

Northwestern Pritzker School of Law

Northwestern Pritzker School of Law Scholarly Commons

JCLC Online

Fall 12-22-2023

Role-Reversible Judgments and Related Democratic Objections to AI Judges

Amin Ebrahimi Afrouzi

Follow this and additional works at: https://scholarlycommons.law.northwestern.edu/jclc_online

ROLE-REVERSIBLE JUDGMENTS AND RELATED DEMOCRATIC OBJECTIONS TO AI JUDGES

AMIN EBRAHIMI AFROUZI*

In a recent article published by this journal, Kiel Brennan-Marquez and Stephen E. Henderson argue that replacing human judges with AI would violate the role-reversibility ideal of democratic governance. Unlike human judges, they argue, AI judges are not reciprocally vulnerable to the process and effects of their own decisions. I argue that role-reversibility, though a formal ideal of democratic governance, is in the service of substantive ends that may be independently achieved under AI judges. Thus, although role-reversibility is necessary for democratic governance when human judges are on the job, it may not be so when AI judges replace them. One broader implication for normative evaluation of disruptive technologies follows: formal and substantive ideals that often align must be independently examined in the evaluation of disruptive technologies. This is because these formal and substantive ideals may no longer align under the factual circumstances that come to govern when such technologies are deployed.

INTRODUCTION

Countless articles, academic and otherwise, portend the arrival of AI judges. Some authors even believe that human judges can be entirely

* Many thanks to Christopher Kutz, Stephen Darwall, Christine Chwaszcza, Weyma Lübke, Simone Zurbuchen, Christian Maurer, and Matthew Braham, Rafael B. Nunes, and Pinchas Huberman, who offered written feedback on earlier versions. And to Jonathan Wolff, Grant Lamond, Julian F. Müller, Jeffrey Kaplan, William Darwall, Agnes Harriet Lindberg, Artur Pericles Lima Monteiro, Elisabeth Paar, Aaron Mendon-Plasek, Marcela Mattiuzzo, and Mila Samdub for conversations about these issues. And to Markus Rabe and Ali Afrouzi for technological expertise and invaluable feedback on earlier drafts. And to audiences at the Information Society Project Workshop at Yale Law School and the Cambridge Legal Theory Discussion Group, where I presented some of these ideas. Finally, I am deeply grateful to the editors of JCLC, especially to Sierra Anderson and Nia Crosley, for insightful substantive and editorial suggestions. All faults remain mine.

replaced by AI judges.¹ But would such a future be justifiable? Much of the literature on AI ethics focuses on the imperfections in existing technology. With technological progress, however, presumably in due course, technologically ideal AI judges can be developed and the ethical concerns arising from their putative technological shortcomings will be outdated. In a recent article published by this journal, Kiel Brennan-Marquez and Stephen E. Henderson make a stronger case against AI judges.² They argue that replacing human judges with AI would violate the role-reversibility ideal of democratic governance. The ideal of role-reversibility requires that “those who exercise judgment should be reciprocally vulnerable to its processes and effects.”³ Not only do human judges ultimately also occupy a space in society as private citizens and in that capacity are themselves subject to the precedents they leave behind, but the roles between those judging and those judged could have been reversed but for a contingent series of events. Unlike human judges, AI judges are not so situated. Hence, Brennan-Marquez and Henderson conclude, deploying AI judges would violate ideals of democratic governance.

The argument of Brennan-Marquez and Henderson cuts deep because it even applies to technologically ideal AI judges. Futurologists like Eugene Volokh maintain that once such technologies are developed, there would be nothing wrong in delegating most, if not all, of the tasks of human judges to AI judges.⁴ Brennan-Marquez and Henderson’s critique dampens such optimism by focusing on constraints that democratic governance imposes on the structure of courts, which the reign of AI judges threatens. Such threats persist, even after purportedly evanescent technological shortcomings were addressed.

Therefore, there is much to admire about and learn from Brennan-Marquez and Henderson’s critique. I am skeptical, however, about whether the violation of role-reversibility will have enough normative weight to support a democratic objection to AI judges. This is not to say that AI judges are immune to critique. Indeed, I share Brennan-Marquez and Henderson’s intuitions against AI judges. What I am skeptical about is whether it is AI

¹ Eugene Volokh, *Chief Justice Robots*, 68 DUKE L. J. 1135, 1191 (2019), <https://perma.cc/7ATF-4ZEC>.

² Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-reversible Judgement*, 109 J. CRIM. L. & CRIMINOLOGY 137 (2019) [hereinafter Brennan-Marquez & Henderson (2019)].

³ Brennan-Marquez & Henderson, *supra* note 2, at 149.

⁴ Volokh, *supra* note 1, at 1135–92 (“[W]hen AI judges become highly effective at crafting persuasive legal arguments, there will be little reason to prefer human judges to AI judges, at least for the overwhelming majority of legal questions, including the law development questions that reach the Supreme Court.”).

judges' violation of *role-reversibility* that justifies our shared intuitions. This is because, as far as I can see, the violation of role-reversibility will have very little or no normative weight under some factual conditions that could come to govern when AIs rather than humans act as judges.

Brennan-Marquez and Henderson maintain that role-reversibility is of “intrinsic value.”⁵ But this idea can be fleshed out further. The ideal of role-reversibility expresses a formal constraint, namely that the role of the judging and the judged parties should be reversible in principle (i.e., but for a contingent series of events). This formal constraint often serves two substantive ends. First, it promotes evenhandedness and public justifiability of the content and application of the law. Second, it promotes equality of rank between those judging and those judged, thereby minimizing hierarchies and relations of domination between citizens, irrespective of social roles they happen to occupy. Now the violation of role-reversibility can be interpreted either in purely formal terms or in terms of the frustration of the substantive ends it serves. Given their discussion, I think Brennan-Marquez and Henderson are concerned about both. In response, I argue that as merely a formal matter, the violation of role-reversibility is too thin to do the normative work they want. Meanwhile, its substantive ends need not be frustrated under the reign of AI judges. Thus, although the formal condition of role-reversibility is a necessary criterion of democratic governance when human judges are on the job, it may not be when AI judges replace them.

The aim of this paper is not exclusively critical, however. Rather, my critical discussion is meant to highlight certain valuable lessons that we can learn by paying due attention to Brennan-Marquez and Henderson's subtle critique. If successful, what the following dialectic (among other things) demonstrates is that when it comes to normatively evaluating disruptive technologies, we must explain exactly what substantive ends may be frustrated if existing formal criteria are violated. For instance, we must show that violated formal criteria that track substantive ends today continue to do so even under conditions that come to be as a result of deploying the technology in question. Because formal and substantive ideals may no longer align as a result of deploying the technology in question, substantive conclusions could not rest on an examination of compliance with formal criteria alone.

My discussion proceeds as follows: in section I, I reproduce Brennan-Marquez and Henderson's argument in a nutshell. In section II, I offer a very brief bird's eye view of the existing literature on AI judges to highlight the brilliance of Brennan-Marquez and Henderson's article. In section III, I

⁵ Brennan-Marquez & Henderson, *supra* note 2, at 7.

critique Brennan-Marquez and Henderson's argument along the lines suggested above. In section IV, I conclude with sketching the broader theoretical implications of my discussion.

I. BRENNAN-MARQUEZ AND HENDERSON'S ARGUMENT

Brennan-Marquez and Henderson argue that the principle of "role-reversibility" is a "facet of democratic legitimacy" and "the normative basis...of the Anglo-American 'jury of peers' ideal."⁶ This principle maintains that "those who exercise judgment should be vulnerable, in reverse, to its processes and effects. And those subject to its effects should be capable, reciprocally, of exercising judgment."⁷ In line with this principle, Brennan-Marquez and Henderson argue that "in a liberal democracy," there must be a sense in which "the participants' roles in the process could always be inverted."⁸ Delegating the tasks of human judges to AI is impermissible, in their view, because it would flaunt role-reversibility, hence undermining the democratic legitimacy of subsequent decisions.

The kernel of Brennan-Marquez and Henderson's argument is a distinction between when decisions about the adoption and application of rules are made from within the moral community and from outside of it. To this end, they emphasize two substantive ends that adherence to role-reversibility serves: First, it promotes the evenhandedness and public justifiability of the content and application of legal rules. Second, it promotes equality of rank between those applying the rules and those subject to them. It is the promotion of these substantive ends that renders decisions made under the conditions of role-reversibility *democratic* acts, even when resultant decisions are no better in content.

Once this much is established, Brennan-Marquez and Henderson proceed to the final step of their argument: even supposing that AI judges produce the same output as human judges, their decisions could not be democratic, because AI judges are not reciprocally vulnerable to the judicial process and its effects. This difference matters because human judges could in principle occupy the position of those who are judged. Not only do human judges occupy the positions of ordinary citizens and are subject to the precedents they leave behind, they could have been in the position of the person they are now judging but for a contingent series of events.

