2019

"Cyborg Justice" and the Risk of Technological-Legal Lock-In

Rebecca Crootof
University of Richmond - School of Law, rcrootof@richmond.edu

COLUMBIA LAW REVIEW FORUM

VOL. 119 NOVEMBER 20, 2019 PAGES 233–251

"CYBORG JUSTICE" AND THE RISK OF TECHNOLOGICAL–LEGAL LOCK-IN

*Rebecca Crootof**

INTRODUCTION

An Estonian team is designing an artificially-intelligent (AI) agent to adjudicate claims of €7,000 or less, with the aim of clearing case backlog.[1] The pilot version will focus on contract disputes: An algorithm will analyze uploaded documents to reach an initial decision, which can then be appealed to a human judge.[2] This is but one of many examples of how AI is increasingly being incorporated into adjudicatory or adjudicatory-adjacent processes, ranging from the relatively minor to the most significant determinations.[3] For example, DoNotPay, an AI-based chatbot originally designed to assist with parking ticket appeals, now offers a panoply of legal services, including helping individuals report suspected discrimination, request maternity leave, and seek compensation for transit delays.[4] Companies rely on automated processes to resolve customer complaints,[5]

* Assistant Professor of Law, University of Richmond School of Law; Affiliated Fellow, Information Society Project. This Piece is based on remarks given at the *Columbia Law Review*'s 2019 Symposium, "Common Law for the Age of AI," in response to Tim Wu's *Law's End?* manuscript. Many thanks to BJ Ard, Douglas Bernstein, Hannah Bloch-Wehba, Jennifer Dayrit, Frank Pasquale, Richard Re, Alicia Solow-Niederman, and Harry Surden for clarifying edits and suggestions. I am also indebted to those who have thought far longer on the immediate and long-term promises and pitfalls of incorporating AI into the judicial process. See, e.g., pretty much everyone thanked above and cited below.

1. Eric Niiler, Can AI Be a Fair Judge in Court? Estonia Thinks So, WIRED (Mar. 25, 2019), https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/ [https://perma.cc/Z8DH-APBL].

2. Id.

3. "Intelligence" is a slippery term. See Shane Legg & Marcus Hutter, A Collection of Definitions of Intelligence 2 (2007), https://arxiv.org/pdf/0706.3639.pdf [https://perma.cc/K54T-DE34] (surveying various definitions of "intelligence"). This Piece uses the term "artificial intelligence" broadly, encompassing varied systems and computer programs that can assist or replace human decisionmakers. Cf. Harry Surden, Artificial Intelligence and Law: An Overview, 35 Ga. St. U. L. Rev. 1305, 1307 (2019) [hereinafter Surden, Artificial Intelligence] (defining AI as "using technology to automate tasks that 'normally require human intelligence'").

4. Sebastian Anthony, Chatbot Lawyer, Which Contested £7.2M in Parking Tickets, Now Offers Legal Help for 1,000+ Topics, Ars Technica (July 14, 2017), https://arstechnica.com/tech-policy/2017/07/donotpay-chatbot-lawyer-homelessness/ [https://perma.cc/KWC5-V6XD].

5. Rory Van Loo, The Corporation as Courthouse, 33 Yale J. on Reg. 547, 564–66 (2016) [hereinafter Van Loo, The Corporation as Courthouse].

cities depend on automated systems to generate citations for minor traffic violations,[6] police departments engage data mining systems to predict and investigate crime,[7] and domestic judges employ algorithms throughout the criminal justice process.[8] Meanwhile, militaries are incorporating increasingly autonomous systems in their targeting and detention decisionmaking structures.[9]

Although AI is already of use to litigants,[10] to legal practitioners,[11] and in predicting legal outcomes,[12] we must be cautious and deliberate when incorporating AI into the common law judicial process.[13] As will be discussed in Part I, human beings and machine systems process information and reach conclusions in fundamentally different ways, with AI being particularly ill-suited for the rule application and value balancing

---

6. Lawrence B. Solum, Artificially Intelligent Law, BioLaw J., no. 1, 2019, at 53, 58, http://rivista.biodiritto.org/ojs/index.php?journal=biolaw&page=article&op=view&path%5B%5D=351&path%5B%5D=317 (on file with the *Columbia Law Review*) [hereinafter Solum, Artificially Intelligent Law].

7. Andrew Selbst, Disparate Impact in Big Data Policing, 52 Ga. L. Rev. 109, 113–14 (2017).

8. See, e.g., John Logan Koepke & David G. Robinson, Danger Ahead: Risk Assessment and the Future of Bail Reform, 93 Wash. L. Rev. 1725, 1729 (2018); Rebecca Wexler, Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System, 70 Stan. L. Rev. 1343, 1346–48 (2018).

9. Rebecca Crootof, The Killer Robots Are Here: Legal and Policy Implications, 36 Cardozo L. Rev. 1837, 1868–72 (2015); Ashley Deeks, Predicting Enemies, 104 Va. L. Rev. 1529, 1561–63 (2018).

10. See Anthony, supra note 4 (describing how an AI-driven chatbot provides free legal services).

11. See Surden, Artificial Intelligence, supra note 3, at 1329–31 (discussing the utility and limitations of AI-assisted document review in discovery).

12. Id. at 1331–32.

13. There are similar and distinct reasons to be concerned about how algorithmic decisionmaking may affect government decisionmaking more broadly. See, e.g., Hannah Bloch-Wehba, Access to Algorithms, 88 Fordham L. Rev. (forthcoming 2019) (manuscript at 6–9), https://ssrn.com/abstract=3355776 (on file with the *Columbia Law Review*) (cataloging proprietary algorithmic governance programs and discussing the potential utility of transparency mechanisms); Danielle Keats Citron, Technological Due Process, 85 Wash. U. L. Rev. 1249, 1252–55 (2008) (considering how to maintain due process protections in an increasingly automated administrative state). Relatedly, concerns regarding these and private uses of algorithmic decisionmaking are prompting calls for regulating it writ large. E.g., Margot E. Kaminski, Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability, 92 S. Cal. L. Rev. (forthcoming 2019) (manuscript at 104–10), https://ssrn.com/abstract_id=3351404 (on file with the *Columbia Law Review*) (proposing a hybrid model of individual process rights and systemic regulation through collaborative governance of algorithmic decisionmaking); Andrew D. Selbst & Solon Barocas, The Intuitive Appeal of Explainable Machines, 87 Fordham L. Rev. 1085, 1089–90 (2018) (cataloging concerns with algorithmic decisionmaking). While relevant to those discussions, this Piece focuses on the issues relevant to the use of AI in U.S. common law adjudication.

often required of human judges. Nor will "cyborg justice"[14]—hybrid human–AI judicial systems that would attempt to marry the best of human and machine decisionmaking and minimize the drawbacks of both—be a panacea.[15] Part II notes the benefits of such systems and outlines how teaming may create new overtrust, undertrust, and interface design problems, as well as second-order, structural side effects. Part III explores one such side effect of hybrid human–AI judicial systems, which I term "technological–legal lock-in." Translating rules and decisionmaking procedures into algorithms grants them a new kind of permanency, which creates an additional barrier to legal evolution. By augmenting the common law's extant conservative bent, hybrid human–AI judicial systems risk fostering legal stagnation and an attendant loss of judicial legitimacy.

