DUQUESNE UNIVERSITY
SCHOOL OF LAW

LEGAL STUDIES RESEARCH PAPER SERIES

AI, On Algorithmic Justice: A New Proposal Toward the Identification and Reduction of Discriminatory Bias in Artificial Intelligence Systems

Emile Loza de Siles
Professor of Law

2020

Duquesne University School of Law Research Paper
No. 2020-10

AI, On Algorithmic Justice: A New Proposal Toward the Identification and Reduction of Discriminatory Bias in Artificial Intelligence Systems

Emile Loza de Siles*
Assistant Professor of Law
Duquesne University School of Law

Abstract:

As powerful and useful as machine learning and other artificial intelligence systems (collectively, "artificial intelligence" or "AI") are and can be, there are grave concerns about racial and other discriminatory bias in the algorithms at the heart of these computational systems. These concerns are well-founded. There are numerous reports of discriminatory algorithmic bias as fact and eventuality.

The use of algorithmic systems is ubiquitous in banking, credit, housing, consumer goods and services, the juvenile and criminal justice systems, and many other sectors and applications. The immense and complex data ecosystems that fuel and propagate these algorithmic systems are constantly evolving and opaquely so. Decision makers and users, including judges and pre-sentencing investigators, for example, may lack adequate technolegal understanding and training as to how algorithmic systems work and the contexts in which these systems are appropriate to or should not be employed. Although technology workers and the IEEE and other professional societies are engaged in impactful actions, standards for the ethical development and governance of algorithmic systems are not yet in place. Similarly, the International Organization for Standardization ("ISO") is developing no fewer than six (6) standards around AI security and trust, but none is near completion.

---

* Emile Loza de Siles is assistant professor of law at Duquesne University School of Law and recipient of the University's 2020 Presidential Scholarship Award. She serves on the Institute of Electrical and Electronics Engineers' ("IEEE's") Artificial Intelligence & Autonomous Systems Policy Committee; Working Group P2863 for Organizational Governance of Artificial Intelligence; and Working Group P2895 for Standard Taxonomy for the Responsible Trading of Human-Generated Data.

Emile also is founder of Technology & Cybersecurity Law Group and has provided trusted legal advice, services, and consultations to Fortune 50 and other tech companies and entrepreneurs since 2003. She is assistant professor (adjunct) in the Cybersecurity Program, The Graduate School, University of Maryland Global Campus. I welcome correspondence at lozae@duq.edu and connections at https://www.linkedin.com/in/emileloza.

This work was presented at the Artificial Intelligence: Thinking about Law, Law Practice, and Legal Education Conference, Duquesne University School of Law in Pittsburgh, Pennsylvania on April 26, 2019.

lozae@duq.edu

As with all digitally transformational technologies, AI systems' technical complexity and rapid adoption in society fast and far outpace the current knowledge and experience of most lawmakers and regulators and operate beyond the envisaged scope of many existing legal doctrines and frameworks. Even if existing law is potentially adequate to address discriminatory algorithmic bias in some domains, the contextual "translation" of those laws to this brave new AI world is indeterminate and uncertain. Further, due to the difficulty of enforcement against illegal algorithmic discrimination, jurisprudential guidance is sparse and slow in coming.

These gaps between the currently limited status of algorithmic law and the realities of the Wild AI West create a void in which unscrupulous or merely unfettered and ambitious algorithm purveyors may pursue enriching, but societally corrosive opportunities. Uninformed persons making decisions about and using AI systems may adopt and deploy such systems without appropriate insight, preparation, or restraint. Further, they may use algorithmic systems in ways contrary to purveyors' guidance, such with recidivism risk systems used in sentencing.

All these factors coalesce to create the potential for discriminatorily-biased algorithms to do exponentially amplified, persistent, and irreparable harm to individuals, communities, and society. The need is urgent for a workable system of algorithmic justice by which to illuminate and eradicate discriminatory computational biases, or at least to more quickly identify them and reduce their incidence and duration. Fostering greater access to justice, public trust in AI technology, and other important policy goals, an algorithmic justice system also would provide empirical mechanisms by which to establish baselines and measure and communicate the status of progress toward eradicating discriminatory algorithm bias in State of the Algorithmic Nation reports, for example. In addition, this algorithmic justice system would provide a framework for crafting meaningful policy, legislative, and regulatory systems for AI and a more accessible and definitive means of enforcing and litigating against illegal algorithmic discrimination.

