALLY (June 25, 2014), https://perma.cc/Q47T-JRPM ("[W]e invoke the 'algorithmic' here . . . not [as] the algorithm per se but the insertion of procedure into human knowledge and social experience. What makes something algorithmic is that it is produced by or related

and is often visible in written "code" materials and other actuarial forms of assessment.[33] For example, long before recent AI advances, codified justice was visible in the federal sentencing guidelines,[34] as well as in many similar administrative[35] and quasi-administrative rubrics. In general, codified justice aspires to establish the total set of legally relevant variables in advance, while discounting other facts and circumstances discoverable in individualized proceedings. The basic goal of such standardization is to reduce space for human discretion in adjudication, thereby diminishing opportunities for arbitrariness, bias, and waste, while increasing efficiency, consistency, and transparency. In short, codified justice sees the vices of discretion, whereas equitable justice sees its virtues. Codified justice accordingly tends to eliminate the need for any explanation, restraint, or legitimation apart from adherence to the standardized procedures themselves. In other words, the judge's power and authority are cast as non-discretionary and derivative of whatever entity created the relevant legal algorithm. So when either analog or digital algorithmic systems supplant any of the multifaceted functions of trial and appellate courts, the judiciary's commitment to discretionary, reasoned decision-making will come under strain.[36]

* * *

In one form or another, tradeoffs between equitable and codified justice are present in any adjudicatory system, whether it involves sports referees, private arbitrators, administrative agencies, or pardon boards.[37] So the

to an information system that is committed (functionally and ideologically) to the computational generation of knowledge or decisions.").

33. *See* Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2222 & n.5 (noting the historical use of actuarial tools by parole boards and, more recently, in other stages of criminal justice risk assessment); Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 67 (2017) (discussing use of actuarial tools and the recent move toward "data-centric approach[es] to prediction in sentencing").

34. *See* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1288-90 (2016) (discussing the guidelines' susceptibility to mechanical application).

35. *See, e.g.*, Heckler v. Campbell, 461 U.S. 458, 458 (1983) (permitting "medical-vocational guidelines" that used a "matrix" to allocate disability benefits).

36. A rich and growing body of work has critiqued code-based processes for attempting to implement democratic norms without adequate transparency and accountability. *See, e.g.*, Koops, *supra* note 19; *see generally* Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CALIF. L. REV. 697 (2018); REGULATING TECHNOLOGIES: LEGAL FUTURES, REGULATORY FRAMES AND TECHNOLOGICAL FIXES (Roger Brownsword & Karen Yeung eds., 2008); *supra* note 21.

37. For a thoughtful analysis of how attitudes toward automation in sports could

point here is not that codified justice is always or necessarily preferable to equitable justice. Nor is the inverse true. Both adjudicatory paradigms have appeal, and the appropriate way to balance discretion and standardization in any given context is debatable. As we will see, however, AI adjudication and its tendency to promote codified justice will raise distinctive concerns.

C. AI Development

Applying the framework set out above, this Subpart argues that AI adjudication will generate new capabilities, information, and incentives that will foster codified justice at the expense of equitable justice. AI adjudication will thereby tend to change the law's values, and not merely its rules, in a manner consistent with Value Updating. This process will influence not just the appeal of AI adjudication, but also its form. In short, AI adjudication's early affinity with codified justice will both accelerate the technology's adoption and encourage its development in ways inimical to equitable justice.

The main strengths of AI adjudication are two hallmarks of codified justice: efficiency (or elimination of waste) and uniformity (or elimination of bias and arbitrariness).[38]

As to efficiency, AI adjudication has a capacity for mass deployment at a scale and speed that far exceeds what any human bureaucracy could achieve. An algorithmic decision procedure that draws on ML could resolve an indefinite number of cases and would not be limited by time and space in the way that a human judge or team of human decision-makers would be. Instead of having to train people, monitor their performance, provide safe workspaces, manage personnel benefits and health insurance, and so forth, a single device provided with ample electricity, maintenance, and the necessary data could almost instantly adjudicate a vast number of cases, limited only by computing power and energy resources. Of course, AI adjudication will still make "mistakes," however defined.[39] But for any given

"inform our understanding of public attitudes toward automation in other domains," see Meg Leta Jones & Karen Levy, *Sporting Chances: Robot Referees and the Automation of Enforcement*, (Mar. 31, 2017) (unpublished manuscript), https://perma.cc/65T9-H8U3.

38. See *supra* note 15 and accompanying text.

39. *Cf.* Joshua A. Kroll et al., *Accountable Algorithms*, 165 PENN. L. REV. 633, 656-95 (2017) (expressing concern that automated judgment systems may deliver procedurally irregular outcomes or arrive at results that depart from substantive policy choices and

level of technically attainable accuracy, use of AI adjudication would lower costs.[40] Governments could simply pocket the resulting savings or else spend it to improve algorithmic accuracy. And because the same AI adjudicator could be used to resolve so many disputes—in principle, a single program could clear an entire nation's caseload—it would afford an otherwise impossible degree of uniformity.[41]

The potential benefits are twofold. First, because an identical algorithm would resolve each relevant dispute, AI adjudication could mitigate, or even eliminate, the arbitrariness that results from drawing a "good" or sympathetic judge.[42] Second, standardization of the adjudication process itself could make good on codified justice's promise to eliminate human bias from judicial decision-making.[43] Already, AI adjudication is being

proposing technical remedies to align automated outcomes with legal standards for fairness).

40. *See, e.g.*, Issie Lapowsky, *One State's Bail Reform Exposes the Promise and Pitfalls of Tech-Driven Justice*, WIRED (Sept. 5, 2017), https://perma.cc/H4A3-4CBT (discussing increased judicial efficiency in New Jersey after a state statute mandated that judges use an algorithmic risk score in making pre-trial bail decisions). *See also* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252 (2008) (discussing cost savings as argument made by proponents of automated agency decision-making).

41. *Cf.* Arthur Rizner & Caleb Watney, *Artificial Intelligence Can Make Our Jail System More Efficient, Equitable and Just*, TEX. REV. L. & POL. 181 (2018) (contending that AI may improve pretrial decision-making, as compared to human judgments). Controversy over the use of such tools has led to at least one controlled study, currently in progress, of judges' use of pretrial sentencing tools. *See Pretrial Release*, ACCESS TO JUST. LAB, HARV. L. SCH., https://perma.cc/5XE2-S388 (archived Apr. 24, 2019) (a study funded by Laura and John Arnold Foundation to "test whether an actuarial risk assessment tool that avoids the need for arrestee interviews produces better pretrial incarceration decisions and associated decreases in [failures to appear] and repeat offenses").

42. *See, e.g.,* Alma Cohen & Crystal S. Yang, *Judicial Politics and Sentencing Decisions*, 11 AM. ECON. J. ECON. POL'Y 160 (2019) (noting sentencing disparities that track the defendant's race and the judge's political association); Jennifer L. Peresie, *Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts*, 114 YALE L.J. 1759 (2005) (providing quantitative evidence that the presence of female judges on appellate panels affected the outcome in a subset of Title VII sex discrimination and sexual harassment cases).

43. Standardization in terms of adjudicatory decision-making would not necessarily generate equality or fairness in outcomes. For example, algorithmic decisions could have a patina of consistency only because they are insensitive to case-specific nuances that more flexibly-minded human adjudicators might react to (either explicitly, as a matter of equitable discretion, or subconsciously). *See generally* Danielle Kehl et al., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, RESPONSIVE COMMUNITIES INITIATIVE, BERKMAN KLEIN CTR. INTERNET & SOC'Y, HARV. L. SCH. (2017), https://perma.cc/G2YY-7C79 (providing an overview of risk assessment in sentencing). For a study of the potential benefits of using machine learning

celebrated—and marketed—on precisely these grounds, even as critics raise concerns about the role of human bias in creating adjudicatory algorithms.[44]

AI adjudication will also create powerful new incentives and opportunities to criticize the discretion inherent in equitable justice and, more generally, in human decision-making.[45] Private developers are driving the creation of AI adjudication, and they will have an interest in marketing their wares,[46] including by criticizing the competition. The need to disparage human decision-makers may be especially acute because AI vendors have to overcome the intuitive aversion to their product's involvement in matters of justice.[47] One straightforward solution is to deride human judgment for concealing arbitrariness and bias behind a cloak of discretion. With ever-mounting evidence that humans do exhibit discrimination and cognitive bias,[48] AI adjudicators can promise a more

in parole decisions, see Richard Berk, *An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism*, 13 J. EXP. CRIM. 193, 193 (2017) ("Risk assessments based on machine learning forecasts can improve parole release decisions, especially when distinctions are made between re-arrests for violent and nonviolent crime.").

44. *See, e.g.*, Chris Stewart, *Hey Watson: Local Judge First to Use IBM's Artificial Intelligence on Juvenile Cases*, MY DAYTON DAILY NEWS (Aug. 3, 2017), https://perma.cc/9TAN-4FM3 (quoting a local judge as celebrating use of AI to enhance efficiency and to "standardize best practices, which are not currently uniform").

45. Consider a recent study of asylum cases suggesting that "decision makers exhibit a fair degree of autocorrelation in their rulings, and extraneous factors such as news and the local weather may be impacting the fate of an asylum seeker," to the extent that "granting asylum is predominantly driven by trend features and judicial characteristics." Daniel L. Chen & Jess Eagel, *Can Machine Learning Help Predict the Outcome of Asylum Adjudications?* PROC. ASS'N FOR COMPUTING MACHINERY CONF. ON ARTIFICIAL INTELLIGENCE & L., JUNE 12-26, 2017, at 237.

46. For instance, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm that was used by the state of Wisconsin and contested in the case *State v. Loomis*, and which EPIC reports is used by many other states, is produced by the private company Equivant (formerly Northpointe). *See supra* note 2, *Algorithms in the Criminal Justice System*. For Equivant's marketing materials, see COMPAS Classification, EQUIVANT, (archived Apr. 22, 2019).

47. Some recent polling data regarding self-driving cars suggests that individuals might hold such transformative automated technology to a higher standard than a human analog. *See, e.g.,* Jerry Kaplan, *Why We Find Self-Driving Cars So Scary*, WALL ST. J. (May 31, 2018), https://perma.cc/YS7S-TFRR. But that intuition is easily challenged. *See* Volokh, *supra* note 5.