Cyborg justice systems are proliferating, but their structure has not yet stabilized. We now have a bounded opportunity to design systems that

---

14. "Cyborgs" are cybernetic organisms, beings with both organic and biomechatronic body parts. "Cyborg justice" systems are less colorfully described as "hybrid human–AI judicial systems." They can be conceptualized generally as augmented decisionmaking systems, in which an AI assists a human judge as a kind of career law clerk, Eugene Volokh, Chief Justice Robots, 68 Duke L.J. 1135, 1141, 1148–49 (2019), or as a tiered system, in which AIs address low-level complaints and cases, which can then be appealed to a human reviewer, Niiler, supra note 1. Of course, these structures could be combined: An AI could make an initial assessment, which could then be appealed to an AI-assisted human decisionmaker. Richard M. Re & Alicia Solow-Niederman, Developing Artificially Intelligent Justice, 22 Stan. Tech. L. Rev. 242, 282–85 (2019), https://www-cdn.law.stanford.edu/wp-content/uploads/2019/08/Re-Solow-Niederman_20190808.pdf [https://perma.cc/3SKY-ULYK].

15. Many scholars are considering the implications of human–machine teaming in the context of adjudicatory decisionmaking. See, e.g., Frank Pasquale, A Rule of Persons, Not Machines: The Limits of Legal Automation, 87 Geo. Wash. L. Rev. 1, 46–49 (2019) (cautioning against AI supplanting human decisionmakers); Re & Solow-Niederman, supra note 14, at 262–78 ("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."); Volokh, supra note 14, at 1161–77 (responding to theoretical and practical objections to the use of AI judges); Tim Wu, Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems, 119 Colum. L. Rev. 2001, 2001–05 (2019) (arguing that AI will be incorporated into, rather than replace, U.S. adjudicatory structures); Kirsten Martin & Ari Ezra Waldman, Privacy and the Legitimacy of Automated Decision-Making 23 (unpublished manuscript) (on file with the *Columbia Law Review*) (finding empirically that "including either notification or human oversight removes the legitimacy penalty of automation"); cf. Paul Scharre, Centaur Warfighting: The False Choice of Humans vs. Automation, 30 Temp. Int'l & Comp. L.J. 151, 151 (2016) (arguing that "[h]ybrid human-machine cognitive architectures will be able to leverage the precision and reliability of automation without sacrificing the robustness and flexibility of human intelligence"); Andrew Selbst, Negligence and AI's Human Users, 100 B.U. L. Rev. (forthcoming 2019) (manuscript at 44–45) [hereinafter Selbst, Negligence], https://ssrn.com/id=3350508 (on file with the *Columbia Law Review*) (highlighting that, while AI decision assistants are "sold as a remedy to the weakness of human decisionmaking," incorporating AI "does not actually solve bounded rationality; rather, it transforms the problem . . . into an inability to completely oversee the decision mechanism").

realize the benefits and mitigate the issues associated with incorporating AI into the common law adjudicatory process.[16]

I. HUMAN VS. AI JUDGES

The conversation around whether and when algorithmic decisionmaking processes might augment or replace human judges is far from new.[17] Over the past decade, however, the rise of big data and advances in computation power have enabled dramatic advances in machine learning,[18] which have fostered a renewed interest in the varied societal roles AI might assume. After all, if an algorithm can file a legal complaint, why can't it decide one?

Certainly, there is much to critique about the justice of human judges.[19] They are famously inconsistent, both as a group and as individuals. Not only will decisions vary from judge to judge,[20] any one judge's sensitivity to context and penchant for leniency may vary dramatically with whether they are hungry, tired, bored, overworked, overwhelmed, or otherwise distracted.[21] Further, human judges are an inherently expensive and limited resource: They must prepare for years,

16. Cf. Gaia Bernstein, When New Technologies Are Still New: Windows of Opportunity for Privacy Protection, 51 Vill. L. Rev. 921, 921 (2006) (noting that the possibility of changing a technological design becomes less likely over time, due both to closure and path dependence); Meg Leta Jones, Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw, 2018 U. Ill. J.L. Tech. & Pol'y 249, 281 (discussing how the design and uses of new technologies are constructed differently in different societies).

17. See, e.g., Anthony D'Amato, Can/Should Computers Replace Judges, 11 Ga. L. Rev. 1277, 1277–78 (1977) (speculating in 1977 about whether the search for the "rule of law" may be answered by computer programs replacing judges, and if so, whether society would benefit).

18. E.g., Tim Hwang, Computational Power and the Social Impact of Artificial Intelligence 4–6 (2018), https://arxiv.org/pdf/1803.08971.pdf [https://perma.cc/Z8EF-UQ24].

19. See, e.g., Re & Solow-Niederman, supra note 14, at 272–73 (discussing how firms interested in marketing AI adjudication will be incentivized to highlight human biases and disparage human decisionmakers).

20. See Pasquale, supra note 15, at 49 ("[T]here are almost always conflicts among the approaches of multiple courts to similar sets of facts, irreconcilable by logic or reason."); Volokh, supra note 14, at 1156–57 (observing that "[h]uman judges, . . . being human, have human prejudices" and detailing the many forms they might take).

21. See, e.g., Ozkan Eren & Naci Mocan, Emotional Judges and Unlucky Juveniles 3 (Sept. 9, 2016) (unpublished manuscript), https://ssrn.com/abstract=2914649 (on file with the *Columbia Law Review*) (finding that unexpected losses in Louisiana State University football games correlated with local juvenile court judges giving increased sentences for the following week).

they take time to decide cases, and they retire. As a result, over and over again, justice delayed becomes justice denied.[22]

Given human limitations, advocates for legal automation tend to view delegating judging to algorithms as the natural next evolutionary step toward "a rule of law, not of men."[23] Unlike arbitrary, expensive, and temporally and spatially limited human beings, algorithms appear relatively consistent,[24] cheap,[25] fast,[26] scalable,[27] and (apparently) impartial,[28] suggesting the possibility of fair access to justice for all.

Absent a breakthrough in machine intelligence or a complete structural shift in what we value about the common law process,[29] however, this idealized AI judge is infeasible. It may be possible to create a program that can calculate taxes or make an initial assessment of whether a contract has been breached and still be impossible to create a generalist program that can apply legal rules in accordance with changing social mores.