This work offers a new model toward an algorithmic justice system. As a beginning to address the likely immense and certainly multiply complex problem of discriminatory algorithmic bias, this new model commences with two foundational processes.

First, the model proposes the creation of Algorithmic Justice Standards ("Standards"). These Standards must be created to identify and elaborate what constitutes impermissible discriminatory bias within algorithmic systems. Overtly discriminatory bias in algorithms should be easier to identify and, one hopes, rarely present. As the greater challenge, however, these Standards must identify implicit biases arising from the use of zip codes, "ethnic" or gendered names, and other data that may function, including in otherwise seemingly innocent combinations, as proxies for race and other protected classifications that, in turn, serve as inputs for algorithmic engines. It is reasonably foreseeable that implicit discriminatory bias in algorithmic systems is as rampant as aversive racism and other implicit bias within human society, at least in the United States. Therefore, the Algorithmic Justice Standards must encompass a great numerosity of deeply complex analyses to get at implicit discriminatory

lozae@duq.edu

biases. Interdisciplinary approaches and collaborations are required to carry out these ambitious, critical, ultimately revelatory, and hopefully transformative exercises.

The Standards development and other processes must involve multiple stakeholders from within multiple disciplines and must incorporate legal constructs and ethical requirements. Part of the Standards processes must include consensus-building as to whether and, if so, what levels of discriminatory bias is tolerable, given the context(s) in which algorithmic systems are to be deployed. Above all, the Algorithmic Justice Standards and all associated processes and information must be transparent and openly available and participatory.

As its model's second cornerstone, this work proposes the creation of robust technical testing and validation mechanisms by which each algorithmic implementation will be tested by accredited laboratories and validated by an appropriate government agency against the Algorithmic Justice Standards prior to its deployment in the market or otherwise. Algorithmic systems may constitute highly valuable or sensitive proprietary property or otherwise warrant protection against unfettered disclosure. To encourage participation by algorithm purveyors having such interests, algorithms submitted for testing and validation must be shielded under exceptions to the federal Freedom of Information Act and state equivalents. Rather, such algorithms must be discoverable only through judicial process and, as applicable to trade secrets and other sensitive or restricted access information, subject to protective orders. The results of algorithm testing and validation, however, must be transparent and readily available. Inspectors general and internal auditors must have the technical capacity and, where needed, additional legal authority to critically review testing and validation processes and the institutions that carry them out and to move forward with enforcement or other corrective measures.

In sum, the two cornerstones of the proposed algorithmic justice system build foundational processes, first, to create a largely, but not purely, community-based law in which the community of stakeholders agree, including as circumscribed by existing law, as to what constitute unacceptable discriminatory algorithmic biases; and, second, to evaluate algorithms for compliance against that law before people, their communities, and society are subjected to them.

At this point in the proposed algorithmic justice system model, enter the National Institute of Standards and Technology ("NIST"). NIST is a highly developed and capable technical agency under the auspices of the U.S. Department of Commerce. A number of Nobel Laureates, other leading experts in mathematics, physics, and other scientific disciplines, and a highly experienced staff fill its ranks. Although non-regulatory in the traditional sense, NIST is one of the leading technical public authorities in the world and certainly in the United States. Further to the point, NIST is the leading government authority on cybersecurity standards and on the testing and validation of encryption, or cryptographic, algorithms. NIST's roles around artificial intelligence currently focus on research and participation in ISO's newly-commenced development projects regarding AI trust and security standards.

lozae@duq.edu

As to the model's first cornerstone process, NIST's Cybersecurity Framework ("Framework") provides an excellent example to emulate for the creation and development of Algorithmic Justice Standards. The Cybersecurity Framework presents a detailed collection of technical, best practice, and other standards as to how to achieve and maintain more cyber secure environments and systems. What's more, like the algorithmic systems to be addressed by the Standards, the Framework reaches across multiple and divergent domains and operational contexts. Broad and sustained collaboration between NIST, industries, governments, and consumers and other stakeholders produced the Framework, and the same inclusive scope, level, and duration of collaboration will be required to formulate and continue to develop the Algorithmic Justice Standards.