48. *See, e.g.*, Dan M. Kahan, *Ideology, Motivated Reasoning, and Cognitive Reflection*, 8 JUDGMENT & DECISION MAKING 407 (2013). For a summary of recent research on cognitive bias and decision-making, see Sean Silverthorne, *You Probably Have a Bias for Making*

even-handed and transparent approach.[49] That marketing pitch would take advantage of computerized decision-making's ostensibly "neutral" rationality.[50] And even when AI adjudicators fail, developers can argue for greater research and improvement, conjuring hopes of adjudicatory perfection that no human judge can offer.[51] True, proponents of human judging can be counted on to emphasize the assertedly inimitable wisdom of human discretion. But those arguments will face headwinds as AI achieves greater, perhaps even superhuman accuracy. Even where human judges stay "in the loop," the prospect of more efficient adjudication means that there will still be substantial market driven-pressure for AI systems to guide or constrain the asserted wisdom of human judges.[52] For many, the pitch to invest in "better, faster, cheaper" justice will prove irresistible.

To be sure, the link between AI adjudication and codified justice is contingent, not inevitable. AI adjudication could—counterintuitively—preserve or even foster equitable justice. Courts might integrate some amount of AI decision-making alongside human reflection and deliberation,

---

Bad Decisions. Here's Why., HARV. BUS. SCH. (May 17, 2018), https://perma.cc/WFS2-WU9M.

49. It seems unlikely that existing programs can fulfill this promise. One recent study found that the COMPAS algorithm was no better than non-expert humans in predicting the rate of recidivism. *See* Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES, Jan. 17, 2018, at eaao5580, https://perma.cc/J3Y2-7D29.

50. *See* Mulligan & Bamberger, *supra* note 36, at 711 & nn.43-45 (citing Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 697-701, 710-12 (2010)). Even advanced legal practitioners are susceptible to the allure of sophisticated-seeming technology. For instance, a 2009 report published by the National Research Council's Committee on Identifying the Needs of the Forensic Sciences Community discussed the tendency for courts to accept seemingly scientific forensic techniques without parsing the validity of forensic evidence presented in a particular case. *See* COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 4 (2009), https://perma.cc/MTE9-36P8.

51. Advocates might even position AI adjudication as an improvement on Article III's grant of life tenure for federal judges. If the idea behind Article III is that such job security removes political pressures that might otherwise sway judicial decision-making, algorithmic creators might claim that they can "perfect" the ability to make neutral decisions based only on legal principles and rules by removing humans from the equation entirely.

52. *See, e.g.*, Matt O'Brien & Dake Kang, *AI in the Court: When Algorithms Rule on Jail Time*, PHYS.ORG (Jan. 31, 2018), https://perma.cc/7B5M-MAG9 (quoting a judge's position that algorithms are "not something where you put in a ticket, push a button and it tells you what bail to give somebody," but rather one additional factor for judges to consider).

with the goal of freeing up more time for human decision-makers to exercise equitable discretion. More fundamentally, positioning AI adjudicators as codified and humans as equitable would be overly simplistic. We have already seen that human judges can adhere to codified justice, such as when they apply grid-like decision frameworks. And an AI adjudicator powered by ML and a rich dataset might take into account more, not fewer, fine-grained distinctions than a human judge, thereby creating greater opportunities to adjust for the facts at hand in a case-sensitive manner.[53] An algorithm could in principle parse an unlimited number of mitigation factors and deliver a highly particularized sentence in a criminal case, thereby rendering an even more finely tailored outcome than a human judge would or could deliver.

But while AI adjudication could in principle develop in ways that would foster greater equitable justice, several factors will tend to push the technology's development path in a different direction. Even if interpretable AI eventually becomes feasible,[54] near-term technological limitations will make it difficult for AI adjudication to match either the perceived or the actual competence of human adjudicators when it comes to equitable decision-making.[55] Apart from the general difficulty of programming relatively refined algorithms and the resource requirements in terms of hardware and data scientists' time, there are only so many types of datasets available for programmers to use in generating ML techniques.[56] Limited bodies of training data might curtail data scientists' ability to play with a model and arrive at a working algorithm that sufficiently exhibits equity. Data limitations are, moreover, even more significant to the extent that the

53. Note that such fine-grained distinctions could occur whether or not the AI adjudicator "decides" to act in a particular way. *See* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 539-40 (distinguishing autonomy, or a decision to act in a particular way, from emergence, or "the coupling of complexity and usefulness, the movement of low-level rules to tasks of apparently high sophistication" (citing STEPHEN JOHNSON, EMERGENCE: THE CONNECTED LIVES OF ANTS, BRAINS, CITIES, AND SOFTWARE 18-19 (2001))).

54. *See* Chris Olah et al., *The Building Blocks of Interpretability*, DISTILL (Mar. 6, 2018), perma.cc/NT7W-9VCT (exploring the possibility of combining interpretability techniques to explain decisions made by machine learning systems).

55. As taken up in more detail *infra* Part III.B, it is an open question whether it is technically possible for an algorithm to efficiently provide case-by-case decision-making at scale, while providing evidence of the legal reasoning that undergirds the decision.

56. *See* James Vincent, *These Are Three of the Biggest Problems Facing Today's AI*, VERGE (Oct. 10, 2016, 9:27 AM), perma.cc/P2MS-URJ6.

relevant adjudicatory algorithms are developed by private actors whose information is proprietary and often hidden from public view.[57] Without access to these proprietary data sets, the universe of available information on which to train an equitable AI adjudicator is even more limited. At least in the near term, then, AI adjudication will not embody equitable justice. As a result, early proponents of AI adjudication are likely to celebrate codified justice as a good that their algorithms are better positioned to offer. And this pattern would be self-reinforcing: if purveyors of algorithmic codified justice gain support for their product and amass market share early on, there would be less incentive to make additional investments required to develop algorithmic equitable justice. Early-stage market success of codified justice would then diminish later-stage demand for equitable justice.[58]

Recent experience suggests that market forces are already pushing AI development in directions that challenge, erode, and change preexisting legal values. Take the problem of "explainability," which is discussed in more detail below.[59] In short, current ML-based systems can generate desirable outcomes without providing explanations that humans can understand.[60] But because people accustomed to equitable justice typically

57. On proprietary code, see Natalie Ram, *Innovating Criminal Justice*, 112 N.W. L. REV. 659 (2018); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018). Conversations about the importance of open data, however, are beginning to coalesce. *See, e.g.*, '*Roadmap Zero' to AI and Data Commons*, ITU NEWS (May 25, 2018), perma.cc/P6NL-7T2H.

58. In theory, a disruptive upstart could offer a new innovation in the form of equitable adjudication. However, given the current state of proprietary data sets and the general benefits of scale, it seems more likely that a limited number of purveyors will emerge as market leaders. Already, the need to compete in the highly-competitive AI market has led formerly non-profit AI entities to pivot and rebrand as commercial actors. *See, e.g.*, Greg Brockman, Ilya Sutskever, & OpenAI, *OpenAI LP*, OPENAI (Mar. 9, 2019), perma.cc/LJ6P-UMSX (discussing non-profit's transition into part non-profit, part "capped-profit" company); Kelli Rhee, *A Letter From Our President*, ARNOLD VENTURES (Jan. 28, 2019), perma.cc/ZY74-BBY4 (announcing Arnold Foundation's organizational restructuring into a limited liability corporation). As discussed in Part IV, strong state-backed alternatives early in the development process might alter this trajectory.

59. *See* text accompanying *infra* note 67.

60. Genuine explanations may be particularly difficult to provide within the context of, say, a trial. *See* Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation* 6-9 (Berkman Klein Ctr. for Internet & Soc'y, Working Paper, 2017), perma.cc/Q6XK-EGPZ (arguing it is technically feasible to obtain explanations from AI systems at least some of the time, while recognizing that obtaining an AI explanation may be challenging in situations, like litigation, in which the topic on which an explanation is

expect explanations for legal outcomes, they might demand an AI product that meets that felt need. The result might be products that deliver "automated rationalizations": computer-generated narratives that sound plausible to a human audience but do not reflect the AI's actual decision-making process.[61] Such a process could even consider the receiver's particular characteristics to deliver a message that is uniquely satisfying to the requestor.[62] If enough people become satisfied by the faux explanations, then AI developers might never have any reason to expend the effort and resources necessary to generate *real* ones. Genuine explanation would have become obsolete.

* * *

Increasing AI capacities will likely create new opportunities for codified justice while enhancing its appeal as compared with equitable justice. As a result, AI adjudication will tend to strengthen forces that already push toward greater measurability, objectivity, and empiricism in the legal system. Those dynamics will only reinforce one another as technology makes it appear more feasible to accomplish what many champions of codified justice have long desired. The result is a self-reinforcing cycle of changes in both technologies and values: AI adjudication will tend to make codified justice more appealing, which will in turn make AI adjudication more appealing, and so forth. Codified justice can thus be viewed as both a consequence of technological change and as an intensifying or exacerbating

---

sought is not clear until after the algorithm is run).

61. *Cf*. Carlos E. Perez, *Deep Learning's Uncertainty Principle*, MEDIUM (Apr. 6, 2018), perma.cc/8Z92-K6PA ("This leads me to the inevitable reality that the best we can do is to have machines render very intuitive 'fake explanations.'"). Consider, for example, credit scoring. When calculating a credit score based on an applicant's credit report, "the credit reporting agency will also provide up to five reasons that are most heavily influencing that particular score." *What is a Credit Score?*, MYFICO, perma.cc/L9AP-6MGE (archived Apr. 22, 2019). These reasons tell a plausible story about the factors that influenced denial, yet they are not a complete exposition of the calculations that led to the denial. Moreover, it does not take much imagination to posit a world in which, rather than presenting five factors that are "most heavily influencing" a given metric, the provider is allowed to determine which factors are most likely to satisfy a particular applicant's request and serve them up on demand.

62. Imagine an AI adjudicator whose "opinions" are leavened with personal touches informed by instantaneous social media research. After discovering that a losing party is a Rolling Stones fan, for instance, the AI might comment that "you can't always get what you want" and then play the hit song's refrain. The song's aphoristic familiarity might be both emotionally comforting and cognitively distracting, despite (or because of) its total lack of explanatory value.

factor in its own right: it will facilitate adoption of AI adjudication and also influence how that technology is implemented.