This inability stems from the fundamental differences in human and machine intelligence. Our familiarity with human intelligence encourages the mistaken assumption that artificial intelligence operates similarly,[30] that it reaches determinations by "engaging in some sort of synthetic computer cognition that matches or surpasses human-level thinking."[31] Instead, today's AI produces results "by detecting patterns in data

---

22. Fred Shapiro, You Can Quote Them, Yale Alumni Mag. (Sept./Oct. 2010), https://yalealumnimagazine.com/articles/2967-you-can-quote-them [https://perma.cc/8239-PDVE] (recounting various origins of the saying, "justice delayed is justice denied").

23. Pasquale, supra note 15, at 4 (citing Margaret Jane Radin, Reconsidering the Rule of Law, 69 B.U. L. Rev. 781, 781 (1989)).

24. Martin & Waldman, supra note 15, at 7 ("Consistency is one of a computer's chief advantages over human decision-makers."); see also Re & Solow-Niederman, supra note 14, at 268–69 ("Flesh-and-blood judges often aspire to consistency throughout their careers, but they are subject to several forms of natural updating.").

25. Re & Solow-Niederman, supra note 14, at 255–56 ("[F]or any given level of technically attainable accuracy, use of AI adjudication would lower costs.").

26. Id. at 255 (noting that "an algorithmic decision procedure that draws on [machine learning] . . . could almost instantly adjudicate a vast number of cases, limited only by computing power and energy resources").

27. Id. ("AI adjudication has a capacity for mass deployment at a scale and speed that far exceeds what any human bureaucracy could achieve.").

28. Id. at 256 ("[S]tandardization of the adjudication process itself could make good on codified justice's promise to eliminate human bias from judicial decision-making.").

29. Id. at 246 (arguing that incorporating AI in the judicial process will foster a shift in societal values, favoring codified justice, which prioritizes standardization, at the expense of equitable justice, which values discretion and mercy).

30. Cf. Jack M. Balkin, 2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data, 78 Ohio St. L.J. 1217, 1223–24 (2017) (discussing the "homunculus fallacy"—the "belief that there is a little person inside the program who is making it work—who has good intentions or bad intentions, and who makes the program do good or bad things").

31. Surden, Artificial Intelligence, supra note 3, at 1308.

and using knowledge, rules, and information that have been specifically encoded by people into forms that can be processed by computers."[32] It "excel[s] in narrow, limited settings, like chess, that have particular characteristics—often where there are clear right or wrong answers, where there are discernable underlying patterns and structures, and where fast search and computation provides advantages over human cognition."[33]

In contrast, the judgment we value in a common law process is a distinctively human skill.[34] Human judges are sensitive to context, both to extenuating circumstances in individual cases and shifts in social norms over time, and can flexibly apply legal rules. While human contextualization may be incorporated during the design or training of an AI system, that is hardly the same as having human contextualization at the time the algorithmic rule is applied, especially as that application may occur in a temporally, geographically, and culturally different context.[35] AI may be consistent, but it is "brittle": "[It lacks] the flexibility humans have to step outside their instructions and apply 'common sense' to adapt to novel situations."[36]

Human judges have a few additional comparative advantages. While both human and AI judges may be black boxes, human judges must give reasons for their decisions.[37] Because human judges live in the real world and internalize social norms, those norms inform and undergird their reasoning, which in turn strengthens those norms (for better or worse, depending on one's view of the norm at issue). Additionally—and possibly relatedly—decisions made by human judges currently have a higher perception of legitimacy than those of algorithmic judges,[38] suggesting

---

32. Id. This description of AI programming roughly encompasses both machine learning and rule-based systems, as most AI systems today incorporate elements of both. Id. at 1319.

33. Id. at 1309.

34. See, e.g., Kaminski, supra note 13 (manuscript at 118–19) (discussing how human decisionmakers fill in context based on cultural knowledge about appropriate decisional heuristics, including by both expanding the decisional context to include information that would be unfair to ignore and circumscribing the decisional context by ignoring information that would be unfair to include).

35. Id. (noting that algorithms often are "fed both goals and datasets by humans who are . . . remote from a particular decision and [which] are often or even inherently culturally or contextually incomplete" and that human contextualization is "absent at the end point when an algorithm is applied to a particular individual").

36. Scharre, supra note 15, at 151.

37. Cf. Katherine J. Strandburg, Rulemaking and Inscrutable Automated Decision Tools, 119 Colum. L. Rev. 1851, 1864 (2019) ("Reason giving is a core requirement in conventional decision systems precisely *because* human decisionmakers are inscrutable and prone to bias and error . . . .").

38. A recent empirical study found that the more involved an AI is in a legal decision, the lower its perceived legitimacy. Martin & Waldman, supra note 15, at 23. However, the data source for the decision—whether the data were gathered specifically for the decision or aggregated by a third party—was *far* more influential to the perceived legitimacy than

that their decisions are more likely to be accepted and enjoy higher compliance rates.[39]

Nor can we translate our statutes and common law into easily applied rules.[40] While it is tempting to imagine law as subject to algorithmic application, reality is far messier.[41] AI can apply unambiguous rules,[42] but even apparently simple laws are far from unambiguous. For example, different groups of coders were tasked with automating the enforcement of New York traffic speed limits.[43] Consider what rules you would include. Would you issue a ticket for every infraction? What if a driver only went four miles-per-hour over the limit? What if they went eight? Would it matter what other drivers were doing? Unsurprisingly, programmers charged with coding the "Letter of the Law" and the "Intent of the Law" differed in many ways, including whether their software ticketed minor infractions,[44] how much the presumed margin of sensor error was biased in favor of the driver,[45] the extent to which "clusters" of violations were considered single violations,[46] and whether there was a minimum interticket time.[47] As a result, for the same scenarios, the "Letter of the Law" group issued 498.33 (standard deviation = 453.42) tickets on average,

---

the decisionmaker. Id. at 4–5, 21; see also Solum, Artificially Intelligent Law, supra note 6, at 59–61 (discussing how democratic legitimacy, transparency legitimacy, and constitutional role legitimacy would all be implicated by the delegation of legal authority to artificially intelligent systems).

39. Wu, supra note 15, at 2002 ("As between a decision made via software and court adjudication, the latter, even if delivering the same results, may yield deeper acceptance and greater public satisfaction."); cf. Tom Tyler, Why People Obey the Law 57 (2006) (arguing that when people view authorities as legitimate, they are more likely to obey their rules).

40. The desire to do so, with the aim of automating law, has given fresh relevance to the "rules vs. standards" debates. Cf. Lawrence Solum, Legal Theory Lexicon: Rules, Standards, and Principles, Legal Theory Blog (Sept. 6, 2009), https://lsolum.typepad.com/legaltheory/2009/09/legal-theory-lexicon-rules-standards-and-principles.html [https://perma.cc/H55J-4KEU] (summarizing the pros and cons of rules, standards, and principles).

