Fordham Law Review

Volume 92 | Issue 5 Article 8

2024

AI, Algorithms, and Awful Humans

Daniel J. Solove
George Washington University School of Law

Hideyuki Matsumi
Health and Ageing Law Lab of the Vrije Universiteit Brussel

Recommended Citation
Daniel J. Solove and Hideyuki Matsumi, *AI, Algorithms, and Awful Humans*, 92 Fordham L. Rev. 1923 ().
Available at: https://ir.lawnet.fordham.edu/flr/vol92/iss5/8

AI, ALGORITHMS, AND AWFUL HUMANS

Daniel J. Solove & Hideyuki Matsumi***

A profound shift is occurring in the way many decisions are made, with machines taking greater roles in the decision-making process. Two arguments are often advanced to justify the increasing use of automation and algorithms in decisions. The "Awful Human Argument" asserts that human decision-making is often awful and that machines can decide better than humans. Another argument, the "Better Together Argument," posits that machines can augment and improve human decision-making. These arguments exert a powerful influence on law and policy.

In this Essay, we contend that in the context of making decisions about humans, these arguments are far too optimistic. We argue that machine and human decision-making are not readily compatible, making the integration of human and machine decision-making extremely complicated.

It is wrong to view machines as deciding like humans do, except better because they are supposedly cleansed of bias. Machines decide fundamentally differently, and bias often persists. These differences are especially pronounced when decisions require a moral or value judgment or involve human lives and behavior. Making decisions about humans involves special emotional and moral considerations that algorithms are not yet prepared to make—and might never be able to make.

Automated decisions often rely too much on quantifiable data to the exclusion of qualitative data, resulting in a change to the nature of the decision itself. Whereas certain matters might be readily reducible to quantifiable data, such as the weather, human lives are far more complex. Human and machine decision-making often do not mix well. Humans often perform badly when reviewing algorithmic output.

---

* Eugene L. and Barbara A. Bernard Professor of Intellectual Property and Technology Law, George Washington University Law School. Thanks to JP Schnabel and Alexa Adalian for excellent research assistance. We also thank the following people for great feedback on earlier drafts of this Essay: Rebecca Crootof, Ben Green, Aziz Huq, Margot Kaminski, James Lau, Michael Risch, and Alicia Solow-Niederman.

** Hideyuki Matsumi, or Yuki, is a PhD candidate/researcher at the Research Group on Law, Science, Technology and Society (LSTS) as well as at the Health and Ageing Law Lab (HALL) of the Vrije Universiteit Brussel (VUB). Member of the New York Bar. This Essay was prepared for the Symposium entitled *The New AI: The Legal and Ethical Implications of ChatGPT and Other Emerging Technologies*, hosted by the *Fordham Law Review* and cosponsored by Fordham University School of Law's Neuroscience and Law Center on November 3, 2023, at Fordham University School of Law.

We contend that algorithmic decision-making is being relied upon too eagerly and with insufficient skepticism. For decisions about humans, there are important considerations that must be better appreciated before these decisions are delegated in whole or in part to machines.

INTRODUCTION

Artificial intelligence (AI) algorithms are increasingly involved in a myriad of decisions about people's freedom, opportunities, and welfare. Algorithms are deployed in decisions as diverse as hiring, criminal sentencing, education, and lending.[1] We are witnessing a profound shift in the way that many decisions are made, with machines taking on greater roles in the decision-making process.

Many policymakers and commentators view this shift with optimism. They often advance two related arguments to justify and encourage the increasing use of automation and algorithms in decisions. The "Awful Human Argument" asserts that human decision-making is often bad and that machines can decide better than humans.[2] Another argument, the "Better Together Argument," posits that machines can augment and improve human decision-making.[3] These arguments exert a powerful influence on law and policy. The arguments share the optimism that delegating decisions in whole or in part to machines will be better than pure human decision-making.

In this Essay, we express skepticism about these arguments for decisions made about humans.[4] Algorithms change the nature of decisions, shifting

1. *See* Hideyuki Matsumi & Daniel J. Solove, The Prediction Society: Algorithms and the Problems of Forecasting the Future 12–19 (June 5, 2023) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453869 [https://perma.cc/PC3C-L5ZS].

2. *See infra* Part I.

3. *See infra* Part I.

4. We are generally not addressing decision-making involving nonhumans, such as the use of algorithms in certain circumstances for autonomous vehicles, manufacturing, agriculture, and so on. Our focus is on decisions producing legal or similarly significant effects on individuals, a distinction made in the General Data Protection Regulation (GDPR)

them toward quantifiable data and away from qualitative elements.[5] Although this shift can bring benefits, there are also significant costs that are often underappreciated. Whereas certain matters might be readily reducible to quantifiable data, such as the weather, human lives are far more complex. Machine decision-making currently cannot incorporate emotion, morality, or value judgments, which are essential components of decisions involving people's welfare.[6] The increased use of automation in decisions can lead to changes in the weight given to certain factors over others or affect how conflicting goals are resolved—not necessarily in better ways. When machine and human decision-making are integrated, the focus of decisions can shift heavily to automated dimensions and neglect the moral issues involved.

We contend that algorithmic decision-making is being relied on too eagerly and without sufficient skepticism. For decisions about humans, there are important considerations that must be better appreciated before these decisions are delegated in whole or in part to machines. Although it is possible that using more algorithmic decision-making could lead to better outcomes, many policymakers and commentators fail to appreciate what is lost when machines replace human decision-makers, as well as the complexity of mixing human and machine decision-making.