As the next Part shows, increasing use of AI adjudication will influence both the content of the law and the relationship between experts, laypersons, and the legal system in democratic society.

III. CONCERNS

AI adjudication will raise at least four kinds of concern: incomprehensibility, datafication, disillusionment, and alienation. Each concern relates to a distinctive aspect of human adjudication—namely, understanding, adaptation, trust, and participation. AI adjudication will disrupt each of those familiar activities by offering new decision-making technologies that can resolve legal issues in a manner that is qualitatively different from familiar modes of human cognition. In some ways, these disruptions could be beneficial, and AI adjudication may prove to be on net preferable to human adjudication. But the concerns discussed below would still merit serious attention, if only to optimize AI adjudication and avert its greatest risks.

A. Incomprehensibility

Perhaps the most widely appreciated risk of AI decision-making is that it could function in ways that are hard or impossible for humans to comprehend.[63] The nature of this concern may vary depending on the type of AI. For expert systems, the primary change from a pre-AI era seems to be the number of factors that a machine can consider, and the use of AI would not necessarily render the system more of a black box.[64] Other AI methods

63. *See, e.g.*, Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018) (discussing inscrutability and non-intuitiveness in algorithmic decision-making systems). *See also* Cade Metz, *Mark Zuckerberg, Elon Musk and the Feud Over Killer Robots*, N.Y. TIMES (June 9, 2018), perma.cc/B26Z-CMAV (quoting Mark Zuckerberg testifying before Congress: "Right now, a lot of our A.I. systems make decisions in ways that people don't really understand.").

64. There may be limitations if the number of factors the algorithm can consider is exponentially greater than earlier decision models such that the human mind cannot fathom and account for them collectively, even if each of them was individually comprehensible to the designer at the time the program was created.

may be more opaque. Notably, the currently dominant AI method, ML, relies on mass correlations within data to infer sophisticated statistical patterns. ML, moreover, often entails deep learning techniques that lack the explicit logical reasoning or causal inferences that mark conventional human explanations. Deep learning "works" in the sense that the algorithmic outputs achieve measurable ends. But people, especially those without technical training, may be unable to understand how the machine makes decisions, or why.[65] The path from AI inputs to outputs is often or necessarily so complex as to be incomprehensible.[66] That result is at odds with equitable justice, which privileges personal explanations for reasoned decision-making. Three distinct worries arise.

First, AI's potential incomprehensibility could reduce the judiciary's accountability to both the public and to individuals. The lay public and lawyerly experts alike are often thought to believe (perhaps naively) that judicial opinions are at least somewhat indicative of actual human decision-making. And even if a great fraction of judicial opinions, or of the reasoning that those opinions contain, actually consists of post hoc rationalizations, it remains possible to challenge those opinions through argumentation.[67] True, the legal system presently features some deliberate incomprehensibility, such as the "black box" of the criminal jury, but those areas of opacity exist by choice and so are susceptible to certain

65. Interpretability techniques presently require a large amount of technical specialization to implement and must be tailored to deliver meaningful outputs in a given context. *See* Olah et al., *supra* note 54. In making arguments about challenges of interpretability, this Article does not mean to imply that interpretability is formally unattainable in all cases, nor does it intend to advance an unconsidered image of AI as an impenetrable technical object. To the contrary, it is critical to approach the question of interpretability on a case-by-case basis that recognizes how data scientists may be able to "play with the data" at early stages in ways that increase explainability once the data is run. *See* Doshi-Velez et al., *supra* note 60; Lehr & Ohm, *supra* note 9. Even with these points in mind, however, incomprehensibility could be functionally troubling for the legal system for the reasons addressed in this Subpart.

66. *See* Selbst & Barocas, *supra* note 63, at 1094 (addressing machine learning's inscrutability, or the tendency to produce "situation[s] in which the rules that govern decision-making are so complex, numerous, and interdependent that they defy practical inspection and resist comprehension").

67. Human discretion may sometimes rest on judgment calls and decisional processes that can be expressed but not "explained" in any deeper sense. In those cases, equitable justice is comprehensible only because—or if—articulated human judgment is itself viewed as explanatory.

limitations.[68] By contrast, the how and why of AI adjudication might have to remain mysterious in a way that would frustrate public debate and obstruct existing modes of public accountability and oversight, such as impeachment or judicial election. A version of this same concern would play out at the individual level if the legal system appears to work well enough (however measured), while simultaneously failing to satisfy the expectations of each individual litigant.[69] Increased forethought among experts as to, for instance, which methods of statistical modelling are appropriate for a particular context could mitigate some of these concerns,[70] at least insofar as a more interpretable option is feasible. But technical interpretability alone would not suffice to overcome incomprehensibility concerns. Even if it becomes technologically possible to make AI adjudication interpretable, the information that corporations use remains proprietary.[71] Further, even if the AI becomes interpretable, and even if it is not within the control of a private actor, it remains potentially inaccessible to members of the public without a technical background.

Second, the incomprehensibility of an AI adjudicator could pose legitimacy or fairness problems for individuals who are the subjects of AI adjudication. When someone is on the receiving end of the state's monopoly of coercive force, principles of due process suggest that the individual is entitled to at least some understanding of what is happening, and not only for the purpose of contesting adverse decisions.[72] To be ignorant of the law's operation is to be disempowered and vulnerable, akin to the plight of Josef K. The individual without comprehension might thus experience special or separate harms, in addition to any harm that may result from the outcome

68. *See* Peña-Rodriguez v. Colorado, 137 S. Ct. 855 (2017) (providing a limited exception to the "black box" of the criminal jury in cases of alleged racial bias).

69. This dynamic could overlap with the delivery of "automated rationalizations," see *supra* note 62 and accompanying text, that serve up "good enough" explanations to permit corporate success in the market for justice—without actually addressing the reasoning involved in a particular case. And looking beyond a given case, "better, faster, cheaper" today might come with a cost down the line if future would-be plaintiffs find themselves unable to comprehend the relevant legal reasoning in the granular detail necessary to make their case. Incomprehensibility at the level of one plaintiff could thereby effectively deprive a later plaintiff of the ability to draw persuasively on relevant precedents.

70. *See generally* Doshi-Velez et al., *supra* note 60; Lehr & Ohm, *supra* note 9 (discussing statistical choices entailed in ML development).

71. *See* Ram, *supra* note 57, at 720.

72. *See* Citron, *supra* note 40.

of the proceedings.[73] The ideal of equitable justice aspires to avoid those harms, and human decision-makers can live up to that aspiration by providing litigants and the public with formal opinions that explain how the court's judgment applies the law in a given case. By contrast, codified justice via AI adjudication's operation in any particular instance could be incomprehensible not only to the lay people being judged, but also to legal and technical experts.[74] Compounding that problem, AI adjudicators might be less able to foster their own legitimacy, as compared to human adjudicators. For instance, human judges often win the respect of litigants by acknowledging that the losing party's views have some force,[75] but it is an open question whether that psychological effect would still arise when a programmed machine behaves in a similar manner.[76] True, a robot judge could express evenhanded respect for both sides—but would a litigant feel respected when a programmed device simply does what it is programmed to do?

Third, the general incomprehensibility of AI adjudicators might preclude optimal degrees, or desirable forms, of incomprehensibility. The ability to understand an adjudicatory process is not an unlimited good since, in many contexts, adjudicators have good reason to conceal their extant uncertainty on a particular issue or to maintain a degree of future unpredictability. For example, human judges might obfuscate the grounds and limits of their precedential decisions today in order to preserve room for jurisprudential maneuvering tomorrow.[77] And judicial rulings can create

---

73. This harm of incomprehensibility would also fall unequally on different groups, as individuals who are disproportionately likely to be on the receiving end of AI adjudication would often lack the background or expertise that would allow them to take advantage of technical interpretability.

74. *See* Selbst & Barocas, *supra* note 63, at 1097 (internal citations omitted) (addressing the possibility that a ML system will be "non-intuitive," making it impossible for even technical experts to "weave a sensible story to account for the statistical relationships in the model").

75. *See* Tom R. Tyler & Gregory Mitchell, *Legitimacy and the Empowerment of Discretionary Legal Authority: The United States Supreme Court and Abortion Rights*, 43 DUKE L.J. 703, 770-72 (1994).

76. *See* Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. DAVIS L. REV. 1067, 1067-68 (2018) (reporting on an empirical study that suggests ways of making people more likely to perceive algorithmic justice as legitimate).

77. *See, e.g.*, Richard M. Re, *Narrowing Precedent in the Supreme Court*, 114 COLUM. L. REV. 1861, 1891 (2014); Alec Stone Sweet, *Judicialization and the Construction of Governance*, *in* ON LAW, POLITICS, AND JUDICIALIZATION 55, 63 (Martin Shapiro & Alec Stone

desirable "acoustic separation" between relatively lucid rules understood by the public and the more complex practices apprehended by legal experts.[78] Deliberate or even inadvertent ambiguity in human judges' reasoning, explanation, and behavior can thus be a valuable aspect of the legal system, at least some of the time.[79] As Jerome Frank famously put it, "Much of the uncertainty of law is not an unfortunate accident: it is of immense social value."[80] Adjudicators should accordingly aim to foster the right amount or kinds of incomprehensibility, as opposed to eliminating it altogether.[81]

Perhaps an AI adjudicator could one day replicate the desirable forms of ambiguity that humans generate, or even attain a superior blend of explanation and obfuscation. But to the extent that automatic decision-making processes are categorically or unduly incomprehensible—as appears to be the case when AI relies on ML techniques, at least from the perspective of non-experts—AI adjudicators would be unable to take advantage of nuanced blends of ambiguity and clarity, transparency and opacity in the same way as human judges. And to the extent that these decisions are incomprehensible even to technical experts, it would be difficult to ensure that there is adequate oversight and calibration of emerging processes and decision patterns.

Finally, the incomprehensibility of AI adjudication could be unequally distributed among different groups in ways that allow the legal system to be gamed. In many contexts, the ability to understand the causes or grounds of a decisional process allows for its manipulation.[82] And efforts to render AI

Sweet eds., 2002).