41. See, e.g., Pasquale, supra note 15, at 18–22 (discussing ambiguities that arise in the context of translating health privacy law into code); id. at 24 (noting that litigants can disagree on the meaning of seemingly obvious words, such as "chicken"). But cf. Volokh, supra note 14, at 1138–39 (arguing that, if software can pass a "Modified John Henry Test," where it performs at least as well as approximately ten average performers in a field, it should be considered an adequate substitute for a human being, including a judge).

42. See Surden, Artificial Intelligence, supra note 3, at 1323 ("AI tends to work well for tasks that have definite right-or-wrong answers, and clear, unambiguous rules."); see also Harry Surden, The Variable Determinacy Thesis, 12 Colum. Sci. & Tech. L. Rev. 1, 6–8 (2011) (proposing guiding principles for automating legal reasoning, on the theory that some legal concepts are relatively determinable).

43. Lisa A. Shay, Woodrow Hartzog, John Nelson & Gregory Conti, Do Robots Dream of Electric Laws? An Experiment in the Law as Algorithm, *in* Robot Law 274, 275 (Ryan Calo, A. Michael Froomkin & Ian Kerr eds., 2016).

44. Id. at 280.

45. Id. at 282–83.

46. Id. at 283–85.

47. Id. at 285–86.

while the "Intent of the Law" group averaged a mere 1.50 (standard deviation = 5.68) tickets for the same scenario.[48] As the experiment's authors concluded, "[E]ven relatively narrow and straightforward 'rules' can be troublingly indeterminate in practice."[49] And much of common law judging requires more than applying unambiguous rules.[50] Can a program master analogical reasoning, with all of its nuanced distinctions and value judgments?[51] Or apply the complex and changing understandings of "cruel and unusual punishment," "reasonable expectation of privacy," or other legal standards that are intentionally vague, as they were intended for "a world where law enforcement was imperfect or discretionary"?[52]

Further, despite its veneer of objectivity, AI will not solve the problem of human bias; it incorporates human bias and adds other kinds. AI systems can produce biased results due to (1) preexisting bias, which is present in the training data sets and encoded in the system design; (2) technical bias, which emerges from a system's limitations, such as loss of context and simplified formulations that accompany attempts to translate reality into code; and (3) emergent bias, which arises from user interaction with specific populations.[53] These sources of bias afflict all AI systems, but some of these issues are extenuated in the context of judicial decisionmaking. Bias in many legal training data sets arises from the fact that "law is one of those domains where high-quality, machine-processable data is currently comparatively scarce except in particular niches."[54] And, as noted above, bias in the system design arises because the same legal rule may be programmed differently—and those differences breed disparate and discriminatory impacts. Bias will also result from technical limitations, as evidenced by the New York traffic law example, and—to the extent the AI decisionmaker is a learning system—from interaction with judges and judged populations.

Unintended glitches and intended interference from malicious actors create other potential sources of error.[55] Any sufficiently complex system will have various parts that interact in unpredictable ways, and AI

48. Id. at 289.

49. Id. at 278.

50. Cf. Pasquale, supra note 15, at 48 ("In both transactional and litigation contexts, it was almost impossible for any truly knowledgeable professional to boil down the sum total of their knowledge and judgment into a series of propositions applicable by machine.").

51. Cass Sunstein, Of Artificial Intelligence and Legal Reasoning, 8 U. Chi. L. Sch. Roundtable 29, 33–34 (2001).

52. Christina Mulligan, Perfect Enforcement of Law: When to Limit and When to Use Technology, 14 Rich. J.L. & Tech. 1, 36 (2018).

53. Batya Friedman & Helen Nissenbaum, Bias in Computer Systems, 14 ACM Transactions on Info. Sys. 330, 333–36 (1996) (discussing preexisting bias, technical bias, and emergent bias).

54. Surden, Artificial Intelligence, supra note 3, at 1316.

55. Volokh, supra note 14, at 1171–73.

programs regularly act unexpectedly, sometimes with dangerous results.[56] Judicial programs will also be a tempting target for entities immediately involved in a suit, politically motivated actors, and foreign entities intent on sowing illegitimacy.[57] Granted, human judges may be influenced by threats, bribes, or internalized prejudice,[58] but this is fundamentally distinct from the complete control that may be exercised over a hacked system. Further, while both human and AI decisionmakers may be "gamed," the erroneous decisions of human judges are contained (if just as critical to the parties to a suit),[59] but scaling AI judges entails scaling their vulnerabilities and errors.[60]

Nor will AI judges' errors necessarily be identifiable.[61] Depending on the structure of the algorithm, it may not be possible to plumb the "reasoning" behind a determination to understand where the error occurred.[62] Some forms of AI are simply too complex to disentangle, while the structure of others render them just as opaque to their designers as a human brain. There are a few potential solutions to the explainability problem, including verifying purpose specifications or testing inputs to see which factors affect outputs.[63] However, in addition to technical barriers to transparency, trade secrets law and other legal

---

56. See Joel Lehman, Jeff Clune & Dusan Misevic, The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities 5 (2018), https://arxiv.org/pdf/1803.03453.pdf [https://perma.cc/3RB6-8AJT] ("[A] well-known result in theoretical computer science is that . . . the outcome of a program *cannot* be predicted . . . .").

57. See Volokh, supra note 14, at 1171–72; see also Miles Brundage et al., The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation 16–22 (2018), https://arxiv.org/pdf/1802.07228.pdf [https://perma.cc/QSN2-WJV7] (noting the possible vulnerabilities of AI technologies to external attacks).

58. Volokh, supra note 14, at 1173.

59. But see id. at 1174 (observing that "many of the human judges' prejudices and predictable errors are systemic").

60. Id. (suggesting this makes them an even more appealing target).

61. Id. (noting that we are relatively accustomed to identifying and checking human judge errors). Extrapolating out, to the extent AI judges are employed and their errors are difficult to identify, their decisions may either be nonappealable or foster the development of an entirely new basis for appeals. (Thanks to Alicia Solow-Niederman for this point.) Cf. Mulligan, supra note 52, at 13 (discussing the difficulty of appealing citations from red light cameras, as opposed to those issued by human beings).

62. Selbst, Negligence, supra note 15, at 44–45. Granted, some have argued that there is no need to understand how a program actually reached a decision, provided it produces a sufficiently persuasive account of its reasoning. Volokh, supra note 14, at 1175–76 ("Requiring that source code be published . . . may make it less likely that back doors will be effectively hidden, and may help expose inadvertent bugs as well.").

