BLACK BOX ARTIFICIAL INTELLIGENCE AND THE RULE OF LAW

JEFF WARD

It seems fitting to explore issues of emerging uses of algorithmic decision-making and artificial intelligence (AI) through an interdisciplinary publication like *Law & Contemporary Problems.* After all, the AI tools at the heart of these articles are being deployed in nearly every industry and in every corner of the globe. This small volume brings together leading thinkers in philosophy, ethics, data science, computer science, and law, who connect with us from Germany, Belgium, England, Columbia, and the United States.

This cosmopolitan and cross-disciplinary approach offers particular value for the exploration of socio-technical systems where AI influences meaningful determinations, distributions, and allocations of rights and responsibilities. Precisely because AI affects personal and professional opportunities, due process, and the rule of law, any narrow exploration set apart from the systems it shapes—where myopically technological inquiries might fail to include broader ethical and sociological scrutiny—could be misguided and potentially harmful. Such narrow explorations might not only fail to prioritize the rights and values we hold dear but might also undermine our abilities then to govern AI and the effects it has on the social and political systems we aim to protect. As such, ideal is a forum like *Law & Contemporary Problems* that brings together lawyers, ethicists, technologist, engineers, and others to consider these socio-technical systems across disciplines. When seeking a positive AI future, it will take a village.

Such a sprawling topic, though, also requires some constraints. As this volume's title "Black Box Algorithms and the Rule of Law" suggests, we have imposed two constraints here.

First, we focus on a particular subset of AI characterized as "black box AI." In his article *The Black Box Society: The Secret Algorithms that Control Money and Information*, contributing author Frank Pasquale showed that black-box systems are those "colonized by the logic of secrecy."[1] His article in this volume adds that "'black box AI' refers to any natural language processing, machine learning, textual analysis, or similar software which uses data which are not accessible to the data subject, or which deploys algorithms which are either similarly inaccessible, or so complex that they cannot be reduced to a series of rules and rule applications comprehensible to the data subject." In other words,

This Article is also available online at http://lcp.law.duke.edu/.

1. FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 2 (2015).

the workings of these black box AI systems lack transparency and may escape our capacities for intuitive explanation.

It should be noted, of course, that *all* kinds of AI tools—even those using simple symbolic or handcrafted algorithms that should be intuitive and accessible to human inquiry—might lack transparency and thus be characterized as black boxes.[2] For example, the algorithms in the infamous *Loomis* case discussed below and herein were opaque and inaccessible *not* because they were complex but rather because they were protected by trade secret and thus off limits to the defendant's review. Nonetheless, much of the present concern about black box AI derives from the modern technological phenomenon that the AI with the greatest predictive accuracy is often a kind of machine learning, including "deep neural networks," that tends to add opacity. These kinds of AI systems do not replicate the step-by-step logic and rules that might be intuitive to us. Instead, they discern patterns from massive sets of examples to produce increasingly accurate predictions. To achieve these predictive capabilities, the networks discern relevant features of the data that usually are not obvious, intuitive, or even explainable to humans. As the use of such tools grows, so too do our concerns about their black-box nature.

Why should we be concerned? If these new AI tools provide us with highly accurate outputs, should we care about their black-box characteristics? A common theme across the articles in this volume is that—in deciding how much we should be concerned—*context matters*. A second way in which this volume constrains the sprawling topic of AI is thus our focus on arenas in which the deployment of AI touches on the rule of law.

Widely shared across the far-ranging jurisdictions of this volume's authorship is a conception of the rule of law where all persons and institutions are held accountable to legal systems that are clear and publicly manifest, with independent judiciaries, and whose legal rules are applied fairly and consistently, enforced without bias, and explained clearly. Quite apparently, black box AI's definitional lack of transparency could challenge such conceptions of the rule of law. Especially in arenas where AI tools assist with decision-making, opacity undermines demands for public transparency and for the kinds of clear explanations that satisfy our due process expectations.

Even more, we need transparency into decision-making processes to ensure their fairness. Black box AI might undermine the rule of law by obscuring biases and allowing unfairness to persist unchecked. There is a growing awareness of both the opacity and bias concerns of black box algorithms in due process contexts, in part because of some recent, troubling cases. Perhaps most well-known, when a predictive AI system labeled Wisconsin resident Eric Loomis at "high risk" for recidivism, Mr. Loomis challenged his resulting six-year prison

2. This volume takes an inclusive (perhaps over-inclusive) approach to what constitutes algorithmic decision-making and artificial intelligence tools such that we include even simple, statistically-based algorithms such as the COMPAS tool discussed herein. Such a broad view assists our mutual project of defending the rule of law.

sentence on grounds that required insight into the workings of the predictive algorithm used as part of the pre-sentence investigation report.[3] As the Wisconsin Supreme Court denied his challenge, the algorithm remained a black box—a decision-making tool unavailable to review by and without meaningful explanation for the one facing deprivation of liberty—because it was protected by trade secret law.[4] For many, *Loomis v. Wisconsin* was a wake-up call: Even if this simple algorithm is not representative of some of the more sophisticated machine learning used widely today, in the wake of this decision, many have worked to raise awareness of the explanation and fairness issues wrought by black box AI. As noted, some have highlighted that opacity itself raises due process concerns where such tools are used in decision-making contexts,[5] while others have emphasized the danger that such tools could undermine fairness and exacerbate inequalities.[6]