78. *See* Meir Dan-Cohen, *Decisions Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 HARV. L. REV. 625, 630-34 (1983).

79. *See* Frank A. Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation,* 87 GEO. WASH. L. REV. 1, 3 (2019) (expressing concern that "[c]ode seeks to eliminate the forms of ambiguity and flexibility characteristic of much language, including legal language" and thereby undermines opportunities for case-specific discretion).

80. JEROME FRANK, LAW AND THE MODERN MIND 7 (1933).

81. Too much transparency regarding judicial processes compromises the legal system. For further discussion of transparency as a value, see generally ARCHON FUNG ET AL., FULL DISCLOSURE (2009); KIM LANE SCHEPPELE, LEGAL SECRETS: EQUALITY AND EFFICIENCY IN THE COMMON LAW (1988).

82. For instance, the IRS does not disclose the "audit flags" used to identify tax returns because such knowledge might allow individuals to game the system. *See* Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. TELECOMM. & HIGH TECH. L. 235, 246 (2011).

adjudication comprehensible, whether through interpretability or another form of transparency, could asymmetrically allow sophisticated actors to adjust their conduct or litigation strategies in ways that would predictably achieve desired results. We might imagine a bifurcated explanatory regime where people receive both a polite-but-useless explanation in readable language along with a detailed technical report that is intelligible only to experts—or to another AI. In that world, only a select set of actors—namely, those with the requisite degree of technical expertise, or the resources to pay someone to provide it—would be able to parse the "real" explanation. AI adjudication's potential incomprehensibility thus raises important distributional considerations.

B. Datafication

The conditions for effective use of AI adjudication will not only influence AI development, but also interact with basic features of the legal system. As we have seen, AI adjudication promises substantial gains in efficiency as well as increased uniformity that promises to root out human bias and arbitrariness.[83] But by focusing attention on seemingly objective data and adapting legal systems to incorporate this information, "datafication," or emphasis on available data and its uses, might undesirably influence the legal system's operation.[84] Datafication most immediately affects the sorts of questions that the system is equipped to address at any given moment. But its effects can also compound over time as the increasing use of AI adjudication creates incentives to accentuate the benefits of big data and reduce its costs, fueling self-reinforcing cycles of datafication that alter the substantive law and affect judicial decision-making. Several concerns may arise.

First, AI adjudication's emphasis on observable data could insulate the legal system from legitimate criticism, thereby allowing bias to flourish. The most often discussed examples involve use of crime data that reflects preexisting racial bias in law enforcement,[85] but in principle any dataset is a product of existing social structures and so is susceptible to analogous

83. *See supra* Part II.C.

84. *See* Citron, *supra* note 40, at 1254, 1271 & nn.146-50 (discussing "automation bias" and compiling sources).

85. *See generally* Dorothy E. Roberts, Book Review, *Digitizing the Carceral State*, 132 HARV. L. REV. 1695, 1708 (2019) (characterizing algorithmic decision-making as "[i]nequality in, inequality out").

problems. To the extent that AI adjudication relies on such biased data, it will recreate or even exacerbate preexisting biases. That process may be insusceptible to effective critique. Existing data, however biased, will be the default basis for decision-making until its defects are shown—not an easy thing to do. Indeed, the algorithms' ostensibly scientific quality will itself placate many observers.[86] Of course, existing human discretion already raises bias problems, and a turn to AI adjudication could make those problems both more visible and more easily corrected.[87] However, promises to improve AI adjudication are a double-edged sword. By appeasing today's critics, these promises facilitate early adoption of the technology. But the decision to adopt a technology can be hard to unwind, such that—once their products are in use—developers will lack a strong incentive to follow through on their stated plans. The result: an elided conversation about *whether* the widespread adoption of the technology is socially desirable in the first instance, coupled with the risk that early reform promises go unfulfilled.

In any event, even if AI adjudication is on balance more perfectible than human judgment, it would still generate a qualitatively new obstacle to future criticism of the legal system. To be effective, would-be critics will have to become more adept at both understanding and explaining the problems with data-based decision-making. That challenge is daunting, as evidenced by early skirmishes over algorithmic fairness, in which critics have jousted with industry participants over the appropriate metric for ascertaining nondiscrimination.[88] While these sorts of expert debates rage on, the allure of efficient, data-based justice will often win out.

Second, AI adjudication that relies on data may be undesirably fixed in ways that human adjudication is not, such that a turn to AI in judicial decisions would ossify the rules of the legal system. This point is most evident at the level of individual jurists. Flesh-and-blood judges often aspire to consistency throughout their careers, but they are subject to several forms of natural updating. For each human judge, updating occurs through

86. *See generally* Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 806 (2014); Bamberger, *supra* note 50, at 711-12 (discussing "automation bias").

87. *See, e.g.*, Simmons, *supra* note 76, at 1070, 1074-75.

88. *See* Sam Corbett-Davies, et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labelled Biased Against Blacks. It's Actually Not That Clear*, WASH. POST (Oct. 17, 2016), https://perma.cc/RV5T-3RJ2.

biological aging as well as exposure to new information and experiences, including non-legal experiences that no lawmaker planned on.[89] And, across judges, turnover from career changes, retirement, and mortality mean that the bench is in a constant state of generational flux. By contrast, AI adjudicators could be fundamentally unchanging, despite substantial exogenous events to which a human judge (or, at longer intervals, a population of such judges) would react. In the aggregate, individual judges' lack of updating might mean that the legal system itself would fail to update.[90] Put more succinctly, AI adjudication might stifle desirable, organic forms of updating, including Value Updating. For example, consider that the U.S. Supreme Court seemed instantly to update its views in light of the attacks on 9/11 and then again after the publicized abuse at Abu Ghraib.[91] Would AI adjudicators be capable of that sort of updating on their own—and in ways that are deemed legitimate? If not, would human governments and programmers be able to agree on how to implement updates to those kinds of events? Further, what might be lost in translation between the formal content of the desired legal updates and the programming language that implements these updates?[92] It is unclear, in the abstract, how to evaluate the loss of natural cultural updating in favor of deliberate software updating. What is apparent is that embracing AI adjudication without incorporating some form of updating risks putting hard data above dynamism and thereby making the legal system too rigid.

Third, increasing use of AI adjudication will accentuate reliance on the data-based modes of decision-making that AI adjudication is capable of performing—at the expense of less quantifiable or data-rich considerations.[93] The very availability of data will create attractive

89. *Cf. Three Supreme Court Justices Later Regretted Supporting the Death Penalty*, AMNESTY INT'L, https://perma.cc/MT32-C3XB (archived May 21, 2019) (reporting that three Supreme Court Justices' "regretted" their 1976 votes to reinstate death penalty).

90. Or, in the extreme, if there were a single "AI judge" for the entire nation's caseload, then there would be no difference between the judge and the system.

91. *See, e.g.*, Linda Greenhouse, *The Supreme Court: Detainees; Access to Courts*, N.Y. TIMES (June 29, 2004), https://perma.cc/2NCF-KZ73 (discussing *Hamdi v. Rumsfeld*, 542 U.S. 507 (2004)).

92. *See, e.g.*, Koops, *supra* note 19, at 160-62 (discussing the translation between "law in the books" and "law in technology").

93. *See* Citron, *supra* note 40, at 1255 (discussing the legislator's incentive to make laws that can take advantage of automation); *cf.* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 132 (2017) ("Using data mining also tends to bias

opportunities to take advantage of that data via ML and other AI techniques.[94] And where there is AI adjudication, there will be pressure to rely on those relatively cheap, comparatively consistent, and at least ostensibly neutral technologies, at the expense of potentially subtle and adaptive human decision-making. AI adjudication could thus resemble the man who looks for his keys only where the light is. These pressures will also influence the nature and perceived purpose of adjudicative decision-making, since AI adjudicators will tend not to focus on variables like personal sincerity or remorse that may not be easily captured by (or inferable based on) observable data.[95] These dynamics, moreover, intersect with market forces given the present role of private entities in developing AI adjudicative techniques.[96] Where data is open source or easily accessible, new entrants will likely seek to use it to compete in potentially lucrative markets; where data is proprietary or difficult to obtain, those who possess it will wish to leverage it to acquire a competitive advantage. In either case, economic incentives will further catalyze the push to develop and deploy data-based decision-making programs.

As AI adjudicators take on a larger role, the kinds of decision-making that those adjudicators can perform will also tend to take center stage, displacing potentially valuable information and forms of decision-making that are not presently observable in data sets. These tendencies will be self-reinforcing, as increasing use of big data will encourage ever greater accumulations of the data types originally possessed—and foster increasing confidence that an AI adjudicator relying on that data can exhaust all relevant considerations.[97] AI adjudication could also "overfit" available data by tailoring its choices too closely to a limited set of data points, yielding

organizations toward questions that are easier for computers to understand." (citing Jon Kleinberg et al., *Prediction Policy Problems*, 105 AM. ECON. REV. 491, 494 (2015))).

94. While the prospect of datafication is particularly salient for AI adjudication that rests on ML, which presently dominates AI approaches and depends on vast amounts of training data, most or all foreseeable forms of AI adjudication will demand copious amounts of quantified information.

95. *See* Roth, *supra* note 34, at 1285-90 ("[T]he removal of judicial discretion to exercise equity and mercy becomes more problematic as the rest of the system, before it reaches the judge, becomes more mechanical as well.").

