ON ADDRESSING ALGORITHMIC DISCRIMINATION THROUGH DEVELOPING RESPONSIBLE ARTIFICIAL INTELLIGENCE CONSORTIUMS

Michael B. Runnels
California State University, Los Angeles

Benjamin Paul Royce
Columbia University

On 25 April 2023, responding to artificial intelligence's (AI) increasing use to make hiring, credit, and housing decisions, regulators across the Biden administration announced a plan to enforce existing civil rights laws against AI systems that perpetuate discrimination (Zakrzewski, 2023). Leaders from the Federal Trade Commission (FTC), Consumer Financial Protection Bureau, the Justice Department, and the Equal Employment Opportunity Center warned the public about the risks of "digital redlining," where faulty data sets and poor design choices develops AI systems that exacerbate ongoing discrimination in the United States economy. Arguing that "AI is being used right now to decide who to hire, who to fire, who gets a loan, who stays in the hospital and who gets sent home" … FTC Commissioner, Alvaro Bedoya, noted that he is "much more worried about those current, real-life uses of AI than potential downstream existential threats (Zakrzewki, 2023)."

POTENTIAL AI-RELATED GAPS IN ENFORCING DISCRIMINATION LAWS

Title VII of the 1964 Civil Rights Act, for example, prohibits discrimination in employment decisions based upon an individual's race, color, sex, or national origin (*Civil Rights Act*, 1964). However, Title VII and other existing U.S. discrimination laws place an evidentiary burden on those suing a U.S. company for discrimination perpetuated by AI that is likely too high to meet. First, to present case of "disparate impact" in employment discrimination under Title VII or in fair lending discrimination under both the Equal Credit Opportunity Act and the Fair Housing Act, the plaintiff must prove that a policy or practice, which appears neutral, has a disproportionately negative impact on an individual based upon their race, color, sex, or national origin. Second, once the plaintiff establishes the factual basis for this discrimination, the burden then shifts to the defendant to provide a "business necessity" defense, which is an argument that the neutral practice creating the disparate impact is an essential practice to the operation of the business (*Civil Rights Act*, 1964). Notably, the U.S. Supreme Court considers the "business necessity" defense as the "touchstone" of a disparate impact case (*Griggs v. Duke Power Co.,* 1971, p. 424), as businesses must be free "to make the practical business choices and profit-related decisions that sustain a vibrant and dynamic free-enterprise system (*Texas Department of Housing & Community Affairs v. Inclusive Communities Project*, 2015, p. 2518)." Third, once the defendant satisfies this defense, the burden then shifts back to the plaintiff to prove that a "less discriminatory" alternative policy or practice that will serve the business's legitimate needs exits (*Texas Department of Housing & Community Affairs v. Inclusive Communities Project*, 2015, p. 2518).

In discrimination suits against businesses using AI, this final prong of the disparate impact analysis will likely prove a difficult burden for a plaintiff to meet, as in order to meet this burden, the plaintiff must prove that (1) there exists an alternative employment practice that is equally valid in meeting the defendant's business necessity; (2) that practice would result in less of a disparate impact; and (3) the defendant refused to adopt this practice (*Civil Rights Act*, 1964; *Jones v. City of Boston*, 2016; *Ricci v. DeStefano*, 2009; *United States v. Brennan*, 2011). Consistent with the business necessity defense, the proposed less discriminatory AI must not be prohibitively costly to implement and must not impose other unreasonable burdens on the employer, such as using alternative data sources that may be even more problematic from a consumer privacy or data security perspective or other technical issues that make the less discriminatory AI difficult to operate. Accordingly, the technical skill and costs likely associated with a plaintiff demonstrating or creating this less discriminatory AI will make this evidentiary burden difficult for a plaintiff to meet. For example, the difficulty in determining how an AI algorithm reached the result it implemented might frustrate the determination of causation, as the number of possible causes, and the identity of the data points and data owners involved, may be so large as to create qualitatively different problems in ferreting out the causes of discrimination than in the past (Burris, 2022; Lederer, 2020; Metz, 2018; McClellan, 2016). For these reasons, current U.S. discrimination laws may ironically hinder a successful discrimination lawsuit against a company utilizing AI. Developing a socially responsible AI, however, can help fill this gap by ensuring that the new age of AI operates in a manner that is consistent with the spirit of U.S. discrimination laws.

ADDRESSING ALGORITHMIC DISCRIMINATION THROUGH NEW GOVERNANCE APPROACHES

"New Governance" is a field of corporate social responsibility (CSR) that grew out of an appreciation of the CSR movement as being fundamentally flawed in its approach to incentivizing responsible corporate behavior (Runnels et al., 2010). Part of this appreciation stems from a traditional debate between those arguing that corporations should do good for goodness's sake and those viewing such arguments as being hopelessly naïve if not tethered to corporate quarterly earnings reports. New Governance approaches often consider the modern CSR movement as little more than an elaborate public relations performance whereby corporations undertake performative social rituals while continuing to conduct business as usual (Runnels et al., 2010). Rather than emphasizing checklist-style compliance with prescriptive laws that often incentivize corporations to find and abuse loopholes, New Governance approaches encourage groups to orient themselves towards the spirit of those laws by continually revising both means and ends to solve problems as they arise in real time. In this way, New Governance approaches envision corporate decision-making as a collaborative, rather than an adversarial process that "provides a rational, systemic alternative to draconian rulemaking and [its] often adverse effects on business (Ford, 2008, p. 45)."