63. E.g., Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, Accountable Algorithms, 165 U. Pa. L. Rev. 633, 637 (2017) (describing "how technical tools for verifying the correctness of computer systems can be used to ensure that appropriate evidence exists for later oversight"); Volokh, supra note 14, at 1175–76 (discussing whether and to what extent AI judges might be made publicly accessible, to expose back doors, bugs, and emergent bias).

shields currently allow companies to refuse to disclose information needed for these evaluations.[64]

There are real potential benefits to employing AI in some sectors of judicial decisionmaking—but we are unlikely to see generalist, AI-only judges in the near future. Instead, as discussed below, AI is being incrementally integrated into adjudicatory systems, which raises other concerns.

II. THE BEST OF BOTH?

Judging in a common law system is a complex task that requires far more than mindlessly applying rules to fact patterns: "Rather, the decisionmaker is assessing the meaning of the facts and the meaning of the law in the situation"[65] in the context of larger social norms and goals.[66] Yet the discretion required for thoughtful judging imports a host of human biases and attendant unfairness. How to create a legal system that maximizes the best of both human and machine intelligence?

The seemingly obvious solution: Cyborg justice! A legal system designed to take advantage of the strengths of both human beings and algorithms to efficiently increase access to justice with a base level of consistency, while preserving the flexibility to address unusual and extenuating circumstances.[67] In many contexts, AI assistants already free human beings from mundane and time-consuming tasks,[68] permitting us to

---

64. Wexler, supra note 8, at 1349–50.

65. Pasquale, supra note 15, at 29–30.

66. Because "[l]aw as a social institution is multifaceted and embedded in particular political systems and traditions," a technology that "reduces a legal relationship to a 'clear prescription that exists prior to its application and that determines appropriate conduct or legal outcomes'" will be inherently flawed. Id. at 45 (quoting Richard H. Fallon, Jr., "The Rule of Law" as a Concept in Constitutional Discourse, 97 Colum. L. Rev. 1, 14 (1997)); see also Kiel Brennan-Marquez, "Plausible Cause": Explanatory Standards in the Age of Powerful Machines, 70 Vand. L. Rev. 1249, 1298 (2017) ("To say that a problem is best resolved by prudence rather than principle is to express doubt about the possibility of fashioning second-order rules for navigating the collision between first-order values. . . . When that happens, case-specific judgements—as opposed to generalized principles—must carry the day."); Wu, supra note 15, at 2005 ("Better results in hard cases may for a long time still depend instead on accessing something that remains, for now, human—that something variously known as moral reasoning, a sensitivity to evolving norms, or a pragmatic assessment of what works.").

67. See supra note 15.

68. AI assistants help us navigate the internet to answer questions, see, e.g., How Do Search Engines Use Artificial Intelligence?, LMGTFY, https://lmgtfy.com/?q=how+do+search+engines+use+artifical+intelligence%3F&s=g [https://perma.cc/2534-MEY7] (last visited Aug. 26, 2019); act as personal shoppers, see Rory Van Loo, Digital Market Perfection, 117 Mich. L. Rev. 815, 817–22 (2019) (discussing the up- and downsides of the current and imminent proliferation of automated personal shoppers); drive our vehicles, see Nat'l Highway Traffic Safety Admin., U.S. Dep't of Transp., Federal Automated Vehicles Policy 9–10 (2016), https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf [https://perma.cc/WBE6-TCNZ] (discussing current and future levels of vehicle

better concentrate on our priorities and the elements of a decision that we do not wish to delegate.[69] And human–machine teaming has enabled "centaur" chess teams to perform better than human or machine players alone.[70] Similarly, teaming human judges with AI assistants may free human judges to be more thoughtful, just, and humane.

Certainly, the use of AI decision assistants in judicial proceedings is unlikely to raise some of the issues associated with human–machine teaming in other contexts. Unlike piloting an airplane or engaging in active combat, there is little danger associated with unexpected communications breakdowns—the proceeding may be paused until the problem is resolved. And while superhuman reaction times may render human supervisors ineffectual and irrelevant in high-frequency trading[71] or responding to a cyberattack,[72] there is no corresponding need for speed in the judicial context (as evidenced by many a backlog!).

But while hybrid human–AI judicial systems would ideally maximize the strengths of human and machine intelligence, they may magnify the drawbacks of both.[73] They also raise distinct "teaming" risks associated with overtrust, undertrust, and interface design issues, as well as second-order structural side effects.

To effectively participate in a human–machine team, the person in the loop must have an appropriately calibrated amount of trust. If the human being trusts the system too much, to the extent that they endorse an algorithm's conclusion in the face of contradictory evidence or an

---

automation); offer medical advice to patients, Claudia E. Haupt, Artificial Professional Advice, 18 Yale J. Health Pol'y L. & Ethics (forthcoming 2019) (manuscript at 3), https://ssrn.com/abstract=3400898 (on file with the *Columbia Law Review*); and diagnoses to doctors and patients, A. Michael Froomkin, Ian R. Kerr & Joelle Pineau, When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning, 61 Ariz. L. Rev. 33, 39 (2019); W. Nicolson Price II, Medical AI and Contextual Bias, 33 Harv. J.L. & Tech. (forthcoming 2019) (manuscript at 3–4), https://ssrn.com/abstract=3347890 (on file with the *Columbia Law Review*).

69. Cf. Scharre, supra note 15, at 154–55 (discussing the various roles humans play in target selection and engagement—essential operator, moral agent, and fail-safe—and arguing that automated assistants could allow human operators to focus on the latter two).

70. Mike Cassidy, Centaur Chess Brings Out the Best in Humans and Machines, The Bloomreach (Dec. 15, 2014), https://www.bloomreach.com/en/blog/2014/12/centaur-chess-brings-best-humans-machines.html [https://perma.cc/V947-ZWX5].

71. U.S. Commodity Futures Trading Comm'n & SEC, Findings Regarding the Market Events of May 6, 2010, at 1–3 (2010), https://www.sec.gov/news/studies/2010/marketevents-report.pdf [https://perma.cc/P7HC-23B4] (describing the "flash crash," when the Dow Jones Industrial Average lost over one trillion dollars over the course of a half hour).

72. See Rebecca Crootof, International Cybertorts: Expanding State Accountability in Cyberspace, 103 Cornell L. Rev. 565, 580 (2018) ("The speed of cyber will nearly always require that in-the-moment defenses [to cyberattacks] be automated or autonomous.").