96. See *supra* notes 42-50 and accompanying text.

97. This consideration is related to the development path of AI discussed *supra* Part II.C: the existence of closed and proprietary data sets early on might permit initial leaders to multiply their starting data advantages, thereby further entrenching confidence in their data-based models.

flawed decisions when the working algorithm is applied in other contexts.[98] Such overfitting would be especially troubling if the initial data reflected systemic social bias, came from entities with relevant pecuniary or political interests, or advanced other discriminatory individual agendas beyond the policies intentionally included in the legal algorithm. And such dynamics could be even more pernicious if an initial mechanized decision triggered follow-on automated steps not amenable to human intervention.[99]

Fourth, the appeal of AI adjudication could create pressure to adapt the law itself to take advantage of what is susceptible to AI treatment.[100] Imagine that AI adjudicators can deliver substantially accurate results in criminal cases by efficiently focusing on measurable data. Imagine, further, that defendants insist on appealing to moral aspects of the law, such as a requirement that murder occurs only when someone kills with a "malignant heart."[101] The human involvement necessary to implement that moral conception of the law could increasingly seem like a distracting inefficiency.[102] A straightforward reform would then present itself:

---

98. *See* AMAZON WEB SERVS., AMAZON MACHINE LEARNING DEVELOPER GUIDE 17-18 (2016), https://perma.cc/KJ6E-Y9XA (archived May 21, 2019) (discussing "overfitting" as well as "underfitting" in machine learning models). This definition is adapted from statistics, where overfitting refers to "the production of an analysis that corresponds too closely or exactly to a particular set of data and may therefore fail to fit additional data or predict future observations reliably." *Overfitting*, OXFORD ENG. DICTIONARIES, https://perma.cc/PSZ2-CSDJ (archived May 21, 2019).

99. Imagine an automated decision about bail that channels a defendant into a particular algorithm for trial on the merits. What if there is an error in the inputs to the initial bail decision, such that a different trial-algorithm should have been used? If the entire process is automated, and perhaps even irrevocable, then initial reliance on faulty inputs will skew the overall process. *Cf.* Jane Wakefield, *The Man Who Was Fired by a Machine*, BBC (June 21, 2018), https://perma.cc/C4BR-UWXD; Ibrahim Diallo, *The Machine Fired Me! No Human Could Do a Thing About It*, ID BLOG (June 17, 2018), https://perma.cc/2ELZ-ZALS (describing the experience of man whose job ended after a flaw in automation lead to his termination, followed by the automated revocation of all of his company credentials).

100. *See* Harry Surden, *The Variable Determinacy Thesis*, 12 COLUM. SCI. & TECH. L. REV. 1, 8 (2011) ("One reason that lawmakers might want to create relatively more determinate contexts is to make them more amenable to computation."); *see also* Citron, *supra* note 40, at 1255.

101. This stance would be consistent with, for instance, contemporary California state law. See CAL. PENAL CODE §§ 187-88 (West 2017) (defining murder as "the unlawful killing of a human being, or a fetus, with malice aforethought" and specifying that malice can be implied when "the circumstances attending the killing show an abandoned and malignant heart").

102. AI adjudication might also erode the appeal of mens rea as a core component of guilt in the criminal justice system, given the opportunity to instead rely on more

demoralize the elements of murder and construct the elements of the crime in a form amenable to codified justice, thus giving free rein to a more "perfect" system of codified justice that AI adjudication makes possible.[103] Similar points could be made about jury nullification or sentencing mercy—both of which have long been substantially curtailed for the sake of codified justice.[104]

C. Disillusionment

The development and use of AI adjudication is already prompting skeptical reconsideration of existing practices.[105] We have seen that proponents of AI adjudication—particularly the profit-motivated firms that develop the technology—will have an incentive to criticize traditional modes of human judging, including its association with equitable justice, and to celebrate the mechanized alternatives linked to codified justice.[106] And those motivated criticisms will often land, given that there are in fact many serious (and often ignored) deficiencies in human adjudicators.

---

efficient, seemingly more objective criteria to determine whether conduct is properly classified as criminal. *Cf*. Richard M. Re, *Imagining Perfect Surveillance*, 64 UCLA L. REV. DISCOURSE 264 (2016) (making a similar point in connection with AI-enabled "perfect" surveillance systems).

103. Related reforms seem likely in areas susceptible to quantifiable measures. Consider recidivism rates and assume for the moment that it is in fact possible to predict, with an acceptable degree of accuracy and without reproducing societal discrimination, the probability that an individual will re-offend. It would be tempting to rely on such codified metrics in making sentencing decisions, and indeed, judges have already begun trying to do so. *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 767 (Wis. 2016). However, that sentencing reform would elide an antecedent policy conversation about whether recidivism is a desirable measure on which to rely in implementing the criminal justice system—notwithstanding recent analyses of its limitations. *See* Jeffrey A. Butts & Vincent Schiraldi, *Recidivism Reconsidered*, PROGRAM IN CRIM. JUST. POL'Y & MGMT., HARV. KENNEDY SCH. (Mar. 15, 2018), https://perma.cc/CS84-UH2C (discussing at least three limitations of recidivism as an "outcome indicator").

104. *See* Andrew D. Leipold, *Rethinking Jury Nullification*, 82 VA. L. REV. 253 (1996).

105. *See*, *e.g.*, Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://perma.cc/G2JH-UN69.

106. For an example of a similar dynamic in the context of surveillance technologies, see Elizabeth E. Joh, *The Undue Influence of Surveillance Companies on Policing*, 91 N.Y.U. L. REV. 101, 102 (2017) ("Through different mechanisms intended to promote their own interests and profits, these [surveillance-technology producing] companies exert control over the police long after their products have been adopted. Private surveillance technology companies wield an undue influence over public police today in ways that aren't widely acknowledged, but have enormous consequences for civil liberties and police oversight.").

Examples include the many cognitive biases, self-interested behaviors, and prejudices that human judges are known to exhibit.[107] Even if AI adjudication also seems flawed and problematic, its relative appeal could still prompt disillusionment as to traditional human judging, diminishing its perceived effectiveness, democratic legitimacy, prestige, and inherent value. AI adjudication, in other words, could cast a bright light on human adjudication—and people might not like what they see.[108] Several implications would follow.

First, disillusionment would erode confidence in the legal system's legitimacy. Insofar as increasing use of AI adjudication prompts people to look more skeptically at human judging, the legitimacy of existing legal activities could be cast into doubt. Whenever a human judge deprives someone of life, liberty, or property, the adversely affected parties could plausibly wonder whether they were victims of the sort of error to which only humans succumb. Criminal convictions are the most salient example, but the principle has no logical stopping point. These trends could catalyze reforms that attempt to check unreliable human discretion within the legal system. For example, legislatures might be more reluctant to delegate authority to judges, and the judiciary itself might feel less willing to issue decisions that might make political waves. This effect is likely to be especially significant for contentious social issues and for high-profile appellate courts like the U.S. Supreme Court, thereby diminishing the reach and impact of the judicial branch.

Second, disillusionment might alter the judiciary's internal composition, culture, and attitudes. One could imagine AI replacing relatively banal judicial functions, leaving room for only a relatively small pool of super-elite and venerated judges. But the appeal of being a human judge may dwindle in a world where human decision-making is criticized and AI adjudicators increasingly make the key decisions. Different sorts of people might then aspire to the bench. Already, commentators have observed the declining

107. For an extreme example, consider empirical evidence that losses by a judge's preferred football team can affect sentencing decisions. *See, e.g.*, Ozkan Eren & Naci Mocan, *Emotional Judges and Unlucky Juveniles*, 10 AM. ECON. J.: APPLIED ECON. 171 (July 2018). See also sources cited *supra* note 42.

108. *Cf*. Mayson, *supra* note 33, at 2251, 2284 ("[P]rediction functions like a mirror. [W]hat prediction does is identify patterns in past data and offer them as projections about future events . . . [P]redictive algorithms transparently reflect inequality in the data from which they are built.").

numbers of political figures appointed to the U.S. Supreme Court, marking a turn toward ostensibly greater legalistic technocracy and a diminished talent for statesmanship.[109] Increasing use of AI adjudication could foster a different and more intensive technocratic turn, as the judiciary's diminishing prestige and authority cast once-venerated human judges as little more than a ministerial support staff for the AI adjudicator—a kind of adjudicatory tech support. So much the worse for the inspiring tradition of American judging as a bulwark of democracy and freedom. These relatively specific changes would dovetail with the previously noted implications of undermining the judiciary's legitimacy: the ultimate effect would be a set of human judges who have less influence, authority, and moxie. Judges might more rarely exercise equitable discretion, and more frequently rely on codified legal rules and standardized norms.

Finally, we can imagine a range of more modest and targeted legal changes resulting from disillusionment. For example, disillusionment specifically with the adversarial system could weaken the political power of practicing attorneys as well as the honor attending lawyerly work. The power of a lawyer's rhetoric, for instance, would count for much less in a legal system where AI adjudicators are capable of ruling on thousands of technically drafted motions for summary judgment. And in the face of its diminished practical utility, any negative characteristics of human lawyerly rhetoric—including the potential to play to bias and ignorance—might become more salient. That more specific form of disillusionment could nudge the legal system away from the adversarial model that characterizes U.S. practice and toward a more inquisitorial system, wherein the judge (or, perhaps, the human judge plus AI assistant) conducts the proceedings, including important aspects of fact-finding and raising legal arguments for consideration.[110] Looking even further ahead, we can imagine a legal system without courts as we know them, wherein contractual disputes, tort claims, and criminal allegations are all posed and "adjudicated" entirely by machine, without the involvement of any human lawyers whatsoever.

---

109. *See, e.g.*, Albert R. Hunt, *The Supreme Court Could Use a Few Good Politicians*, BLOOMBERG NEWS (July 18, 2018), https://perma.cc/Z3UD-NAA9.

110. These dynamics might play out differently in differently structured judicial systems, particularly ones with inquisitorial systems in the first instance. For example, in a system where judges are traditionally viewed as functionaries subordinate to the legislature, AI adjudication might have a smaller effect on the balance of power between branches of government, as well as a diminished impact on the prestige of the judiciary and its human participants.

To be sure, disillusionment could be desirable, particularly where it reflects greater appreciation of the truth. The respect that human judging presently enjoys may be a product of the fact that there is no other means of adjudication available. Human judges' black robes, august courtrooms, sworn oaths, and lengthy opinions may simply obscure the current system's flaws, most notably false transparency, arbitrariness, and discrimination. Disillusionment could then be cast as a necessary first step to reform—including, potentially, greater use of AI.[111] However, disillusionment by its nature pierces a prior construct in which faith has been placed. And here, the cost of pulling back the judicial curtain could be substantial, particularly if the judiciary's current procedures and norms provide some measure of instrumental legitimacy for the legal system. Seen this way, the incorporation of AI adjudication within existing human systems may need to account for the costs of promoting a more cynical view of how judging "actually" functions when it is left to human decision-makers.