Here, a moment of caution is warranted for all readers of this volume and all discussants about the role of AI in our society. To be sure, even our *pre*-AI social systems have involved opacity and biases. For example, as our authors De Mulder, Valcke, Vanderstichele, and Baeck make clear in the title of their article *Are Judges More Transparent Than Black Boxes?*, there is a long, well-documented history of bias in sentencing and judicial decision-making. In this way, new forms of algorithmic decision-making—by freeing us from the limits of deeply-entrenched human biases—could potentially offer the opportunity to *enhance* fairness. Still, even where achieving such promise might be possible, many observers raise concerns about the use of AI tools in due process contexts,[7] especially where systems trained on historical data might learn and then continue to replicate existing biases.[8] The fear, in short, is that AI systems might not eliminate biases but instead crystallize and obscure them. It is no wonder, then, that our evolving expectations would demand transparency and meaningful opportunities to challenge decisions that are based in some way on the workings of AI systems.[9] Not to do so might impede the full achievement of AI's promise

3. State v. Loomis, 881 N.W.2d 749, 760–61 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017).

4. Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1369–70 (2018).

5. *See*, *e.g.*, Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 105–06 (2019). ("automation. . .raises significant due process concerns, involving lack of notice and the opportunity to challenge the decision.").

6. *Criminal Law—Sentencing Guidelines—Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*—State v. Loomis*, 881 N.W.2d 749 (Wis. 2016)*, 130 HARV. L. REV. 1530, 1530–37 (2017).

7. *See, e.g.*, Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 636, 680, 692 (2017); *see* Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15, 46, 48 (2019).

8. *See* Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224 (2019) ("To adapt a computer-science idiom, 'bias in, bias out.'"); *see also, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

9. *See, e.g.*, Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1844–45 (2019); *see also*, Kiel Brennan-Marquez, "*Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1251, 1256 (2017).

and ensure some measure of its peril.

With this context in mind, this volume's collection of articles provides a brief tour of discussions around black box AI and the rule of law that is end-to-end—from design to deployment; wide-ranging—exploring areas of civil, criminal, administrative law, and more; and cutting-edge—challenging what's next in AI systems where black-box issues might arise.

In our first article, *Beyond the Prediction Paradigm*, authors Helm and Hagendorff look not only at the predictive policing technologies (PPTs) that are already widely used in many jurisdictions but also to the future as PPTs are inevitably deployed in more complex areas of criminal policing. These uses raise important technological challenges for the probabilistic logic of prediction and concomitant ethical concerns. Even beyond the insights this article offers in this specific criminal policing domain, it demonstrates the power and importance of technologists, ethicists, and others working together to address those challenges and concerns.

Páez's *Negligent Algorithmic Discrimination* then goes straight to the core of the aforementioned concerns about opacity and bias. In an arena like employment where discrimination is a primary concern, evidentiary issues are always significant, and the recent rise of the use of algorithmic tools has heightened the challenges of establishing disparate treatment or disparate impact. Páez suggests a potential creative evolution of the law of negligence as a means to meet the challenges wrought by the use of AI tools in employment settings, and, in this way, the article serves as inspiration for the kind of forward-looking discourse that the widespread deployment of AI demands.

Pasquale's *Normative Dimensions of Consensual Application of Black Box Artificial Intelligence in Administrative Adjudication of Benefits Claims* offers a helpful taxonomy of black box AI in administrative functions first by separating the analysis of such systems between those cases where claimants in administrative adjudications consent to their use and those cases where claimants either do not consent or else feel obligated to consent. The article then offers a normative evaluation of cases where an claimant's consent is clear. This kind of careful evaluation is increasingly necessary as black box AI is used more broadly, implicates benefits and detriments for more and more people, and calls upon larger, surveillance-driven data sets. Few, if any, have done more in recent years to elevate awareness and demonstrate meticulous analysis than this author.

As noted above, in *Are Judges More Transparent Than Black Boxes?*, De Mulder, Valcke, Vanderstichele, and Baeck help us to recall that judicial decision-making has always been fraught with black-box issues, even before the use of modern AI tools. In another clear demonstration of the benefits of cross-disciplinary approaches to the issue of black box AI and the rule of law, the

---

I also note that some challenge the very notion that transparency requirements or our current conceptions of explanation into black box AI would be meaningful to those harmed by AI-driven decision-making. *See, e.g.*, Lillian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 67 (2017).

authors explore the relationship between traditional judicial decision-making and the mathematical function maximization behind AI-driven tools. Their insights help us to ask better questions about rule of law and the ongoing challenges of uncertainty in legal outcomes.

In the final article *Seconds to Impact?: Regulatory Reform, New Kinds of Legal Services, and Increased Access to Justice*, authors Sandefur, Clarke, and Teufel make clear one of the United States' most widespread rule of law failings: the justice gap and the persistence of unmet legal needs for many. By providing an analysis of a current experiment in the reformation of legal services regulation, where AI tools might play a very significant role in closing the justice gap, they end this volume with an appropriate reminder of AI's potential promise and the stewardship that will be needed to fulfill it.

I mentioned at this outset that this volume is an international and cross-disciplinary affair, and that describes our team of special editors for this volume, too. I'd like to offer my warm thanks to my co-special editors: Luciano Floridi is a leading scholar on digital ethics, the philosophy of information, and the philosophy of technology who serves as Professor of Philosophy and Ethics of Information, University of Oxford and Fellow of Exeter College, Oxford. Cynthia Rudin is a computer scientist who explores how machine learning can be used to help humans make better decisions and who serves as Professor of Computer Science, Electrical and Computer Engineering, and Statistical Science at Duke University, where she directs the Prediction Analysis Lab.

Once again, achieving AI's best future will take a village.

Jeff Ward
Duke University
Durham, NC
Clinical Professor of Law
Associate Dean for Technology & Innovation
Director, Duke Center on Law & Tech