73. See Re & Solow-Niederman, supra note 14, at 284–85 (noting that hybrid human–AI judicial systems may result in the "worst of both worlds" if the wrong policy trade-offs are struck—and that market forces, rather than considered or democratic evaluation, are likely to dictate design choices).

obviously unfair result, the human in the loop isn't performing their needed role. Such overtrust—which is related to "automation bias"—risks both overreliance[74] and skill fade.[75] If we wish to elicit the benefits of human reasoning,[76] teaming systems must be designed so that the human in the loop understands the AI program's capabilities and limitations, has reason to exercise valued human skills, and is actively engaged in the decisionmaking process. Absent this, the person may become little more than a figurehead (or scapegoat, should something go wrong).[77] Meanwhile, although undertrust—an unwillingness to accept an algorithm's determinations—can be useful,[78] if the human judges do not trust their AI counterparts, we may overinvest in useless or even harmful infrastructure. One of the aims of incorporating AI into a judicial process is to identify relevant but counterintuitive indicia. But if the human in the loop sometimes employs that information (due to justified trust or overtrust) and sometimes ignores that information (due to undertrust or alert fatigue), that discrepancy may introduce more unpredictability into the decisionmaking process than a human judge acting alone.[79] Finally, in all too many cases, interface design issues have fostered overtrust in the machine system, sometimes resulting in tragedy.[80] Given how much can be lost in translation when human and machine intelligences exchange information, the interface must be designed to both convey information and appropriately calibrate trust.

We learn to trust through experience: Only after repeated engagements do we determine whether and when another entity can be trusted

---

74. See Scharre, supra note 15, at 158 (discussing how human operators of the Patriot air defense system trusted it too much, leading to two 2003 fratricides); P.W. Singer, Robots at War: The New Battlefield, Wilson Q., Winter 2009, at 30, 40 (recounting the role overtrust played in the *USS Vincennes*'s downing of a civilian passenger jet).

75. See Steve Casner, Dumbing It Down in the Cockpit, Slate (Dec. 12, 2014), https://slate.com/technology/2014/12/automation-in-the-cockpit-is-making-pilots-thinking-skills-duller.html [https://perma.cc/8QP2-XSUT] (discussing how pilots' ability to interpret unusual data and to independently keep track of their location worsened after the introduction of autopilots).

76. See Wu, supra note 15, at 2015 (citing 1 William Blackstone, Commentaries *40; Ronald Dworkin, Taking Rights Seriously 23–28, 81–88 (1977)).

77. See Madeleine Clare Elish, Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction, Engaging Sci. Tech. & Soc'y, 2019, at 40, 41–42, https://estsjournal.org/index.php/ests/article/view/260/177 (on file with the *Columbia Law Review*) ("I articulate the idea of a *moral crumple zone* to describe how responsibility for an action may be misattributed to a human actor who had limited control over the behavior of an automated or autonomous system.").

78. See Marc Bennetts, Soviet Officer Who Averted Cold War Nuclear Disaster Dies Aged 77, Guardian (Sept. 18, 2017), https://www.theguardian.com/world/2017/sep/18/soviet-officer-who-averted-cold-war-nuclear-disaster-dies-aged-77 [https://perma.cc/7KM7-D68V] (recounting how Stanislav Petrov likely averted nuclear war when he determined that the reported launch of U.S. missiles was actually a malfunction of the Soviet early warning system).

79. Thanks to Richard Re for this point.

80. See Elish, supra note 77, at 44, 46 (discussing how interface design issues contributed to the Three Mile Island and Air France Flight 447 accidents).

with a particular task.[81] But while experimentation is critical to calibrating trust, "it is difficult to experiment in a space where human liberties are at stake."[82] So while we want hybrid human–AI judicial systems to fail enough to enable the human team member to calibrate trust appropriately and stay sufficiently actively engaged to prevent skill fade, we don't want them to fail in ways that harm litigants or replicate human discretion issues that these systems are intended to correct. This will be a challenging needle to thread.

Hybrid human–AI judicial systems may spur structural changes. We have traditionally celebrated equitable decisionmaking, which values discretion, tailoring, and mercy. However, AI systems naturally favor codified justice, which values measurability, objectivity, and empiricism. The more we incorporate AI systems in the common law process, the more we will encourage a shift in societal values and expectations around adjudication.[83] This may alter the judiciary's composition: If judgment is increasingly understood as rote application of rules to facts, rather than a complex weighing of equitable values and pronouncement of social norms, the "appeal of being a human judge may dwindle," which will affect the composition of the bench.[84]

Further, depending on whether hybrid human–AI judicial systems are embedded in the market or a government decisionmaking structure, their decisions will be monitored differently, which may incentivize counterintuitive or undesirable actions.[85] For example, many market decisionmaking systems, like TurboTax, will be challenged when they undercomply with the law; in contrast, government decisionmaking systems, like one that determines welfare benefits, will be challenged when they overcomply by adopting strict interpretations of requirements.[86] This will incentivize market decisionmaking systems to overcomply and government decisionmaking systems to undercomply with their respective governing rules.[87] Relatedly, if human judges face more scrutiny and critique should they decide against an algorithm's recommendation, they will be consciously and unconsciously incentivized to favor it.

---

81. See Martin & Waldman, supra note 15, at 7–8 (distinguishing between dispositional, situational, and learned trust).

82. Re & Solow-Niederman, supra note 14, at 289.

83. See id. at 259–60 ("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."); id. at 288–89 ("Increasing use of AI adjudication will fundamentally alter practical capabilities, institutional incentives, power relationships, and, ultimately, the views [of the law] of experts and laypersons alike."). Such a shift would alter how society views adjudication and its purposes. Id. (detailing possible implications).

84. Id. at 273–74.

85. See Susan C. Morse, When Robots Make Legal Mistakes, 72 Okla. L. Rev. 213, 230 (2019).

86. Id. at 215.

87. Id. at 216.

Simultaneously, these systems will likely continue the trend of "the 'law' of the firm"—industry lawmaking, monitoring, and enforcement[88]—replacing the "law of the state," as companies use technology to assume roles once reserved to states and state actors.[89] For example, Facebook is creating a "content review board, designed to rule on the hardest of speech-related questions,"[90] institutionalizing a second U.S. system to adjudicate disputes about harmful speech.[91] The board's decisions may create influential precedent that affects the public understanding and social construction of the First Amendment.

The following Part considers another possible side effect of hybrid human–AI judicial systems: the likelihood of legal stagnation, as technological–legal lock-in increases barriers to common law evolution.

III. TECHNOLOGICAL–LEGAL LOCK-IN

The common law has long been celebrated for combining stability with the ability to adapt to social shifts and new technologies.[92] It evolves

---

88. Margaret Jane Radin, Regulation by Contract, Regulation by Machine, 160 J. Institutional & Theoretical Econ. 142, 143 (2004) [hereinafter Radin, Regulation by Contract]; see also Rebecca Crootof, The Internet of Torts, 69 Duke L.J. (forthcoming 2019) [hereinafter Crootof, Internet of Torts] (manuscript at 3–10), https://ssrn.com/abstract=3342499 (on file with the *Columbia Law Review*) (discussing how terms of service and Internet of Things devices grant companies a newfound power to create governance rules, monitor compliance, and enforce consequences for violations).