In contrast to human judges, AI judges could not even in principle occupy the position of those they judge. So, Brennan-Marquez and

⁶ Brennan-Marquez & Henderson, *supra* note 2, at 142, 163.

⁷ *Id.* at 140

⁸ *Id.*

Henderson conclude, AI decisions would be undemocratic, even if extensionally identical to those made by human judges. In short, when the task is delegated to AI judges, “the rules and outcomes may be functionally identical, but the acts of deciding would be different. They would not be democratic acts.”⁹ Or so they argue.

II. THE ROLE-REVERSIBILITY CRITIQUE: BRILLIANCE

Brennan-Marquez and Henderson’s critique of AI judges is laudable because it cuts really deep. Importantly, the vast majority of critical assessments of AI judges in the existing literature rely on some difference, albeit small, between the work products of humans and AI judges. But futurologists can respond to such critical assessments by promising that, given enough technological advancement, the relevant differences will dissipate.

For instance, Tania Sourdin and Richard Cornes, who are by many accounts among the strongest authorities in the field, have argued that AI judges should not be used because, lacking empathy, they could not be as responsive to the emotions underlying the matters that come before them. I think they are right. But futurologists can object that their stance rests on the presupposition that AI does not perform as well as humans, which may not be true after sufficient advancement in AI technology. Sourdin and Cornes write that “the role of the human judge... is not merely that of a data processor,” adding that “to reduce judging to [what a data processor could do] would be to reject not only the humanity of the judge, but also that of all those who come before them.”¹⁰ This is certainly true today and perhaps continues to be true in the indefinite future. But it is not conceptually true, because conceptually, it may be possible for AI technology to produce judicial decisions that are indistinguishable from those written by humans. Yet, the assumption that animates the argument of Sourdin and Cornes against AI judges is that their work products will be of a lesser quality. Though forceful against today’s AI judges, their critique may not apply to AI judges of some future whose work products are substantively identical to the work products of human judges.

Joshua Davis similarly advocates against using AI judges by arguing that sometimes judges look to the law as a source of moral instruction, which requires the ability to reason morally. But given that AI cannot (yet) engage

⁹ Brennan-Marquez & Henderson, *supra* note 2, at 141.

¹⁰ Tania Sourdin & Richard Cornes, *Do Judges Need to Be Human? The Implications of Technology for Responsive Judging*, in *THE RESPONSIVE JUDGE: INTERNATIONAL PERSPECTIVES* 87, 87 (Tania Sourdin & Archie Zariski, eds. Spring Press 2018).

in moral reasoning, his argument goes, AI could not fulfill some such functions that fall within the responsibilities of human judges.¹¹ I tend to think that Davis is also right. But futurologists may still press that this critique too does not cut deep enough, given that Davis assumes that a disparity between machine and human reasoning causes the machine to yield work products of lesser quality.¹²

Brennan-Marquez and Henderson's critique, on the other hand, cuts much deeper because it does not rest on any *dissimilarity* between the work products of AI and human judges. Thus, the assumption that in some future, the holdings of AI judges will be extensionally identical to the holdings of human judges, could not even mitigate their critique, let alone resolve it. Their critique rests instead on the democratic ideal of role-reversibility. It rests on the purported violation of a democratic constraint on the structure of courts that arises whenever human judges are replaced with AIs, irrespective of how well they perform. As they note:

one appealing feature of our account—even putting all other benefits to one side—is that it resists the ‘humanity-fetishism’ (or ‘speciesism’) that often looms over conversations about humans and machines. There is nothing special, in our view, about a human decision-maker. Rather, what matters is whether the decision-maker could swap positions with the affected party.¹³

The critique cuts deep because it bites AI regardless of how well it performs compared to humans.¹⁴ It does so because it applies equally to human judges who are not part of the moral community. AI judges are therefore not subject to the critique because of some disability assumed about them. They are subject to the critique because of the conceptual role they would and would not occupy in the society. This critique is especially important given that optimism about technologically ideal AI could even

¹¹Joshua P. Davis, *Law Without Mind: AI, Ethics, and Jurisprudence*, 55 CAL. W. L. R. 165, 172 (2019).

¹² Proponents of AI judges can also readily concede that the technology should only be used when it passes a certain threshold of quality in its outcome. See e.g., Volokh, *supra* note 1, at 1191 (limiting his endorsement of AI judging to “*when AI judges become highly effective at crafting persuasive legal arguments*”) (emphasis added).