In no small measure, due to its technical foundations under NIST's leadership and its transparent and collaborative origins, the Framework is one of the world's most powerful and universally accepted examples of soft law within a global technology-driven domain. The Framework's initially voluntary standards have since entered into the positive law to bind federal government agencies and suppliers under the Federal Information Processing Standards and other authorities. Increasingly, the Framework's standards apply to non-government actors under consumer protection, tort, fiduciary, and other legal doctrines and theories.

The proposed Algorithmic Justice Standards should follow the process used to create and continue the development and maturation of the Cybersecurity Framework. As with the Framework, NIST is ideally suited to lead and house this effort, although with respect for and collaboration with other jurisdictionally-relevant federal and other agencies. Among the many subsequent steps to elaborate the proposed algorithmic justice system model, fruitful next steps must include a review of NIST's existing legal authority as to artificial intelligence and analyses as to whether it may require amendment to encompass the development of Algorithmic Justice Standards and also may encompass the consideration of IEEE and other technical ethical standards to determine their place in the law, if any.

Pending legislation, such as the National Institute of Standards and Technology Reauthorization Act of 2018, also provides a statutory point of initiation or expansion where the best practices language in the bill's artificial intelligence section could be clarified or expanded to encompass NIST's leadership toward the new Standards initiative. Even so, the development of Algorithmic Justice Standards must encompass significantly more evaluation and incorporation of existing law, such Title VI of the Civil Rights Act of 1964 and associated jurisprudence, including under the U.S. Supreme Court's *McDonnell Douglas v. Green* framework, to ensure that, for example, what constitutes intentional discrimination is considered and interpreted within relevant AI contexts.

Toward the second cornerstone process of the proposed model, NIST also fields a powerful infrastructure of technical expertise and systems that could be leveraged and harnessed. Under its Cryptographic Algorithm Validation Program and associated programs, NIST has a well-developed network of algorithm testing laboratories, a certification program for those

lozae@duq.edu

laboratories, and a robust process by which algorithms are validated against various technical cryptographic standards. NIST also already has in place a robust and transparent publication process as to the results of such validations, including the companies or other entities submitting each of the algorithmic implementations.

Here, the proposed model rests upon two assumptions that must be tested and confirmed, debunked, or modified. First, it assumes that NIST's cryptographic algorithm infrastructure and human capabilities are capable and suitable for expansion to fulfill similar roles as to the Algorithmic Justice Standards. If that is true, the model next assumes that such an expansion would be effective, efficient, and politically achievable. Certainly, significant foundational work will be required, including technical work such as the aggregation and evaluation of more diverse data sets of, for example, historical housing, credit, and recidivism risk, sentencing, and parole data; and new, challenging, and fascinating domains of legal and policy development and legal and judicial education.

Of necessity, the creation of an algorithmic justice system is an interdisciplinary undertaking with lawyers, policymakers, computer and cognitive scientists, software engineers, statisticians, and others working together. Those collaborations will produce meaningful Algorithmic Justice Standards and a workable means by which to test and validate algorithms as non-discriminatory prior to society's exposure to them. Rich educational, consumer protection, civil rights, and other benefits also will ensue. The elimination of discriminatory algorithmic bias presents one of the most daunting and compelling calls to action of the digital age. By leveraging and building upon NIST's capabilities, the algorithmic justice system model proposed here maps out a practically achievable way forward to prevent and enable enforcement against illegal AI discrimination and to ensure transparency, order, and essential human protections in an increasingly artificially intelligent world.

With this, the model seeks to improve and infuse greater justice in the algorithmic state of the nation and help create a more perfect union.

lozae@duq.edu