89. See, e.g., Kate Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, 131 Harv. L. Rev. 1598, 1601–03 (2018) (highlighting how the content moderation policies of online platforms are shaping what speech is allowed in the public sphere); K. Sabeel Rahman, The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept, 39 Cardozo L. Rev. 1621, 1625–27 (2018) (arguing that, as with early utilities, platforms are entities with coercive powers that are not subject to democratic constraints); Van Loo, The Corporation as Courthouse, supra note 5, at 549–54 (discussing how companies' internal dispute processes are replacing arbitration and trials). For a vision of how industry and states might work together to govern the algorithms, see Kaminski, supra note 13 (manuscript at 108).

90. Wu, supra note 15, at 2004–05; see also Anupam Chander, A Facebook Supreme Court?, Balkinization (May 31, 2019), https://balkin.blogspot.com/2019/05/a-facebook-supreme-court.html [https://perma.cc/3BE3-97Q9] (outlining alternatives to Facebook's content review board and concluding that "a Facebook Supreme Court is the worst idea, except for all the others").

91. See Thomas E. Kadri & Kate Klonick, Facebook v. *Sullivan*: Public Figures and Newsworthiness in Online Speech, 93 S. Cal. L. Rev. (forthcoming 2019) (manuscript at 23), https://ssrn.com/abstract=3332530 (on file with the *Columbia Law Review*) (discussing two governance systems: a public one for torts in courts and a private one for content moderation on online platforms); Kate Klonick, Facebook v. *Sullivan*, Knight First Amend. Inst. (Oct. 1, 2018), https://knightcolumbia.org/content/facebook-v-sullivan [https://perma.cc/GML5-GU92] (noting that "the United States now has two systems to adjudicate disputes arising from harmful speech about other people" and discussing what actors in each system can learn from the other).

92. See, e.g., Comment, The Dead Hand of the Common Law, 27 Yale L.J. 668, 672 (1918).

gradually and organically, responding to new problems first with analogies and then with revised rules. Precedential cases gain strength by virtue of being built upon: While none of the cases in a particular chain may be weighty enough to be memorialized in a casebook, incremental changes in interpretations and applications can dramatically shift or fundamentally set the path of the law.[93] Meanwhile, human judges are regularly "updating."[94] Individuals are "expose[d] to new information and experiences," while "turnover from career changes, retirement, and mortality mean that the bench is in a constant state of generational flux."[95]

Both legal stability and legal evolution are critical to law's legitimacy. Predictable decisions comporting with prior ones creates the legitimacy associated with time-tested rules and the fairness of like cases being treated similarly; flexible applications of those rules are necessary to accommodate changing social norms and extenuating circumstances in particular cases.

But hybrid human–AI judicial systems risk fostering stability at the expense of evolution.[96] As laws, norms, policies, and decision processes are translated into algorithmic programs, they are cemented. Although regulation by code has been celebrated for its flexibility,[97] the technological, economic, and political edifices within which code is entrenched may be far more rigid. Path dependence, bureaucratic inertia, and special interest lobbying by concentrated groups may make it just as difficult to change algorithms implementing law as it already is to change written law, creating a kind of "technological deep state" that resists a new administration's desired policy changes.

Further, regulation by code is arguably harder to change than precedent or statutes because—like other forms of architectural regulation—it constructs and constricts our options while remaining hidden.[98] We are aware that laws are passed or modified through interpretation, that social norms change from community to community, and that prices are set. Architectural regulation is far more insidious, as it can cloak nudges and

93. Cf. Brian Soucek, Copy-Paste Precedent, 13 J. App. Prac. & Process 153, 153 (2013) (discussing how nonprecedential decisions act as precedent and, over time, can magnify ambiguities and mistakes).

94. Re & Solow-Niederman, supra note 14, at 268–69.

95. Id.

96. These concerns are equally—if not more—applicable to AI-based adjudicatory systems that do not have a human in the loop. But, for reasons articulated above, we are far more likely to have hybrid human–AI judicial systems than AI judges in the near future.

97. See Lawrence Lessig, Code: Version 2.0, at 139 (2006).

98. See id. at 136–37. Of course, architectural regulation can be made obvious, as anyone who has faced a password request or had speedbumps installed on a familiar route can attest.

mandates with the legitimacy of apparent self-determination.[99] After 1948, for example, communities built highways, laid railroad tracks, and created zoning constraints to fortify segregation, enabling "what would clearly be an illegal and controversial regulation without even having to admit any regulation exists."[100] While the segregating effects of these architectural decisions are still felt today, it is easy to forget that they were constructed. Instead, "[t]he continuing segregation of these communities is described as the product of 'choice.'"[101] Similarly, digital code directs and limits what we can do online[102]—and, as more and more items are added to the "Internet of Things," code increasingly determines what we can do in the physical world.[103]

Nor will the AI programs in the system automatically update over time.[104] Pattern-recognition AI systems are trained on past data to predict future success—but while they may be excellent at identifying traits that were once relevant in achieving a goal, AI programs cannot easily adapt what once worked to new situations.[105] Instead, they reflect back prior human bias, as when Amazon's CV-reviewing program learned to penalize women candidates based on initial training data about who had traditionally been successful in the company.[106] To minimize this issue, the AI

---

99. See id. at 136 (noting that, by regulating indirectly through code, governments "can achieve regulatory ends, often without suffering the political consequences that the same ends, pursued directly, would yield").

100. Id. at 135. To the extent architectural regulation obscures accountability, it impedes public knowledge, granting additional information advantages to interest groups. Cf. BJ Ard, The Limits of Industry-Specific Privacy Law, 51 Idaho L. Rev. 607, 615 (2015) (explaining why public choice theory predicts concentrated groups will often prevail over the general public in "organizing around rapidly changing, technically dense, and largely invisible issues").

101. Lessig, supra note 97, at 136.

102. E.g., Julie E. Cohen, Cyberspace as/and Space, 107 Colum. L. Rev. 210, 224–25 (2007) (arguing that the design of online environments constricts users' behavior); Steve Woolgar, Configuring the User: The Case of Usability Trials, *in* A Sociology of Monsters: Essays on Power, Technology, and Domination 59, 69 (John Law ed., 1991) (noting how technological designs limit users' activities).

103. See, e.g., Aaron Perzanowski & Jason Schultz, The End of Ownership: Personal Property in the Digital Economy 140–41 (2016) (noting the increased functionality of, and reliance on, smart devices); Crootof, Internet of Torts, supra note 88 (manuscript at 23–24) (underscoring the physical nature and physical consequences of Internet of Things–enabled corporate remote interference).