¹³ Brennan-Marquez & Henderson, *supra* note 2, at 142. See also *id.* at 149 (“What matters, then, is not the fact of humanness per se. What matters is whether decision-makers are situated to imagine themselves into the role of an affected party, and vice versa—such that both participants, and in some sense the entire moral community, can understand judgment as a democratic act.”).

¹⁴ On whether the arising shortcoming could be compensated by, say, extremely better performance of AI, see Amin Ebrahimi Afrouzi, *AI Decisions: Faster, Cheaper, and more Accurate, yet not Good Enough?* [hereinafter *AI Decisions*, manuscript forthcoming, on file with author].

pave the path of using *imperfect* AI technology with the promise that its flaws resolve themselves while in use.

III. THE ROLE-REVERSIBILITY CRITIQUE: LIMITATIONS

The contribution of Brennan-Marquez and Henderson's argument should not be underestimated. In applying to technologically ideal AI judges with equal force, their argument preempts much debate in the literature. It also pioneers the critical assessment of AI technologies from a higher level of abstraction than what may have been thought possible before.¹⁵ There is much to admire about and learn from their critique, therefore, both in substance and methodology. Nevertheless, there is a limitation to the normative upshot of their critique as it applies to a technologically ideal future. Importantly, the normative force of their critique seems to dissipate if we assume that, in such a future, AI judges decide in a way that is substantively identical to (or on a par with) decisions made by human judges who would be operating under the conditions of role-reversibility.

Consider first the substantive ends role-reversibility serves. As discussed above, having a judge that is of the people and vulnerable to the process and effects of her own decision on the one hand promotes evenhandedness and public justifiability of the law's content and its application. On the other hand, it promotes the equality in the rank of citizens, thereby diminishing hierarchies and relations of subordination between those judging and judged.

The justifiability of the content is part of what concerns Brennan-Marquez and Henderson when, for instance, they draw on Rawls and Scanlon to argue that the conditions of role-reversibility is likely successful in promoting the adoption of impartial rules and their impartial application.¹⁶ But Brennan-Marquez and Henderson also maintain that role-reversibility is of "intrinsic value."¹⁷ What they mean by this seems to be that decisions made from within the moral community rather than from outside of it promote equality of rank among citizens.¹⁸ To this end, they argue that decisions made by peers and under the conditions of role-reversibility will minimize hierarchies and relations of domination. This renders those

¹⁵ *But see* Ian Kerr & Carissima Mathen, *Chief Justice John Roberts is a Robot* 8, (U. Ottawa Working Paper, 2014), <https://perma.cc/7ZQE-EY7S>, which is to my knowledge the first article to argue that even technologically perfect AI would be disqualified to serve as judge.

¹⁶ Brennan-Marquez & Henderson, *supra* note 2, at 149–52.

¹⁷ *Id.* at 142, 163.

¹⁸ *Id.* at 152–56.

decisions democratic, even if they prove to be no better in content. Brennan-Marquez and Henderson write:

even if they apply the same criteria and reach the same outcomes, it is fundamentally different for a king or queen, standing above the law, to cast judgment on one of their subjects, or for the high-born, in a caste system, to decide the fate of the low-born. And for the same basic reason, it would be fundamentally different for a machine to have the ultimate say over decisions. The rules and outcomes may be functionally identical, but the acts of deciding would be different. They would not be democratic acts.¹⁹

The idea here is that decisions made by a judge foreign to the community under her rule could subject that moral community to her arbitrary will. Even if this foreign judge made her decisions on the basis of the values of the community subject to her rule, she would still occupy an elevated role compared to them. Role-reversibility ensures against this sort of structural hierarchy.

Suppose (for now) that promoting the two substantive ends just discussed exhaust the value of role-reversibility. It would then be natural to think that, if these substantive ends could be guaranteed by other means, we would have no reason to additionally fulfill role-reversibility for its own sake. My contention is that all such substantive ends could indeed be fulfilled in a future with sufficiently sophisticated AI technology. Accordingly, the violation of role-reversibility will have no normative upshot.