In Part I, we discuss optimistic claims about AI, which are often cast in terms of comparison to flawed human decision-making. In Part II, we argue that there are enormous challenges for the successful use of machine decision-making about people. In Part III, we contend that more thought must be devoted to decisions about humans; adding machines might appear to be a solution, but often the machines create new problems and take focus away from difficulties with the underlying decision.

I. BETTER WITH MACHINES?: THE RISE OF AI OPTIMISM

The Awful Human Argument and the Better Together Argument are frequently invoked to encourage the replacement or augmentation of human decision-making with algorithms and automated systems.

Human decision-making is fraught with problems. Humans are infected with bias; they have a limited range of experience to draw on; they often are slow and inefficient when making decisions; they are swayed by emotion and irrational factors; they can be impulsive; and they exhibit various cognitive biases and heuristics that can lead them astray.

---

to determine when the law should intervene with automated decisions. *See* Council Regulation 2016/679, General Data Protection Regulation, 2016 O.J. (L 119) 1, 46. In some cases, initial decisions about nonhumans can have significant effects on humans, such as when an autonomous vehicle might make a moral choice regarding an accident, like in the famous "Trolley Problem." *See* Iria Giuffrida, *Liability for AI Decision-Making: Some Legal and Ethical Considerations*, 88 FORDHAM L. REV. 439, 453–54, 454 n.89 (2019). In cases in which decisions have effects on humans and involve moral choices, our skepticism is also warranted.

5. *See infra* Part II.A.

6. *See infra* Part II.B.

The Awful Human Argument contends that, when compared to how badly humans decide, algorithms are superior.[7] Machines offer the promise of cleansing away the stains of bias; machines are consistent, fast, and efficient. According to Professor Cary Coglianese and Lavi Ben Dor, "new digital technologies that rely on machine-learning algorithms to process vast quantities of data are making highly accurate predictions that often outperform humans in executing important tasks."[8] Professor Cass Sunstein contends that algorithms can eliminate "noise"—"unwanted variability in judgments."[9] Algorithms "prevent unequal treatment and reduce errors."[10] Unlike humans, algorithms "do not use mental shortcuts; they rely on statistical predictors, which means that they can counteract or even eliminate cognitive biases."[11]

The Better Together Argument contends that humans and machines make better decisions together than human decision-makers make alone. Algorithms provide additional information to aid human decisions, and more information is better than less information. For example, in *State v. Loomis*,[12] a criminal defendant challenged a trial court's use of Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a recidivism risk assessment algorithm.[13] The defendant contended that using the algorithm violated his right to due process in part because it was based not on data about him but on recidivism statistics of others.[14] The Wisconsin Supreme Court rejected the defendant's arguments and held that the algorithm was merely providing the sentencing judge with "more complete information."[15] Similarly, in *Malenchik v. State*,[16] the Indiana Supreme Court held that sentencing assessment tools "enable a sentencing judge to more effectively evaluate and weigh several express statutory sentencing considerations."[17]

Many commentators express optimism that machine and human decision-making can be integrated successfully. As Professor Orly Lobel argues: "[D]espite its risks and flaws, digitization can and must become a powerful force for societal good—for fairness, inclusion, economic growth, expanded opportunities, innovation, and, above all else, equality. AI can be a force for debiasing, and it can provide early detection of discrimination and

7. Jenna Burrell & Marion Fourcade, *The Society of Algorithms,* 47 ANN. REV. SOCIO. 213, 222–23 (2021) (critiquing arguments that algorithmic systems are more fair than human decision-makers).

8. Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. 791, 791 (2021).

9. Cass R. Sunstein, *Governing by Algorithm?: No Noise and (Potentially) Less Bias*, 71 DUKE L.J. 1175, 1177–78 (2022).

10. *Id.* at 1177.

11. *Id.*

12. 881 N.W.2d 749 (Wis. 2016).

13. *See id.* at 756.

14. *Id.* at 764.

15. *Id.* at 765.

16. 928 N.E.2d 564 (Ind. 2010).

17. *Id.* at 574.

abuse."[18] Sunstein notes that "[a]lgorithms can encode or perpetuate discrimination, perhaps because their inputs are based on discrimination or because what they (accurately) predict is infected by discrimination."[19] However, he contends that "properly constructed algorithms nonetheless hold a great deal of promise for the administrative state."[20] In a similar fashion, as Sam Corbett-Davies, Professor Sharad Goel, and Professor Sandra González-Bailón argue: "Poorly designed algorithms can indeed exacerbate historical inequalities, but well-designed algorithms can mitigate pernicious problems with unaided human decisions."[21] Many other commenters note the potential for algorithms to be created in ways to reduce bias.[22]

Unfortunately, we are more pessimistic. Finding a happy harmony between human and machine decision-making will be immensely difficult, and we have yet to see a workable blueprint for doing so. We are sorry to spoil the party, but the stakes are high, and the dangers are significant.

II. CHALLENGES FOR MACHINE DECISION-MAKING ABOUT PEOPLE

On the surface, the Awful Human Argument and the Better Together Argument seem to be relatively uncontroversial. Of course, human decision-making can be awful. Of course, algorithms can avoid many of these problems and augment human decision-making.

Although these arguments appear unassailable on the surface, they often ignore considerations that should give us pause. Humans are bad, but this fact should not lure us into concluding that algorithms are better.[23] Machine decision-making might appear to be a perfection of human decision-making, cleansing it of bias and making it more efficient, fast, objective, and data-driven. But machine decisions are fundamentally different from human ones.[24] Comparing human to machine decision-making is akin to comparing

---

18. ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE 3 (2022).