D. Alienation

As AI adjudicators play a larger role in the legal system, human participation will change and, in some respects, decrease. Those developments raise the prospect of alienation, or the tendency for some or all people to cease participating in the legal system and even lose interest in its operations. Extreme forms of alienation are imaginable, such as a fully autonomous legal system that operated without any human involvement whatsoever.[112] But much more realistic and modest scenarios also pose risks of alienation. In time, AI adjudicators will likely become capable of performing many discrete tasks presently assigned to human lawyers, judges, and juries, such as making legal arguments, ascertaining the credibility of witnesses, and setting the form and severity of punishment.[113] These tasks could simply shift, en masse, toward AI adjudicators. The result would be a trend away from the language-based deliberation that marks equitable justice and toward a codified system that is controlled,

111. Optimism bias, novelty, techno-appeal, and so on will also prompt overly favorable pictures of how AI adjudication might improve on human counterparts. *See supra* note 86 (collecting sources, including on "automation bias").

112. *See* Lawrence B. Solum, *Artificially Intelligent Law*, 1 BIOLAW J. 53 (2019).

113. *See generally* Volokh, *supra* note 5.

understood, and participated in by computer scientists, corporations, and other technically sophisticated actors—not lawyers or the public. At present, at least some theories of procedural justice cast public engagement with the legal process as an intrinsic good, above and beyond any instrumental benefits,[114] such that alienation would erode the legal system's normative legitimacy. In addition, alienation will generate several practical concerns.

One straightforward reason to worry about alienation is that it would leave important aspects of social life without sufficient public participation and oversight. Whether civic engagement is seen as an instrumental or intrinsic good, alienation threatens existing modes of civic duty and public accountability, even if incomprehensibility problems (discussed above) are overcome. Juries pose an especially apt example. Few people look forward to jury service, yet the jury is often thought to be an important means of public control within the U.S. legal system.[115] If juries were largely supplanted by AI adjudicators, then the public would lose out on a form of legal education. Even apart from knowledge, the public would also sacrifice the experience of governing, including its sense of empowerment and responsibility.[116] Similar if less pat stories could be told about public participation and debate over state court elections for human judges, or about entertainment programs (like *Judge Judy*) that propagate simplified versions of legal judgment. Again, as these familiar, human modes of doing law subside, the operation of law might seem that much less interesting, relevant, and subject to the control and care of everyday people. Alienation thus leads to a concern about the distribution of decision-making power, as

114. *See, e.g.*, Lawrence B. Solum, *Procedural Justice*, 78 S. CALIF. L. REV. 181, 275-81 (2004) (discussing "the participation that is essential for legitimacy," arguing that "[p]rocedures that purport to bind without affording meaningful rights of participation are fundamentally illegitimate," and contending that "the legitimacy of adjudication depends on affording those who are to be bound a right to participate, either directly or through adequate representation" (internal citations omitted)). Issues of public participation are likely to present even greater challenges in non-judicial contexts, particularly the legislative branch. *See infra* Part V (Conclusion).

115. *Cf*. Powers v. Ohio, 499 U.S. 400, 406 (1991) ("The opportunity for ordinary citizens to participate in the administration of justice has long been recognized as one of the principal justifications for retaining the jury system." (internal citations omitted)).

116. Juries have long been theorized as central to the American system of democratic governance. *See generally* ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 284-91 (Henry Reeve trans., 1850).

sophisticates—particularly corporate actors—play a larger, and apparently more legitimate, role in managing the legal system.[117]

A comparison might be drawn with complex administrative regulations that ostensibly rely on specialized expertise while remaining largely inaccessible to the public.[118] The agency comparison suggests that AI adjudication could allow public participation to thrive in a new form. Rather than participating at the point of law-application in trials as they are presently carried out, the public could participate and exert analogous control via higher-order regulation, much as currently takes place, at least in principle, via public notice-and-comment practice connected with agency regulations. For example, legislators could campaign on various modifications to an AI adjudicator's decisional programming. For that sort of solution to work, however, members of the public would have to understand the stakes and effects of various forms of AI adjudication.[119] Experience with the extant notice-and-comment process suggests that such an administrative approach has at best limited potential to foster public participation[120]—and would not closely replicate the trial process that presently occupies such a large role in the public consciousness.

* * *

To some extent, AI adjudication's appeal cannot be disentangled from fundamental questions of justice and the state's duty to preserve the legitimacy of the justice system. Clearly, both equitable justice and codified justice have virtue, and the law has long struggled to place the two in a desirable balance. Perhaps AI adjudication's distinctive features, including its tendency to promote codified justice, will move the legal system toward a more desirable new equilibrium in which more efficient and uniform

117. *See* Pasquale, *supra* note 79 (arguing that an automated legal system necessarily "shifts personal responsibility from attorneys, regulators, and judges, to those coding their would-be replacements").

118. Here, the AI can be cast as the administrative "expert." If enough of the public accepts and/or private forces and state actors endorse the benefits of AI adjudication, notwithstanding these concerns about alienation, then the legitimacy of the AI method seems likely to spread from the judiciary to administrative and legislative settings.

119. A critical underlying question is whether there could be adequate public participation in and oversight of outcomes without the ability to understand the highly technical process. Put differently, if the *legal process* shifts to a black box while the public retains some oversight of *inputs* and *outputs*, would that form of public participation mitigate alienation? The answer likely turns in part on the underlying theory of the value of participation for procedural justice and democratic legitimacy.

120. *See, e.g.*, Thomas O. McGarity, *Some Thoughts on "Deossifying" the Rulemaking Process*, 41 DUKE L. J. 1385, 1386 (1992).

justice helps to combat human limitations and biases. But even if so, we have seen that AI adjudication would still pose an array of concerns that merit a deliberate response. The next Part surveys the options.

IV. RESPONSES

We have now seen that AI adjudication will tend to facilitate its own adoption by altering underlying legal values in ways that tip the scales toward codified justice. But equitable justice still has its supporters, and AI adjudication already has critics. Further, the legal system's increasing focus on codified justice is bound to generate new pushback. In many contexts, people often leap to embrace new technologies that promise perfect enforcement of formal rules, only to gain a new appreciation of imperfect enforcement.[121] So AI adjudication's tendency to promote codified justice will likely face opposition rooted in both existing values and countervailing social forces. In general, concerns about the rise of AI adjudication and the likely decline of equitable justice prompt several possible responses, but each raises its own difficulties. This Part discusses four types of response before briefly considering how they might operate in tandem to maximize the chances of preserving an important role for equitable justice.

A. Experimentation

In the face of so much uncertainty regarding future technological and social change, the most prudent responsive strategy may be to postpone decisive action and trust existing institutions and systems to muddle through. For example, someone worried about codified justice's potential for disillusionment and alienation might hold out hope that presently unforeseen and unforeseeable adjustments in the legal system might arise to dampen any adverse effects. After all, there is always some lag time before the legal system recalibrates to new technology or any other systemic change. Still, there are ways of making the lag time shorter and the relevant transition period less socially disruptive.

121. *See, e.g.*, Daniel Solove, *Do We Really Want Perfect Law Enforcement?*, PRIVACY & SECURITY BLOG (Oct. 12, 2005), https://perma.cc/S8KG-LH3C (discussing opposition to traffic cameras); David Pozen, *What Are the Rules of Soccer?*, BALKANIZATION (June 20, 2019), https://perma.cc/A3DH-HN9Y (criticizing use of Video Assistant Referees to "close the gap between the paper rules and real rules of penalty kicks" in soccer).

The most promising way to facilitate gradual, imperfectly foreseen legal adaptation is to embrace uncertainty through a policy of experimentation. To some extent, experimentation will happen naturally as different jurisdictions and governmental agencies interact with competing companies. But experimentation can also operate as part of a deliberate program, in which it might take either a regulatory form (for instance, sunsets on particular processes, along with pre-specified metrics to assess the efficacy of a particular approach) or a technical form (such as beta testing of technology in low-stakes cases, perhaps with litigants' consent).

However, it is difficult to experiment in a space where human liberties are at stake while simultaneously preserving the stability of the legal system and minimizing risks of disillusionment and alienation on a broader scale. When highly individualized interests are adjudicated in particular cases, a process of trial and error may be ethically unacceptable, particularly from the standpoint of the individual participants in the legal system. And in the meantime, the population as a whole could become disillusioned with a system that makes bad judgment calls as it moves toward a new balance.[122] The severity of these risks may depend on how long it will take to reestablish an acceptable equilibrium as well as in the contexts in which the AI tool is first or relatively quickly deployed.

Assuming that experimentation is desirable or at least practically inevitable, what is its optimal pace and scale? Adjusting to rapid AI advances would likely require ongoing experimentation as new technologies and regulations are put in place,[123] and that fluidity counsels in favor of fast

122. Imagine an ML system that government officials rely on to determine when the risk of domestic abuse rises to the level that a child should be taken out of the custody of their parents. If researchers revealed that this system's generally effective pattern correlation sometimes relied on discriminatory proxies that human oversight could have avoided, the sense that mechanized injustices had been committed might create disillusionment with the child welfare or legal system as a whole. For a real-life example of an algorithmic tool that interacts with human decision-makers to determine which families are most in need of intervention, see Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. TIMES (Jan. 2, 2018), https://perma.cc/NKM2-CSGG ("In August 2016, Allegheny County became the first jurisdiction in the United States, or anywhere else, to let a predictive-analytics algorithm . . . offer up a second opinion on every incoming call, in hopes of doing a better job of identifying the families most in need of intervention.").

123. *See* Allan Lavell et al., *Climate Change: New Dimensions in Disaster Risk, Exposure, Vulnerability, and Resilience* 48-56, *in* MANAGING THE RISKS OF EXTREME EVENTS AND DISASTERS TO ADVANCE CLIMATE CHANGE ADAPTATION (Christopher B. Field et. al, eds. 2012) (discussing how to cope with and adapt to changing risks and complex, ever-changing dynamics in

cycles of change and assessment. Yet the need for consistency and predictability—important aspects of the rule of law—would push for less frequent or substantial changes over time.[124] It is hard to know, moreover, which decision-makers would control these changes, and with what degree of accountability. Lay persons, technical experts, and legal experts all have different perspectives on and knowledge of AI, and ongoing regulation of AI adjudication would presumably require that all of those groups participate. But how to facilitate such engagement, over time, is not obvious. Experimentation may be attractive because it at least offers a framework for accommodating the extensive uncertainty in this area. But that observation only postpones the question of whether and how to preserve equitable justice in the face of changing adjudicative technologies. In other words, what kinds of experimentation are most appropriate? The remainder of this Part explores that deeper question.