Brennan-Marquez and Henderson overlook this fact because they assume a particular technological embodiment of an AI judge, one that places it *a priori* outside of the moral community. In their assumed hypothetical, judging is delegated to *artificial general intelligence* or AGI, which presumably has a mind of its own in deciding the fates of those before the court.²⁰ But an ideal AI judge need not have a mind of its own, nor, god forbid, have the capacity of becoming our “machine overlord” having “more to say about our systems of criminal justice than we had planned.”²¹ Suppose instead that the AI judge of our futuristic hypothetical simply predicts what a human judge would do had she been presiding over the case. Existing AI technology does just this, though inaccurately.²² But suppose that in some future, AI judges do so accurately. Now all that is needed for the AI’s work

¹⁹ *Id.* at 141.

²⁰ *Id.* at 144–45.

²¹ *Id.* at 145.

²² See, e.g., Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiuc-Pietro & Vasileios Lampos, *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, 2 PEER J. COMPUT. SCI., Oct. 24, 2016; Ryan W. Copus, *Statistical Precedent: Allocating Judicial Attention*, 73 VAND. L. R. 605 (2020).

product to fulfill the substantive ends served by role-reversibility is that the *human judge it mimics* be a member of the society it serves.²³

Under these circumstances, the content of law will be just as even-handed and as publicly justifiable as it would have been had the human judge decided the case herself. Nor would this institute any hierarchy or relation of domination among citizens, for remember that the human judge who the AI mimics is but herself a member of society, and in her capacity as an ordinary citizen is equally vulnerable to the process and its effects. In short, under these circumstances, it is not important for the AI to be subject to its own decisions, so long as its decisions mirror decisions of human judges who are subject to their own decisions. For then, all the substantive ends that role-reversibility serves will still be fulfilled.

This analysis absolves AI judges if role-reversibility was of value only for the substantive ends it serves. But perhaps Brennan-Marquez and Henderson mean to claim something stronger when they write that role-reversibility is of intrinsic value.²⁴ Perhaps they mean to claim that even when all the substantive democratic ends that role-reversibility currently fulfills could be guaranteed by other means, we would still have reason to strive for role-reversibility for its own sake. That would be to claim that promoting the sort of substantive ends discussed above does *not* exhaust the value of role-reversibility. I doubt that this claim holds water. I am also not sure that this is indeed what Brennan-Marquez and Henderson mean to claim. But let me address this possibility lest some readers wonder about it.

Consider for instance the idea that male judges should not be able to decide cases about abortion law because they would never be vulnerable to their decisions. Though this line of thought mirrors the role-reversibility discussion of our concern, it does not have universal uptake among jurists or legal theorists. What's important is not that this idea is controversial but that even its proponents would likely concede that it would be appropriate for male judges to decide such cases *if* they had the ability to decide them in a way indistinguishable from how their female counterparts would have done so, had they been presiding. Disagreements here, I think, do not turn on formal reciprocity but on whether male judges indeed have such an ability.

The example shows that a purely formal requirement of role-reversibility proves too much, which is to say that it not only speaks against AI judges but against much of our existing practices that we regard as

²³ An AI could do what an actual human judge who presides over similar cases would do or it could do what an average human judge in the abstract would do. In either case, the AI's work product would fulfill the substantive ends served by role-reversibility. I thank Grant Lamond for pressing me to clarify this point.

²⁴ Brennan-Marquez & Henderson, *supra* note 2, at 142, 163.

unproblematic. This makes me doubt whether we have reason to pursue role-reversibility as a purely formal matter and for its own sake.

It is also worth pointing out that even as a purely formal matter, role-reversibility may be more closely met under AI judges, so long as the human judges whom they mimic are themselves part of the community. Consider once again that a male human judge likely would not even in principle be subject to the abortion decisions he makes,²⁵ even if he really *did* have the ability to decide cases in the exact same way as those affected by his decisions. AI judges on the other hand, can mimic judges who would themselves be affected by the decisions. Here, only a human judge who successfully tried to mimic his female colleagues had they been presiding would reach the threshold of role-reversibility in abortion cases that an AI judge could reach. Yet, this is not what human judges ordinarily do or are particularly good at.

Thus, so long as AI judges decided not based on their own best discretion but on the basis of what they predict human judges would do if presiding, even a purely formal notion of role-reversibility will be in some sense satisfied. This is especially so, given that there is no reason to think that in a future where AI judges are deployed, humans would be barred from becoming judges. The position of the judge can remain open to all in a full democratic fashion, though it may have no human aspirants.