19. Sunstein, *supra* note 9, at 1177.

20. *Id.*

21. Sam Corbett-Davies, Sharad Goel & Sandra González-Bailón, *Even Imperfect Algorithms Can Improve the Criminal Justice System*, N.Y. TIMES (Dec. 20, 2017), https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html [https://perma.cc/TKV4-ZK6W].

22. *See, e.g.*, Carolin Kemper, *Kafkaesque AI?: Legal Decision-Making in the Era of Machine Learning*, 24 INTELL. PROP. & TECH. L.J. 251, 285 (2020) ("Algorithms can be designed to alleviate biases . . . ."); Mirko Bagaric, Dan Hunter & Nigel Stobbs, *Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire*, 88 U. CIN. L. REV. 1037, 1040 (2020) (positing that it is "achievable" to ensure that algorithms "do not discriminate directly or indirectly").

23. Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1538 (2019) ("Human decision-making can be deeply, terribly flawed . . . . It is thus tempting to believe that machines will be better.").

24. *See* Mireille Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, THEORETICAL INQUIRIES L., Jan. 2019, at 83, 83–84.

apples and oranges, not rotten apples to fresh ones. The differences between human and machine decisions make successful integration exceedingly complicated. Although machines can help human judgment, this is far from certain, and there are many reasons why increasingly turning to machines might make decisions worse.

A. Quantifiable and Qualitative Judgments

Machines excel at processing quantifiable data. Quantification is transformative and can skew and distort decisions that have qualitative dimensions. Problems emerge when too much quantitative data is relied on to the exclusion of qualitative data because not everything is readily quantifiable.

In *Seeing Like a State*, Professor James Scott argues that making things "legible" (more capable of being understood and analyzed) involves simplification—a "narrowing of vision" that "brings into sharp focus certain limited aspects of an otherwise far more complex and unwieldy reality."[25] This simplification can make things "more susceptible to careful measurement and calculation."[26] When modern institutions seek to "see" things by gathering quantifiable data, they are seeing through a particular lens, and much is excluded from the picture. Scott warns that "thin, formulaic simplifications" can lead to tragic failures.[27]

Quantification can certainly lead to insights that we might not otherwise recognize. But the fact that we can see certain things through quantification does not mean that quantification is a superior way of knowing or that it should be the only way of examining things. Unfortunately, it is easy to become beguiled by large datasets because they help us see in exciting new ways. Lambert Adolphe Jacques Quetelet, an early pioneer in statistics, enthusiastically proclaimed: "The 'greater the number of individuals observed, the more do individual particularities, whether physical or moral, become effaced, and leave in a prominent point of view the general facts, by virtue of which society exists and is preserved.'"[28]

Quetelet believed that statistics was a more refined way of understanding humankind than considering the irregularities of specific individuals.[29] But focusing on "general facts" omits the rich tapestry of peculiarities of individuals and human nature.[30] As Professor Laurence Tribe warns,

25. JAMES C. SCOTT, SEEING LIKE A STATE: HOW CERTAIN SCHEMES TO IMPROVE THE HUMAN CONDITION HAVE FAILED 11 (1998).

26. *Id.*

27. *Id.* at 309. Professors Marion Fourcade and Kieran Healy note how personal data is being used to measure, score, and analyze individuals. Marion Fourcade & Kieren Healy, *Seeing Like a Market*, 15 SOCIO.-ECON. REV. 9, 10–11 (2017).

28. CHRIS WIGGINS & MATTHEW L. JONES, HOW DATA HAPPENED: A HISTORY FROM THE AGE OF REASON TO THE AGE OF ALGORITHMS 26 (2023) (quoting ADOLPHE QUETELET, A TREATISE ON MAN AND THE DEVELOPMENT OF HIS FACULTIES 6 (Robert Knox & Thomas Smibert trans., Thomas Smibert ed., 1842)).

29. *See id.*

30. *Id.* (quoting ADOLPHE QUETELET, A TREATISE ON MAN AND THE DEVELOPMENT OF HIS FACULTIES 6 (Robert Knox & Thomas Smibert trans., Thomas Smibert ed., 1842)).

"readily quantifiable variables" could "dwarf those that are harder to measure."[31]

Certainly, statistics can be quite useful, and particular attempts to rank, score, or infer based on aggregated standardized data can be valuable. But these practices can be fraught with danger because algorithmic systems do not just see the world; they simplify it. As Professors Chris Wiggins and Matthew Jones contend, "[s]tatistics doesn't simply represent the world. It transforms how we categorize and view the world. It transforms how we categorize others and ourselves. It changes the world."[32]

For example, ranking systems and scoring systems are designed to make things more comparable so hierarchies can be established. But such systems are reductive; they create the illusion of comparability for things that are not necessarily comparable.[33] Writing about the IQ test, Professor Neil Postman argues that it "assume[s] that intelligence is not only a thing, but a single thing, located in the brain, and accessible to the assignment of a number."[34] Yet some have posited that we have multiple types of intelligence, not just a singular type.[35]

When data is ingested into AI systems, the data must be shucked of extraneous information that isn't digestible for the algorithm.[36] When the rich details of life are turned into quantifiable data, essential information can be lost. As one of us has argued, the "information in databases often fails to capture the texture of our lives. Rather than provide a nuanced portrait of our personalities, compilations of data capture the brute facts of what we do without the reasons. . . . In short, we are reconstituted in databases as a digital person composed of data."[37]

Qualitative judgments are hard to make, and humans often succumb to the temptation to find shortcuts, which frequently involve quantified data. Consider law school rankings. A magazine, *U.S. News and World Report*, has ranked law schools using a formula that has been vehemently criticized by law schools as being reductive, inaccurate, and unfair.[38] Schools have

---

31. Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1393 (1971).