B. Coding Equity

A more technically ambitious approach would integrate some measure of equitable justice into AI adjudication by "coding equity" into the AI adjudicator itself.[125] A "program of discretion" might then enable case-specific discretion, at least in some contexts. This response could take two basic forms. It could be hardwired at a particular point in time to reflect social or legal consensus regarding an issue. Alternatively, and perhaps preferably given the risk of locking in a baseline definition of equity that is "aligned" with extant values, the coding might be updated at regular intervals (presumably through intentional human action but conceivably by an autonomous AI) to ensure that it evolves in tandem with any changed values. A machine capable of dispensing "AI equity" could also mitigate the problem of datafication by being even more responsive than human judges

---

the climate change context).

124. *Cf.* Burnet v. Coronado Oil & Gas Co., 285 U.S. 393, 406-07 (1932).

125. On AI's ability to foster equity, see text accompanying *supra* notes 53-54. As one example, in the criminal law context, the value of equity might appear in appeals to "mercy." In a civil law context, it may hark back more generally to the Aristotelian ideal of preventing rigid, rule-bound application of the law in unjust ways. For further discussion of "equity," see *supra* note 30. *Cf.* Doug Glanville, Opinion, *Baseball's Unwritten Rules*, N.Y. TIMES (Apr. 6, 2018), https://perma.cc/RSL3-294T (discussing "grace" as a value in baseball). Of course, greater discretion could just as easily allow for greater retributive harshness, instead of merciful forbearance.

when it comes to a case's subtle factual nuances or changes in social values. And this approach also has the longer-term virtue of preserving space for equity even if AI brings about an autonomous legal system that functions without any human involvement.

Coding for equity is not a straightforward fix, however, in either a technical or a normative sense. It is not clear whether it is even technologically possible to code for nuanced equitable correction in cases when strict application of a legal rule might seem unjust.[126] And the feasibility of coding equity ultimately may turn more on the social and political meaning of "equity," however defined and implemented, than upon technical considerations. In particular, lay people and experts alike may want to distinguish between what we currently recognize as (human) legal judgment and the technical process whereby a machine makes a "judgment call" based on programmed considerations or obscure patterns in datasets.[127] Codified equity, in other words, may not seem equitable at all.

There would, moreover, be further related concerns involving incomprehensibility, disillusionment, and alienation. Unless the code is comprehensible to a human, it is not clear how to ensure that consideration of equity is what the AI system is actually doing,[128] yielding a lack of public trust and oversight. And it would be challenging to reach meaningful democratic agreement on the content of coded equity, which inevitably implicates evolving notions of justice. Trust and participation in the legal

126. Part of the difficulty stems from the structure of the existing legal system and its constraints on human decision-making. Presently, there is a recognized difference between preexisting legal rules and judges who apply the rules to reach substantive outcomes. The gap between predetermined rules and substantive outcomes is where judges exercise discretion. But when it comes to AI judges, that separation between process and substance may no longer exist: an automated protocol would embody *both* the predetermined rules *and* the decision-maker that applies them to reach outcomes. As a result, there may no longer be room for judgment or equitable justice. *Cf.* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1061 (2019) (discussing "substantial residual discretion" that remains "even when a written protocol is used" and noting that "[a]lgorithmic criminal justice represents a categorical rejection of . . . ad hoc, situated judgments as an instrument of regulations").

127. *See* Mireille Hildebrandt, *Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics*, 68 U. TORONTO L.J. (Supp. 1) 12, 15 (2018) (distinguishing a machine learning system's "performance rating" from "the performativity of a legal judgment by a court of law"). See also *supra* note 67 and accompanying text. *Cf.* James Moor, *Are There Decisions Computers Should Never Make?*, 1 NATURE & SYS. 217, 217-22 (1979) (assessing whether computers can "make decisions" in the first instance).

128. *See supra* Part III.C.

system would grow more difficult, as public debate increasingly took on a technical aspect inaccessible to the public and even legal experts.

Finally, there is a more basic problem with hoping to instantiate equitable justice within a computerized process. Efforts to codify equity attempt to solve the problems posed by AI's development path by demanding a more equity-driven kind of AI program. But that demand, operating alone, will not alter AI adjudication's tendency to diminish the appeal of equity itself. Unless something alters the development path—not merely the AI program at any given point in time—AI adjudication will operate in a technical and social context more conducive to long-term prioritization of codified justice. Coding for equity is thus a tempting but ineffective stand-alone solution.

C. Division of Labor

A third tack would endeavor to preserve a traditional role for humans within systems of AI adjudication, even if that role introduces increased opportunities for bias, arbitrariness, error, and cost. The point of this division of labor is to allow human participation to mitigate the concerns associated with AI adjudication and codified justice. This approach can be viewed as a compromise that would aspire to achieve the best (and avoid the worst) of both adjudicatory paradigms. In general, combined human and AI decision-making could take two basic forms that, respectively, emphasize blending or separating the two decision-makers within a given decisional stage.

First, human and AI judges might collaborate by operating in tandem at specified stages of the judicial process, either by functioning with a human in-the-loop or by preserving an extra measure of human oversight and involvement at particular points.[129] AI adjudication with a human in-the-loop could take many forms. One possibility is to insist upon paired AI/human decision-making at key stages of a judicial proceeding that seem to pose an especially acute threat to life, liberty, or property interests at the core of equitable justice. That sort of approach might also alleviate concerns about a lack of public participation in adjudication and thereby better preserve the legitimacy of the legal system. Alternatively, human oversight

129. *Cf.* Joi Ito, *Society in the Loop Artificial Intelligence*, JOI ITO BLOG (June 23, 2016), https://perma.cc/6C9U-X93S; Iyad Rahwan, *Society in the Loop*, MEDIUM (Aug. 12, 2016), https://perma.cc/9HT3-57DX.

of mechanized learning systems might be especially key at the front-end or back-end of a legal decision—a sort of legal corollary to the role of technical decision-makers who interact with data during supervised machine learning or reinforcement learning. This latter vision is in keeping with current risk assessment methods that provide an automated "score" that a human judge can implement at their discretion.[130] Either way, this strategy would blend human and machine within the same decisional stage, such that human judges would neither act alone nor wholly delegate their authority to the machine.

A second form of human/machine division of labor would apportion discrete types of judicial decision-making to human as opposed to mechanized actors.[131] The resulting separation could be based on subject matter, such as a rule barring automated judging in criminal cases. Or it could derive from more fine-grained determinations about which parts of a legal decision raise concerns about equitable and codified justice.[132] For example, some types of fact-finding could be well-suited for mechanization, without a commensurate cost in disillusionment and alienation, so long as there is a human judge who engages in the analytically severable task of applying the facts to the law. Even within appellate courts, a split in judicial function between human rule-generation and mechanized rule-application might be desirable. More broadly, codified justice already marks key aspects of many bureaucratic legal systems, and AI adjudicators might simply offer a better version of codified justice, limited to those contexts.[133]

The prospect of striking a division of labor is appealing and, in some form, may be the most auspicious response available. Notably, AI is already

130. *See, e.g.*, Loomis, 881 N.W.2d at 767 (finding no due process violation where risk score was not "determinative" factor in sentencing, but merely advisory information for judge). However, the substantial risk of automation bias makes it uncertain whether human judges will in fact override automated recommendations. *See* Citron, *supra* note 40, at 1254, 1271 & nn.146-50 (discussing "automation bias" and compiling sources).

131. Though the main text refers to a singular human or AI judge for expository clarity, several different human judges or different AI adjudicators could act at different stages.

132. Frank Pasquale recently emphasized the "limits of legal automation" throughout the entire trial process, arguing that automation may "elide or exclude important human values, necessary improvisations, and irreducibly deliberative governance." Pasquale, *supra* note 79, at 1.

133. Again, the federal sentencing guidelines are an especially prominent example. See *supra* text accompanying note 34 and sources cited therein.

making great inroads in pre-trial and post-trial judicial decision-making but has not yet entered the trial itself.[134] That pattern is partly a result of AI adjudication's current technological limitations. But it also has to do with the trial's special role within the legal system: at least in popular imagination, jury trials offer a paradigmatic context for equitable justice.[135] The idea of mechanized verdicts, especially criminal verdicts, therefore seems to cut at the heart of democratic self-government, as well as due process.[136] So long as that intuition persists, the trial could serve as a focal point for efforts to preserve a role for human judgment within an increasingly automated legal system. AI adjudication, in other words, could be constrained by existing legal practices and norms.

But there are challenges here as well. Most fundamentally, we have already seen that the values underlying reverence for juries are themselves mutable. Even without AI adjudication, the realities of trial-level adjudication are a far cry from idealistic or dramatic narratives characterized by exercises of equitable justice.[137] And we have seen that AI adjudication will put increasing pressure on the legal system's overall commitment to equitable justice. The process of Value Updating could be depicted as a choice: the legal system must either reconcile AI adjudication with long-held values, such as by adopting a division of labor, or else reconceive those values. Given the dynamics we have already discussed, the trend might well be toward compromising or recasting values relating to equitable justice, including jury trials and due process. So even if the jury trial is among the last places where equitable justice will still reign supreme, its erosion even in that forum may only be a matter of time.

There are, moreover, pragmatic difficulties that will make it difficult to divide human and AI tasks in ways that will desirably preserve human discretion. For one thing, any division of labor would have to preserve certain predetermined roles for humans, and there is no reliable way to know the right balance of human and AI activity until various combinations are attempted via experimentation. Retaining a human in the system, for

134. *Cf.* Huq, *supra* note 126 ("Algorithmic tools are used in three main criminal-justice contexts: policing, bail decisions, and post-conviction matters.").

135. *See, e.g.*, 12 ANGRY MEN (Orion-Nova Productions 1957).

136. *See* Brennan-Marquez & Henderson, *supra* note 5.