One could insist, of course, that still *something* would be lost, even in decisions by human judges when role-reversibility is violated even in a purely formal sense. This may seem more dooming for AI judges because, in their case, formal role-reversibility would be necessarily violated, not just in abortion-type cases but across the board. After all, it is the AI rather than the judge it mimics who *actually* hands down the decisions. I concede this point. My contention is that this “something” could not be any more than a ritual—a ritual that, as far as I can see, has little or no *normative* significance in that future world where human-mimicking-AI is on the job.²⁶

To summarize, then, much hangs on what role-reversibility amounts to.²⁷ I have offered three interpretations of it: (1) as an instrument for securing even-handed and publicly justifiable rules, (2) as an instrument to secure equality of rank among citizens, or (3) as a purely formal ritual worth pursuing for its own sake. The first and second interpretations are motivated

²⁵ Except, of course, being born male is a contingent matter, a possibility that I could not dismiss but in the service of simplifying our current discussion.

²⁶ Thanks to Rafael B. Nunes for pressing me address the ritualistic violation of role-reversibility as a formal matter.

²⁷ As Aaron Mendon-Plasek put it to me in conversation, Brennan-Marquez and Henderson’s and my discussion is as much about role-reversibility as it is about AI.

by Brennan-Marquez and Henderson's own discussion. On these interpretations, AI judges prove unproblematic. The last and purely formal interpretation of role-reversibility, though possible, is less plausible. Even so, this last interpretation seems to me as either hollow or too thin to underpin a democratic objection against AI judges, at least not of the kind that mimics human judges. None of this is meant to suggest that there could not be other reasons against AI judges, but only that role-reversibility does not provide such a reason.²⁸

Before closing, one point needs emphasis, namely that role-reversibility succeeds in putting a constraint on AGI or any trans-human AI to act as judges.²⁹ As discussed above, and as Brennan-Marquez and Henderson correctly see, judicial decisions made from outside the moral community would not be democratic. This in effect constrains how much accuracy can be purchased with AI. Suppose then, as Brennan-Marquez and Henderson (following Max Tegmark) do, that decisions made by AGI acting as judge would be "more accurate and consistent."³⁰ Brennan-Marquez and Henderson's argument shows that this added accuracy or consistency must be purchased at the cost of rendering the entire system undemocratic. That bargain, they convincingly show, is not worth it. This point is worth emphasizing because a key argument in favor of AI judges is that they *improve* on human performance. Brennan-Marquez and Henderson effectively show that even if such improvement were more than imaginary, it would not be democratically permissible. For even if my skepticism of their critique is right, AI judges can at best perform only as good as the human judges they mimic. This is not to say that there would be no gain in accuracy. Perhaps AI judges would mimic only human judges who aren't underqualified, sleep deprived, over-worked, and so on.³¹ Nevertheless, they

²⁸ For a series of other reasons against endorsing AI judges, see, for example, John Tasioulas, *The Rule of Algorithm and the Rule of Law*, in 3 VIENNA LECTURES ON LEGAL PHILOSOPHY 17–39 (2023), available at <https://perma.cc/D48Q-SAF3>.

²⁹ Thanks to Artur Pericles Lima Monteiro for pressing me to emphasize this point.

³⁰ Brennan-Marquez & Henderson, *supra* note 2, at 149. See also *id.* at 138 (citing MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 105 (2017) ("What are the first associations that come to your mind when you think about the court system in your country? If it's lengthy delays, high costs and occasional injustice, then you're not alone. Wouldn't it be wonderful if your first thoughts were instead "efficiency" and "fairness"? Since the legal process can be abstractly viewed as a computation, inputting information about evidence and laws and outputting a decision, some scholars dream of fully automating it with robojudges: AI systems that tirelessly apply the same high legal standards to every judgment without succumbing to human errors such as bias, fatigue or lack of the latest knowledge.").

³¹ For tradeoffs that arise from the resulting marginal gains in accuracy and efficiency, see *AI Decisions*, *supra* note 14.

would not rid the system of human error, insofar as they would at best perform as good as the best human judges.

IV. GENERAL IMPLICATIONS

Normative evaluations of various phenomena can involve either or both formal and substantive criteria. Though conceptually distinct, many such criteria may be interdependent in practice. Disruptive technologies, however, sometimes change this fact. In evaluating such technologies, therefore, the analysis of substantive and formal criteria can and do come apart. Thus, the violation of a formal criteria, which may be sufficient to show the violation of substantive criteria in the absence of a technology, may not be sufficient to show the same in its presence. When evaluating such technologies, therefore, we must pay extra attention as to whether the violation of one can be inferred from that of the other.