32. WIGGINS & JONES, *supra* note 28, at 57.

33. STEFFEN MAU, THE METRIC SOCIETY: ON THE QUANTIFICATION OF THE SOCIAL 33 (2019).

34. NEIL POSTMAN, TECHNOPOLY: THE SURRENDER OF CULTURE TO TECHNOLOGY 130 (1992).

35. *See generally* HOWARD GARDNER, FRAMES OF MIND: THE THEORY OF MULTIPLE INTELLIGENCES (1983).

36. *See* Dan L. Burk, *Algorithmic Legal Metrics*, 96 NOTRE DAME L. REV. 1147, 1158 (2021) ("To reconfigure otherwise incompatible data, analytical processing imposes a radical decontextualization on the data, paring away extraneous information and meanings.").

37. DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 49 (2004).

38. *2023–2024 Best Law Schools*, U.S. NEWS & WORLD REP., https://www.usnews.com/best-graduate-schools/top-law-schools/law-rankings [https://perma.cc/3CGT-L7TS] (last visited Mar. 3, 2024).

attempted to boycott the rankings by withholding data, but the magazine continues to rank schools by using an even more simplistic formula.[39]

Can law schools be reduced to a singular ranking? People crave rankings, partly because they simplify complex things and appear to make choices easier by ignoring incommensurate dimensions. But this simplicity is just a mirage; rankings distort in rather crude ways. One can reduce a Shakespearean play to a simple plot summary, but this removes the most important part—the artistry and beauty of the language. History can be reduced to dates and events, but that does not make it useful or insightful.

Any scoring or ranking system involves certain value judgments about which elements matter the most. Skewing occurs because scoring and ranking systems favor quantifiable elements over ones that are hard to measure and quantify, even if the quantifiable elements are not the ideal ones to use.

Scoring, ranking, and other quantifiable ways of making judgments can be useful heuristics. They are a way to address the great complexity of making certain decisions, but there are situations in which the usefulness of certain heuristics comes at the cost of distortion that nullifies or even outweighs the benefits. Machine decision-making pushes decisions more toward quantifiable methods when the opposite direction might be more desirable.

B. Emotion, Morality, and Value Judgments

Human judgment has important dimensions that machine decision-making lacks, such as emotion and nonrational elements. Machines decide fundamentally differently than humans do. These differences are especially pronounced when decisions have a moral component or involve human lives and behavior. Making decisions about humans involves special emotional and moral considerations that algorithms are not yet prepared to make—and might never be able to make.

Some might be tempted to equate thinking to rationality, with emotions clouding lucid thought, but there are many dimensions to human decision-making beyond rationality, and these nonrational elements are often underappreciated. Professor Martha Nussbaum aptly argues that "emotions are suffused with intelligence and discernment" and involve "an awareness of value or importance."[40] Emotions are "part and parcel of the system of ethical reasoning."[41]

Yet algorithms do not experience emotions. Algorithms might be able to mimic what people might say or do, but they do not understand emotions or feel emotions, and it is questionable how well algorithms will be able to incorporate emotion into their output.

39. Stephanie Saul, *U.S. News Releases Its Latest, Disputed Rankings of Law and Medical Schools*, N.Y. TIMES (May 11, 2023), https://www.nytimes.com/2023/05/11/us/us-news-rankings-law-medical-schools.html [https://perma.cc/879Q-6H5G].

40. MARTHA C. NUSSBAUM, UPHEAVALS OF THOUGHT: THE INTELLIGENCE OF EMOTIONS 1 (2001).

41. *Id.*

Based on where the technology currently is, algorithms should not be seen as akin to better-thinking humans. They do not think, and they do not feel. As Professor Aziz Huq notes, "machine decisions are not presently appropriate for decisions with ethical or normative components."[42] Although emotion can certainly make decisions worse, it can also make them better. Nonrational elements can play a positive role when decisions are made about humans.

Humans can make decisions about moral and normative issues in ways that machines cannot. Making normative choices can involve tradeoffs and compromises between different values, which can be incommensurate and conflicting. Consider, for example, college admissions decisions. There is a value in making decisions consistently. The most consistent way to decide might be to look only at test scores and grade point averages. But some applicants might have unique experiences or skills that are highly valuable. Other values in the admissions process involve establishing a balance in the overall class—gender balance, diversity, and a variety of interests, among other things. Negotiating all these considerations is complicated because no one consideration might dominate. In selecting students, we might want a little bit of everything. The ideal might be some degree of consistency, but also some swerves for special cases, as well as some skewing to navigate conflicting values. This is a difficult recipe for machines to follow. Machines are most adept at consistency; they struggle when told to do inconsistent things.

It is easy to point to grades and scores as more objective, even though these more quantifiable things can also be influenced by privilege and bias.[43] In the end, college admissions decisions, as well as many other types of decisions, involve value judgments and tradeoffs that machines are currently unable to make.

Many decisions about humans are fraught with complex moral and policy issues that elude an easy answer. For example, in sentencing, is recidivism an appropriate factor to focus on? If so, how much weight should it be given? Algorithms can try to predict recidivism. But they cannot answer the normative question of how much recidivism should factor into the sentencing calculation.

The use of machine decision-making tools can not only skew decision-making toward quantitative elements and away from qualitative ones, but also shift attention away from larger normative issues that

42. Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 685 (2020).