137. For example, both civil and criminal trials are dwindling in frequency, and jurors are discouraged or disabled from using their power of nullification on equitable grounds.

instance, could succeed in preserving the legal system's preexisting public legitimacy—but only by objectionably sacrificing efficiency and uniformity that pure AI adjudication would otherwise offer. Pursuit of human-AI collaboration, in other words, could end up being more like the worst of both worlds than the best if the wrong policy tradeoffs are struck. And the guiding policy choice may not flow from measured consideration; instead, undemocratic market dynamics might end up dictating how those hard choices are made—a problem discussed in the next Subpart.

D. Market Intervention

A final set of responses would focus on "the market for justice," or the development and procurement of jurisprudential AI tools. As we have seen, AI adjudication fosters efficiency and so will exert a gravitation force on both public and private actors, prompting them to emphasize codified justice. But proponents of equitable justice could attempt to entrench their values by influencing the development of AI adjudication.

One approach would be to remove profit-seeking actors from the market for jurisprudential tools, such that the judiciary would purchase jurisprudential tools only from non-profit entities who have ideological commitments to balance the allure of efficiency. More ambitiously, the government itself could produce a "public option" jurisprudential tool for key purposes, such as criminal justice. Such an algorithm might take advantage of government data that has not been publicly disclosed,[138] thereby offering an institutional counterweight to proprietary datasets. That approach would also be more democratically legitimate insofar as it is created by actors who are motivated by more than profit alone, particularly if the state were willing to accept financial losses in order to advance non-market goals such as maintaining citizen faith in the political system.

Another straightforward way to entrench equitable justice values would involve formally establishing certain values-protecting requirements, either by statute or via a comparably permanent mode of regulation. A state or federal statute might, for instance, create a set of transparency or accountability requirements to win a government bid in a particular adjudicatory setting. Or the law might impose requirements

138. *See* DATA.GOV, https://perma.cc/5DU4-KXRU (archived May 21, 2019) (providing 300,295 federal, state, and city datasets).

relating to adjudicatory fairness, distributional equality, and other values that sometimes come into tension with the more market-oriented values of efficiency and profit. But legislators employing that general approach will be caught in a bind. If they impose rule-like regulations, then rapid advances in technology will quickly render the legislators' efforts obsolete. And if they instead adopt flexible standards, then the legislators will have failed to insulate the legal system from gradual changes that erode the core equitable values.[139] After all, judges and others applying the standards would, over time, feel the same pressure to prioritize codified justice. For similar reasons, even a constitutional rule—the most permanent mode of regulation available—would be subject to gradual erosion.

To mitigate those tendencies, present-day regulators could adopt a more dynamic approach. Rather than trying to set fixed, lasting requirements, legislators might in effect delegate regulatory authority to institutions that are independently committed to the preservation of existing legal values. Those protective institutions would need to have both the sophistication to keep up with changes in technology and an interest in resisting pressures toward codified justice. In addition, legislative controls, either in the form of substantive commands or detailed procurement procedures, would need to be put in place to ensure that such protectors become durably entrenched.[140] This institutional entrenchment is the key: a relatively fixed and foreseeable role in the development process will tend to create an incentive for private firms to develop algorithmic processes capable of meeting more equitable standards than would prevail in an unmediated market.

Some regulations already attempt this mode of protective entrenchment. Consider Allegheny County's Family Screening Tool (AFST), "a predictive risk modeling tool designed to improve child welfare call screening decisions."[141] Before adopting the tool, the Allegheny County

139. This problem is related to the Collingridge dilemma, which involves the challenge of properly timing any state intervention in a sociotechnical process. *See* DAVID COLLINGRIDGE, THE SOCIAL CONTROL OF TECHNOLOGY (1980).

140. *Cf.* Catherine Crump, *Surveillance Policy Making By Procurement*, 91 WASH. L. REV. 1595 (2016).

141. *Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening Decisions*, ALLEGHENY COUNTY ANALYTICS (Mar. 2017), https://perma.cc/XH94-4SN4 (archived June 3, 2019) (including three reports: (1) a report on the model's development by a team of data scientists; (2) an ethical analysis prepared by Tim Dare

Department of Human Services consulted with an independent team of ethicists and has subsequently made a summary of the analysis publicly available. This strategy might be more broadly applicable: by engraining a formal role for independent ethicists, a legislature could privilege actors that are less likely to be influenced by a turn toward new values, such as codified justice. Similarly, the California Money Bail Reform Act of 2017,[142] which abolished the state's money bail system and adopted a new system that relies on algorithmic risk assessment,[143] grants discretion to county-level superior courts to create a local risk assessment instrument that is selected from a list of tools validated by the statewide Judicial Council."[144] Such upfront state validation could embed public-minded values in any tool that is later adopted in a particular judicial context.[145]

But ultimately, efforts to shape the market for justice are best viewed as efforts to level the playing field, rather than as mechanisms to lock in a predetermined outcome. Because AI technologies are bound to create spokespersons with vested interests in the "faster, better, cheaper" model of adjudication, the most that a proponent of contemporary legal values can realistically hope for is a counterbalancing institutional home for less market-oriented considerations, like distributional equality. By establishing those institutions now, the law can take significant if incomplete steps toward demarcating criminal justice and other areas as domains in which public actors must both lead and exercise ongoing oversight. This kind of

---

of the University of Auckland and Eileen Gambrill of the University of California, Berkeley; and (3) Allegheny County's response to the ethical analysis).

142. Implementation of this statute and associated judicial rules has been stayed, pending the results of a November 2020 popular vote on Referendum 1856. *See* Referendum to Overturn a 2018 Law That Replaced Money Bail System with a System Based on Public Safety Risk, Referendum 1856 (18-0009) (2020), https://perma.cc/95K5-9HP7; Pretrial Release or Detention: Pretrial Services, S.B. 10, 2018 Cal. Stat. ch. 244, https://perma.cc/T7EU-MVXP.

143. CAL. PENAL CODE § 1320 (West 2017).

144. CAL. PENAL CODE § 1320.7(g) (West 2017) ("The entity, division, or program, at the option of the particular superior court, may be employees of the court, or employees of a public entity contracting with the court for those services as provided in Section 1320.26, and may include an entity, division, or program from an adjoining county or one that provides services as a member of a regional consortium. In all circumstances persons acting on behalf of the entity, division, or program shall be officers of the court.").

145. However, substantial questions remain regarding both the specific validation process contemplated under this statute and the overarching technical and policy tradeoffs that such upfront validation entails. For more detail, see Alicia Solow-Niederman et al., *The Institutional Life of Algorithmic Risk Assessment*, 35 BERKELEY TECH. L.J. (forthcoming 2019), https://perma.cc/KAZ4-S9Y2.

dynamic response would probably do more than any other to preserve a role for equitable justice, at least in particularly sensitive domains.

* * *

To the extent that AI adjudication poses concerns, no single response offers a perfect solution. Instead, each available response has inherent limitations, faces feasibility challenges, or threatens new problems. Moreover, none can entirely remove the pressures pushing AI adjudication toward codified justice. Yet considered collectively, the four responsive approaches do suggest the outlines of a plausible path toward preserving a significant role of equitable justice while simultaneously reaping the genuine benefits of codified justice. In short, the various approaches' strengths seem capable—at least in principle—of mitigating their respective weaknesses.

To illustrate as much, consider the experimentation and division of labor approaches. On the one hand, experimentation can help illuminate how to divide authority between AI and human decision-makers. And, on the other hand, a division of labor could mitigate the problems with experimentation by, say, limiting the initial zone of AI experimentation to legal functions or cases with relatively low personal stakes. A similar mutually reinforcing relationship exists between coding equity and market interventions. Equity-preserving code, once in operation, could help to lock in place certain legal practices, thereby serving as a bulwark against the erosion of equitable justice that would take place in a totally open "market for justice." At the same time, market interventions can play a critical role in prompting and enabling non-profit organizations to update code over time, thereby allowing equity-preserving programming to keep up with changing social dynamics.

In sum, efforts to preserve a significant role for equitable justice and address concerns with AI adjudication should draw on all four responsive approaches outlined above, using each to offset the others' liabilities.

V. CONCLUSION: COURTS AND BEYOND

Increasing use of AI adjudication will fundamentally alter practical capabilities, institutional incentives, power relationships, and, ultimately, the views of experts and laypersons alike. The likely result will be the ascendance of values associated with codified justice. And that development will in turn foster further use of AI adjudication, creating a self-reinforcing

cycle that poses concerns relating to incomprehensibility, datafication, disillusionment, and alienation.

AI adjudication thus offers both an illustration of how technological change can yield Value Updating and an important case study in the ways that the turn to AI will influence human decision-making, both individual and social. Yet increasing AI adjudication characterized by codified justice is not an inevitable endpoint. Rather, AI adjudication's critics, as well as proponents of equitable justice, may respond in ways that preserve their preferred values in particular contexts, or even allow for their improved realization.

Similar dynamics will also affect the appeal of codified justice within other government institutions. After all, if codified justice becomes preeminent in the judiciary, where the opposing tradition of equitable justice is strongest, then AI's influence on values will likely be even greater in executive bureaucracies and administrative agencies.[146] Take international humanitarian law and the law of war, which is heavily grounded in moral determinations but generally implemented outside courts. If codified justice ascends in the judicial branch, would it be possible to resist its influence in this other, perhaps even more sensitive domain?[147] Or would a desire to avoid cognitive dissonance and the push to take advantage of efficiencies lead government officials and the public to embrace new kinds of decision-making in, say, executive military commands?

As AI's increasing use in the judiciary makes codified justice more appealing in other contexts, its downsides are likely to be reproduced, too. Problems analogous to the ones discussed above will likely arise. And the basic menu of responses, with all their limitations, is also likely to recur. Finding a path forward will require attention not only to technology and law, but also to technology's impact on conceptions of justice, in both its human and artificially intelligent forms.

---

146. *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L. J. 1147, 1184-91 (2017); *see also* William Boyd, *Environmental Law, Big Data, and the Torrent of Singularities*, 64 UCLA L. REV. DISCOURSE 544, 546-47 (2016).

147. *Cf.* Ashley Deeks, *Predicting Enemies*, 104 VA. L. REV. 1529 (2018) (discussing the effects of military use of algorithmic prediction and noting transparency and other concerns).