104. See Re & Solow-Niederman, supra note 14, at 269; see also Volokh, supra note 14, at 1187–88 (proposing that AI judges be replaced every three years to accommodate changing attitudes about legal values).

105. See Re & Solow-Niederman, supra note 14, at 245 n.9; cf. Rise of the Machines, Saturday Morning Breakfast Cereal, https://www.smbc-comics.com/comic/rise-of-the-machines [https://perma.cc/K7JH-E6V9] (last visited Aug. 26, 2019).

106. Jeffrey Dastin, Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women, Reuters (Oct. 9, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/6UTX-RJDZ]; see also Re & Solow-Niederman, supra

in the system will need to be regularly retrained on updated data, but absent enforced requirements for frequent updates and some form of planned obsolescence—as well as a dedicated budget line for both—hybrid human–AI judicial systems are likely to remain consistently behind the times.

But won't the human in the loop correct for this issue? Of course, human beings will continue to "update" over time, both as individuals and collectively.[107] But one of the aims of hybrid human–AI judicial systems was to minimize human discretion, save for outlier cases. However, social shifts will render some apparently unfair decisions common and systemic, rather than aberrations or exceptional cases justifying human override.[108] Alternatively, thanks to the insidious nature of architectural regulation, judges and others who interact with AI decision assistants in the adjudicatory process may come to accept their biases, nudges, and limitations as the natural, and possibly only, way of structuring the process.[109] Additionally, some combination of overtrust, skill fade, and shifts in who becomes a judge and what kind of justice is valued[110] may affect how likely a human being is to question an AI's recommendation or call for an override, even when the program's determination is clearly problematic.

Granted, there are potential benefits that accompany technological–legal lock-in. Just as Odysseus's choice to be lashed to the mast was a manifestation of his autonomy, and just as states' acceptance of the obligations and responsibilities detailed in difficult-to-modify constitutions and treaties is an expression of their sovereignty, societies can decide to elevate certain rules or avoid erosion of certain norms through

---

note 15, at 273 & n.108 ("AI adjudication, in other words, could cast a bright light on human adjudication—and people might not like what they see.").

107. Re & Solow-Niederman, supra note 14, at 249–50.

108. Cf. Deborah J. Vagins & Jesselyn McCurdy, ACLU, Cracks in the System: 20 Years of the Unjust Federal Crack Cocaine Law 2 (2006), https://www.aclu.org/sites/default/files/field_document/cracksinsystem_20061025.pdf [https://perma.cc/YJA7-8PR4] (discussing the example of how cocaine possession statutory sentences differed dramatically, depending on whether it was sold in solid or powder form).

109. While this Piece is primarily concerned with the role of the human judge in the hybrid human–AI judicial system, the design of digital decisionmaking environments will also influence litigants, particularly self-represented parties. Ayelet Sela, e-Nudging Justice: The Role of Digital Choice Architecture in Online Courts, 2 J. Disp. Resol. 127, 128–29 (2019) (arguing that, "[w]hether purposefully or inadvertently, the design of digital environments often steers their users' behavior" in ways that are nontransparent and sometimes manipulative, requiring online court designers to take care that the architecture comport with the core values of the judicial system).

110. Re & Solow-Niederman, supra note 14, at 274 ("[T]he ultimate effect [of hybrid human–AI judicial systems] 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.").

legal codification.[111] Further, thoughtful limitation can sometimes result in more generativity: Poets can find that the constraints of the sonnet, limerick, or haiku forms free them to be more creative in other ways, while programming or application advances may depend upon locked-in standards or sub-levels.[112] If the design allows for it, judges may be incentivized to "innovate" within the structures necessitated by human–machine teaming.

But there are also obvious problems. Who decides what gets locked in, what are their biases, and what are they incentivized to prioritize? Today, all too often, profit-motivated firms are making these determinations,[113] raising all of the issues associated with prioritizing the law of the firm, which aims to profit the firm, over the law of the state, which reflects varied interests and presumably exists to benefit the general public.[114] Even if these decisions are instead vested with democratically-elected legislative bodies, and assuming that technologists can adequately translate desired policy into digital code, the legislature of the moment will be privileged at the expense of future polities. And regardless of which entity is making these determinations, technological–legal lock-in will entrench today's biases, resulting in legal decisions that are increasingly out of step with future social norms.

CONCLUSION

Fully aware of the importance of legal evolution, Thomas Jefferson wrote:

> I am certainly not an advocate for frequent and untried changes in laws and constitutions. . . . But I know also, that laws and institutions must go hand in hand with the progress of the human mind. As that becomes more developed, more enlightened, as new discoveries are made, new truths disclosed, and manners and opinions change with the change of circumstances, institutions must advance also, and keep pace with the times. We might as well require a man to wear still the coat which fitted him when a boy, as civilized society to remain ever under the regimen of their barbarous ancestors.[115]

---

111. Id. at 288 (suggesting that "[e]quity-preserving code" could "serv[e] as a bulwark against the erosion of equitable justice that would take place in a totally open 'market for justice'").

112. Jonathan Zittrain, The Future of the Internet and How to Stop It 63–64 (2008).

113. In many cases, industry is developing automated complaint-resolution mechanisms to serve its own purposes. See, e.g., Van Loo, The Corporation as Courthouse, supra note 5, at 564–65. AI-adjudication developers also have incentives that may be at odds with traditional judicial values. Re & Solow-Niederman, supra note 14, at 272.

114. Radin, Regulation by Contract, supra note 88, at 147.

115. Letter from Thomas Jefferson to "Henry Tompkinson" (Samuel Kercheval) (July 12, 1816), https://founders.archives.gov/documents/Jefferson/03-10-02-0128-0002 [https://perma.cc/P272-VXTD]; see also Jack M. Balkin, Living Originalism 3 (2011) (arguing that

Given their sensitivity to changing social norms,[116] human judges are uniquely able to oversee legal evolution and ensure that our judicial system "keep[s] pace with the times."[117] Accordingly, to the extent they are employed, hybrid human–AI judicial systems must be designed to ensure opportunities for human judges to engage in the value balancing and norm incorporation necessary to maintain an evolving and legitimate common law.

---

the U.S. Constitution is best understood as an incomplete framework for government, designed to be built out over time).

116. Cf. Ronald Dworkin, Law's Empire 350 (1986).

117. See Brennan-Marquez, supra note 66, at 1256 (discussing, in the context of criminal justice, how judges are able "to consider the plurality of values implicated by the exercise of state power" and "to resolve conflicts between those values in a context-sensitive way"); id. ("[T]he rationale for individualized review, costly and inefficient as it may be, is that in some settings *we cannot be sure in advance which values will be implicated by the exercise of power*. And when that is true, decisionmaking resists automation."); see also Volokh, supra note 14, at 1183 (noting that, to the extent common law development requires prediction, we may never have Supreme Court or appellate-level AI judges).