43. Greg J. Duncan, Ariel Kalil & Kathleen M. Ziol-Guest, *Increasing Inequality in Parent Incomes and Children's Schooling*, 54 DEMOGRAPHY 1603, 1623 (2017) (finding that parental income level affects children's educational success); Pamela E. Davis-Kean, Lauren A. Tighe & Nicholas E. Waters, *The Role of Parent Educational Attainment in Parenting and Children's Development*, 30 CURRENT DIRECTIONS PSYCH. SCI. 186, 186 (2021) (finding that parental education affects children's success).

algorithms cannot address.[44] In criminal sentencing, for example, focusing on whether an algorithm is accurately predicting recidivism can detract focus from the weight recidivism should be given in a sentencing decision.[45] As Professor Jessica Eaglin notes, "the institutionalization of actuarial risk assessments at sentencing reflects the extension of a larger, historically situated push to move judges away from passing moral judgment on individual defendants and toward basing sentencing on population-level representations of crimes and offenses."[46] Thus, humans might rely too much on the factors the algorithm computes because algorithmic output is overly trusted and wrongly viewed as objective.

C. Goals and Tradeoffs

The goals for many decisions are often contested and even contradictory. Machines cannot readily resolve these conflicts.[47] Machine decision-making can lead to certain goals being privileged over others because they are more achievable by machines, but this might not necessarily be the optimal resolution to a tradeoff between goals.

Algorithmic decisions are often touted as more "accurate" than human ones.[48] But it remains questionable whether algorithms are really as accurate as their evangelists contend.[49] Moreover, we should not assume that accuracy is the key goal, and it is unclear what an "accurate" decision even means. Often, the goals of a particular decision are unclear, contested, or even conflicting. Accuracy is often a key goal, but so is having decisions be free from bias. The challenge is that these goals are often in tension with each other. Quantifiable data does not consist of neutral facts; such data is created and curated by humans, which introduces bias into the algorithm.[50] Hypothetically, as AI optimists might propose, the data could be scrubbed of all bias. But even if the most obvious biases are removed from the data, the

---

44. Alicia Solow-Niederman, YooJung Choi & Guy Van den Broeck, *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L.J. 705, 708 (2019) ("[L]ooking at issues such as fairness or bias in a tool in isolation elides vital bigger-picture considerations about the institutions and political systems within which tools are developed and deployed.").

45. *See* Jessica M. Eaglin, *Population-Based Sentencing*, 106 CORNELL L. REV. 353, 364–65 (2021) (arguing that, in a criminal case, relying on a population-based sentencing tool to measure recidivism risk "obscures the peculiar facts of the case and the individual characteristics of the actual defendant").

46. *Id.* at 357.

47. Ben Green, *Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness*, PHIL. & TECH., Oct. 8, 2022, at 1, 3 (noting that "it is impossible for an algorithm to satisfy all desirable mathematical definitions of fair decision-making" because "[a]n algorithm that is fair along one standard will inevitably be unfair along another standard").

48. Matsumi & Solove, *supra* note 1, at 5.

49. *Id.*

50. *See, e.g.*, Ngozi Okidegbe, *Discredited Data*, 107 CORNELL L. REV. 2007, 2010–13 (2022) (noting that pretrial risk assessment algorithms reproduce inequities produced by human decision-makers because of the "racial and socioeconomic bias in their data sources").

data still comes from a society of humans in which bias infects nearly everything.[51]

Often, trying to design an algorithm to produce accurate and unbiased results is asking it to do two contradictory things. If accuracy means reflecting society, then the accurate output will be biased because society is riddled with bias. An algorithm often cannot produce an accurate and unbiased decision; it might be able to produce one or the other.

It is not clear what "accuracy" means for many decisions or if it is the right goal for these decisions. Regarding algorithmic predictions about recidivism, we have argued elsewhere that such predictions cannot be verified because the future has not occurred yet, and the predictions lead to interventions that alter the future.[52] Algorithmic predictions can become a self-fulfilling prophecy.[53] Thus, with many algorithms, it is not possible to assess accuracy. There is often no one right answer out there in a cosmic answer key.

If the decision is whether a person is guilty of a crime, then there can be a correct answer. If the decision is how long a criminal sentence should be, then there is not necessarily a correct answer. The answer depends on one's theory of punishment as well as other goals in the criminal justice system. It is far from clear that we can achieve an "accurate" punishment. Instead, we should strive for a *just* punishment, one that is fair and that achieves valuable social ends. There is no clear right or wrong answer when it comes to an appropriate punishment, as it involves balancing many considerations, such as fairness to individuals and societal interests. Important values might be in conflict; tradeoffs must be made. The decision teems with normative issues.

Algorithms may attempt to predict recidivism, but recidivism is dynamic and influenced by the sentence itself and how it is carried out.[54] Recidivism is a product of many factors, not something inherent in a person. It can be shaped by a person's prison experience, by education and training during incarceration, by interactions with other inmates, and by opportunities and social acceptance upon release.[55] Of course, the person plays a role too, but

51. *See, e.g.*, *id.* at 2024–26; Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, 26 WM. & MARY BILL RTS. J. 287, 289 (2017); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224 (2019).

52. *See generally* Matsumi & Solove, *supra* note 1.

53. *See id.* at 29–31.

54. *See* AVINASH SINGH BHATI, URB. INST., JUST. POL'Y CTR., STUDYING THE EFFECTS OF INCARCERATION ON OFFENDING TRAJECTORIES: AN INFORMATION-THEORETIC APPROACH 5 (2006) (finding that the "risk of recidivism is not a static but a dynamic measure.").

55. *See* David J. Harding & Heather M. Harris, *Introduction*, *in* AFTER PRISON: NAVIGATING ADULTHOOD IN THE SHADOW OF THE JUSTICE SYSTEM 6 (David J. Harding & Heather M. Harris eds., 2020) (noting that "[t]he social contexts, institutional involvements, and social supports" present on release from prison affect recidivism); *Impact of Prison Experience on Recidivism*, NAT'L INST. JUST. (Oct. 3, 2011), https://nij.ojp.gov/topics/articles/impact-prison-experience-recidivism [https://perma.cc/WY8V-CLU5] ("[B]eing exposed to individuals who have higher propensities to crime may increase criminal behavior or reinforce antisocial attitudes."); Holly Wetzel, *Research Finds Prison Education Programs Reduce Recidivism*, MACKINAC CTR. FOR PUB. POL'Y (Jan. 26. 2023), https://www.mackinac.org/pressroom/2023/research-finds-prison-education-programs-reduce-recidivism [https://perma.

recidivism is produced by an interaction of many different factors on an individual. There is no static answer as to whether a person will recidivate, and it depends on the very sentencing decision in which recidivism is a factor.

Additionally, bias-free decision-making is often an impossible ideal. Algorithms may be programmed to eliminate biases that the law forbids, such as race, religion, or gender. But algorithms can create new categories of disfavored traits such as height, weight, eye color, or athleticism. Ultimately, algorithms will invariably identify certain traits to favor and disfavor. New winners and losers will be created based on these characteristics.[56]

No matter whether bias is traditional or new, algorithms make the problem of bias worse because they can make decisions en masse at a scale and consistency that humans cannot achieve. Regarding the use of algorithms in hiring decisions, Stijin Broecke contends that "even though human beings can be biased when making hiring decisions, the adverse impact of AI could be far greater by virtue of the volume and velocity of the decisions it takes, which could systematize and multiply bias."[57] In contrast, human decisions are often isolated to a smaller number of cases and are much less consistent. Human inconsistency can lessen the harm of bias; if bias is inconsistent, it is not as bad as consistent bias. Moreover, inconsistent human decisions might reflect differing societal values and ambivalence about a particular issue. Certainly, inconsistency can be problematic, but it also has virtues. A completely consistent set of decisions might not be optimal; for example, it could be oppressive if it systematically disfavors certain people.

D. Tensions Between Humans and Machines

Algorithmic decision-making is not necessarily better than human decision-making. It is quite different, and it can be better or worse depending on the circumstances and goals. Moreover, it is complicated to determine how to measure "better" or "worse." Since antiquity, a debate has raged in jurisprudence regarding whether to follow the rule of law rigidly or whether to make exceptions out of compassion or justice. What is better or worse depends on one's philosophical and moral views.

But could human and machine decision-making be combined to produce the perfect mix? Maybe, but it is far from simple. Merely stirring humans and machines together will not readily make an inspired blend of the best each has to offer. In many cases, human and machine decision-making do not mix very well.[58] At the very least, blending requires a complicated

---

cc/P5B3-Q76Q] (noting that research shows that "prison workforce and education programs reduce the likelihood of recidivism by 14.8%").

56. *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 3 (2016).

57. Stijin Broecke, *Artificial Intelligence and the Labour Market: Introduction*, *in* OECD EMPLOYMENT OUTLOOK 2023: ARTIFICIAL INTELLIGENCE AND THE LABOUR MARKET 93, 96 (Andrea Bassanini & Stijn Broecke eds., 2023), https://read.oecd-ilibrary.org/employment/oecd-employment-outlook-2023_08785bba-en [https://perma.cc/4PSF-Y73C].

58. *See, e.g.*, Kaminski, *supra* note 23, at 1538–40, 1546–47.

recipe, and such recipes are lacking in law and policy and have been elusive thus far.

The hope that humans and machines can decide better together is not just vague and unsubstantiated; in fact, strong evidence demonstrates that there are significant problems with combining humans and machines in making decisions. Humans can perform poorly when using algorithmic output because of certain biases and flaws in human decision-making.[59] Far from serving to augment or correct human decision-making, algorithms can exacerbate existing weaknesses in human thinking, making the decisions worse rather than better. As Professors Rebecca Crootof, Margot Kaminski, and Nicholson Price observe, a "hybrid system" consisting of humans and machines could "all too easily foster the worst of both worlds, where human slowness roadblocks algorithmic speed, human bias undermines algorithmic consistency, or algorithmic speed and inflexibility impair humans' ability to make informed, contextual decisions."[60]

Professor Ben Green points out an even more fundamental conflict between algorithmic and human decision-making: algorithms offer "consistency and rule-following" whereas humans offer "flexibility and discretion."[61] When policymakers call for humans to oversee algorithms, they often do not recognize the "inherent tension" between these things and fail to provide sufficient guidance about how to resolve this tension.[62]

Many commentators have pointed out that people are prone to "automation bias"—to trust algorithms without sufficient skepticism.[63] As one court declared in 1968: "Mathematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not cast a spell over him."[64] Tribe contends that "the very mystery that surrounds mathematical arguments—the relative obscurity that makes them at once impenetrable by the layman and impressive to him—creates a continuing risk that he will give such arguments a credence they may not deserve and a weight they cannot logically claim."[65]

Various methods of quantification and automation create an illusion that humans are not involved in shaping the output, which will then be free of the

---

59. *See* Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 468 (2023).

60. *Id.*

61. Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, COMPUT. L. & SEC. REV., Apr. 26, 2022, at 1, 12.

62. *Id.*

63. Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1253 (2008) ("Computer programs seamlessly combine rulemaking and individual adjudications without the critical procedural protections owed either of them."); Burk, *supra* note 36, at 1161; Katrina Geddes, *The Death of the Legal Subject*, 25 VAND. J. ENT. & TECH. L. 1, 36 (2023) (defining "automation bias" as "the tendency of human decision makers to accept, rather than challenge, quantitative assessments"); Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1961 (2021) ("Humans may exhibit an 'automation bias' that creates overconfidence in machine decisions, and an ensuing bias against challenges to those decisions." (footnote omitted)).

64. People v. Collins, 438 P.2d 33, 33 (Cal. 1968).

65. Tribe, *supra* note 31, at 1334.

taint of human bias or animus. But this is far from the case; machines hide the human element but do not eliminate it.[66] AI is far from human-free; humans are involved in nearly every aspect of AI at every stage of development.[67] As Kaminski observes, behind each machine, there are humans who make a myriad of decisions involving the design, goals, inputs, and testing of algorithms.[68] Professor Ifeoma Ajunwa aptly analogizes AI to the infamous Mechanical Turk, an automatic chess playing machine that bedazzled audiences in the late nineteenth century.[69] The machine was a fraud—a human was concealed inside.[70] Much like the Mechanical Turk, machine decision-making appears to be human-free but actually involves humans who are often invisible and unaccountable. Hiding the human element perpetuates the false trust that machine decision-making is objective, neutral, and unbiased.

Humans struggle to oversee algorithms. Humans lack the ability to process large datasets or understand everything that algorithms are doing.[71] Empirical studies show that people readily defer to automated systems, overlook errors in algorithms, and deviate from algorithmic output in ways that render a less accurate result.[72] Moreover, Green notes, "people cannot reliably balance an algorithm's advice with other factors, as they often over-rely on automated advice and place greater weight on the factors that algorithms emphasize."[73] Studies show that "[a]utomation can also create a diminished sense of control, responsibility, and moral agency among human operators."[74]

Of course, humans can be educated and warned about these problems, but it is not clear that this can save the day. In one study, researchers gave participants the same warnings about COMPAS's limitations as required by

---

66. Kaminski & Urban, *supra* note 63, at 1969 (noting that AI "obfuscates bias with layers of ostensibly objective mathematical authority"); Burk, *supra* note 36, at 1161 ("Scored data thus impart an illusion of objectivity because they are rendered by machines, and their human initiators appear distantly removed from the ultimate output.").

67. *See generally* KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021). *See* Kaminski, *supra* note 23, at 1538–39 ("[F]or any kind of algorithm, human choices and assumptions go into its construction, training, and oversight—or lack thereof.").

68. Kaminski, *supra* note 23, at 1538–39.

69. Ifeoma Ajunwa, *The Paradox of Automation as Anti-bias Intervention*, 41 CARDOZO L. REV. 1671, 1704–07 (2020).

70. *Id.*

71. *See* John Zerilli, Alistair Knott, James Maclaurin & Colin Gavaghan, *Algorithmic Decision-Making and the Control Problem*, 29 MINDS AND MACHINES 555, 560 (2019) ("Humans are often at a severe epistemic disadvantage vis-à-vis the systems they are tasked with supervising.").

72. Green, *supra* note 61, at 7; *see also* Meg Leta Jones, *The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles*, 18 VAND. J. ENT. & TECH. L. 77, 91 (2015) ("Rare, abnormal conditions are difficult to detect when inappropriate deference and trust of the machine (automation bias) builds in human operators interacting with a well-functioning system.").

73. Green, *supra* note 61, at 9.

74. *Id.* at 7.

the Wisconsin Supreme Court in *Loomis*.[75] The researchers found that these warnings had "no significant effect" on the decisions made by participants.[76] Humans can become "complacent" and "over-reliant" when working with an autonomous system—a tendency that "seems to afflict experts as much as novices, and is largely resistant to training."[77]

Moreover, there might be significant pressures for humans to accept algorithmic output rather than deviate from it by using their discretion and emotion or relying on intangible considerations. Deviations from the algorithm will likely be viewed quite suspiciously, require extensive justification, and take much effort and moxie for human decision-makers. Not only are humans internally more inclined to rubber-stamp algorithms, but external pressures might also push them toward doing so.

Algorithms can make decisions appear more scientific, more "accurate," and more reliable and trustworthy. Even if the algorithms are used to make decisions better (assuming this can be determined), they might be an overall negative because they are dressing up a fraught activity in scientific-looking vestments. During the heyday of phrenology—the false idea that physical characteristics were linked to behavior—there was an attempt to dress it up with science.[78] This thin veneer of science was based on faulty studies.[79] Proponents raced forward with hope and enthusiasm; any contrary evidence was ignored in their zeal to succeed.[80] A similar story is unfolding with algorithms. Today, algorithms are widely being used to make many decisions about people, and it is far too soon to be doing so on such a large scale.

In some situations, adding automated tools to assist and improve human decisions (and human performance more generally) can lead to harm to humans. Automated systems can function as surveillance mechanisms that micromanage humans in tyrannical ways. For example, in the context of truck driving, Professor Karen Levy explores how AI is used "to address human 'weakness' through constant, intimate, visceral monitoring."[81] She notes how various automation technologies used to augment truck drivers adversely affect the drivers by creating an "intimate invasion into their work and bodies," resulting in an "uneasy, confrontational" relationship between worker and machine.[82]

75. Christoph Engel & Nina Grgić-Hlača, *Machine Advice with a Warning About Machine Limitations: Experimentally Testing the Solution Mandated by the Wisconsin Supreme Court*, 13 J. LEGAL ANALYSIS 284, 286–87 (2021).

76. *Id.* at 323.

77. Zerilli, et al., *supra* note 71, at 556.

78. *See* Pierre Schlag, *Law and Phrenology*, 110 HARV. L. REV. 877, 886–87 (1997).

79. *See id.* at 891–92.

80. For more background on phrenology, see generally *id.* For an excellent account of how automated video interviewing systems are resurrecting a form of phrenology, see IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE 138–52 (2023).

81. KAREN LEVY, DATA DRIVEN: TRUCKERS, TECHNOLOGY, AND THE NEW WORKPLACE SURVEILLANCE 148 (2023).

82. *Id.*

Certainly, automation can help improve human decision-making, but it can come at a high cost to the humans it purportedly helps. Human-machine unions are not necessarily happy ones, and automation can become a tool of oppression.[83] Perhaps there might be ways to restructure human-machine relationships to minimize these negative effects. Perhaps any unavoidable negative effects are a justifiable price to pay for improved decision-making quality. But perhaps in some cases—maybe many cases—the benefits are not worth the costs.

Ultimately, much more study must be devoted to figuring out the proper ways for humans and machines to interact. As Crootof, Kaminski, and Price observe, the mere combination of humans and machines is far from a solution—it is a problem itself yet to be solved.[84]

III. RETHINKING THE HUMAN-MACHINE RELATIONSHIP

Currently, a good blueprint is lacking for how machine and human decision-making should be integrated. As Crootof, Kaminski, and Price aptly observe, "policymakers often assume that adding a human to a machine system will result in the best of both worlds. There is a seductive simplicity to this 'slap a human in it' approach."[85] They argue that the law often fails to address the problems that emerge when humans interact with machines.[86] Merely uniting humans and machines is naïve and simplistic. Extensive thought must be given to the roles humans do play and should play in the process, as well as where they should be added and how they should perform their roles.[87]

Policymakers must address the hidden human dimensions of machine decision-making tools and the way these tools are wrongly perceived as objective. Policymakers must address the hidden bias in the data that feeds AI algorithms and the way that quantified data skews the nature of decisions. Policymakers must find ways to combat the perception of AI output as more objective, to fight against humans being seduced by AI's anthropomorphism, and to appreciate that AI does not think like humans do. Combating these perceptions is all the more difficult when AI technologies are being designed to create them.

We must never forget that the law's overarching goal should be to ensure good decisions, or at least aim to prevent bad ones. Many attempts to integrate machine and human decision-making fail in large part because of problems with the structure and goals of the decisions themselves.[88] As we

83. *See supra* notes 81–82 and accompanying text.

84. Crootof, Kaminksi & Price, *supra* note 59, at 508 ("Human involvement is no panacea. If anything, it creates new problems to solve.").

85. *Id.* at 437.

86. *See id.*

87. *See id.* at 487.

88. Ben Green and Yiling Chen, *Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-States Government Contexts*, PROC. ACM ON HUM.-COMPUT. INTERACTION, Oct. 2021, at 1, 2 (noting that although predictions of risk can

discussed earlier, there are overarching problems with sentencing decisions that exist before the decisions are even made.[89] The definition of recidivism and the weight given to recidivism both play a major role in the decision, and these are issues that are often fraught with problems.[90] Humans must resolve these issues for a good decision to be made. Adding a machine into the equation often exacerbates the problems and takes the focus away from them.

Good qualities in decision-making include a commitment to the scientific method, humility, feedback loops, fairness, morality, lack of bias, empathy, due process, listening to all stakeholders, diversity, practicality, accuracy as to facts, critical reflection, philosophical depth, open-mindedness, awareness of context, and much more. Some decisions might call for more accuracy, but others less so. For landing a plane, we want high accuracy, but for decisions about school admissions, credit scoring, or criminal sentencing, other values are also quite important.[91] There is no one-size-fits-all approach to regulating AI, as the decisions it will be employed to help make are quite different and demand different considerations.[92]

Algorithms focus on correlation, not causation, and rarely do those designing the algorithm or using the algorithm ask why the correlations exist.[93] Human overreliance on algorithms can deter critical reflection, further study, and deeper thought about causal relationships. For example, in a decision to admit a student into law school, the algorithm might be programmed to look for students who will pass the bar and obtain a lucrative job. But this can overlook students who might struggle with the bar but will eventually pass. It can dismiss students who could succeed at the bar if given the right training or opportunities. It might fail to consider students who use the degree to go into a nonlegal career (e.g., politics, writing, or journalism) or who go into a low-paying public interest position.

Ultimately, the goal should be good decisions, and such decisions are quite varied and contextual. There is no substitute for good judgment. Awful humans produce awful results, whether alone or combined with a machine.

CONCLUSION

When we focus on how awful human decision-making can be, it is tempting to rush to embrace the machines. But machines decide differently than humans. Algorithms might bring greater uniformity but might also lead to rigidity when flexible decisions might be preferable. There are certainly

---

be "optimized for accuracy," policy decisions cannot, and "using risk assessments to improve people's predictions may not necessarily improve people's decisions").

89. *See supra* notes 45–46 and accompanying text.

90. Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 72 (2017).

91. *See supra* note 1 and accompanying text.

92. *See supra* Part II.B.

93. As Mireille Hildebrandt aptly notes, "profilers are not very interested in causes or reasons, their interest lies in a reliable prediction, to allow adequate decision making." Mireille Hildebrandt, *Defining Profiling: A New Type of Knowledge?*, *in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES 17, 18 (Mirielle Hildebrandt & Seth Gutwirth eds., 2008).

times when we want a standardized decision, but there are other times when we want a wise and creative one.

In the end, good decisions depend on good humans. We need much more thought and study about when and how to integrate machine with human decision-making. Ultimately, the success of this project depends on humans and how carefully they reflect about the nature and effects of algorithmic decisions.