One New Governance approach recasts the conception of the socially responsible corporation as one operating at the nexus of certain situational factors, or "modes of social control" (Epstein, 2007). Those modes are law, consortium regulation, self-regulation, ethical precepts, the media, and an engaged civil society. In brief, law is the articulation of public policy enforced by government. Consortium regulation refers to standards of behavior established by members of a particular profession, such as medicine. Self-regulation regards voluntary adherence to standards set by nongovernmental entities concerned with specific issues, such as climate change. The expectation with self-regulation is that corporations will comply with standards voluntarily, and in

good faith. Ethical precepts are beliefs derived from religion, humanistic philosophy, social customs, mores, and traditions that often inform or inspires laws. A vigilant and responsible media responds to information about corporate misbehavior by providing material that brings corporate behavior to light. An engaged civil society refers to direct citizens' action through applying direct political pressure on government officials.

These modes, used individually or together, operate as a practical framework that allows business leaders to engage in a systematic reconsideration of ways to incentivize responsible corporate behavior (Runnels et al., 2010). Indeed, while law has traditionally served as the centerpiece for incentivizing this behavior, some increasingly recognize that corporate misbehavior is highly context-specific, e.g., that in each situation, such as enforcing discrimination laws in corporations using AI, some modes may work better than others in encouraging socially responsible corporate behavior. Moreover, and in particular regards to the technology industry, the creation of law is an often extended process where laws created for companies operating in fast-paced business environments often arrive as dated answers to yesterday's problems.

CONCLUSION: A RESPONSIBLE AI CONSORTIUM AS A NEW GOVERNANCE SOLUTION

New Governance approaches emphasize the importance of matching the appropriate mode of social control with the specific challenges facing business. Rather than focusing on overarching regulation and adversarial enforcement, New Governance approaches consider ongoing deliberation between those who have access to local information and context-specific understanding of business problems as the most effective mechanism for making decisions in complex fast-paced business environments (Runnels, 2011; Runnels & Burton, 2012; Runnels & Giampetro-Meyer, 2011). In this way, the CSR movement alone is fundamentally incapable of incentivizing socially responsible corporations. This argument is not based on some belief in an inherent malevolence of the corporate form, but rather, on a recognition of the incentives that form the modern corporate environment, which are incentives that corporations must abide by (Runnels et al., 2010). An increasingly competitive global economy, driven by legal duty to maximize shareholder profits every quarter, creates environments where companies using AI are incentivized to hoard their data. Such data hoarding in the exponentially evolving world of AI is problematic, since companies are less likely to share potentially sensitive training data, particularly involving fields such as customer information, sales, or proprietary metadata that could be critical in evaluating their potential use for addressing algorithmic discrimination. This leads to a fundamental paradox whereby stakeholders expect corporations to produce ever increasing efficiencies through AI while the market incentivizes the hoarding of the very training data within which a potential solution to the problem of algorithmic discrimination resides.

A Responsible AI Consortium (RAIC) is one solution to this data hoarding paradox. Here, RAIC regulation is defined as standards of behavior established by members of a particular industry, such as medicine. Indeed, associations like the American Medical Association establish standards of behavior for their members, impose those standards on their own industry, and do so voluntarily for the good of both their industry and customers. A RAIC combines the benefits of collaboration with a competitive business environment and incentivizes AI operating companies to be proactive in complying with U.S. discrimination laws by establishing formal non-governmental organizations at the industry level that facilitates two efforts:

1. A working data exchange to help competing companies share non-personal, aggregate information about their AI models and where they are either not performing or are perpetuating discrimination. This allows companies to learn from their competitors' mistakes and prevent discrimination at a much larger scale than any potentially successful discrimination lawsuit can.
2. Manage and standardize which datasets should be used and in which ways that are specific to their industry/sector and applicable law(s). An industry standard schema of data for training AI models is a significant first step in systemically ferreting out and preventing algorithmic discrimination.

Such a proposed RAIC would take the form of an industry level consortium that validates and inspects data that may contain inherent bias, which is then trained into AI systems in a way that may not reveal itself until after the training. Though not an ideal fix, if competing companies can learn from each other, without compromising either their proprietary data or competitive edge, they will be able to identify and mitigate disparate impacts before the results of any successful lawsuit.

The case of using ZIP codes in this context is illustrative, as ZIP codes have been standardized by the US Postal Service since 1963. They are in every consumer facing company's data and make it easy for communications, deliveries, and customer personalization. ZIP codes, however, are often at the root of the discriminatory effect of racial profiling and race-based automated decision making (National Fair Housing Alliance, 2014). Simple solutions, such as reducing the ZIP code from five digits to three, can remove much systemic bias in the data. Indeed, why use five digits when one can use only three? Since delivery addresses and billing addresses require all five digits, instances of discrimination can be prevented by using all five digits only in cases of delivery. This could be a standard whereby the industry based RAIC provides industry-level guidance without compromising their data or competitive edge. If there were a non-profit and industry specific organization for retail services that could provide clean and nearly bias-free data sets for industry members to contribute to and utilize, companies could prevent the customer segmentation models that inadvertently prioritize one racial group over another. Such an approach can substantively address many discriminatory concerns before they arise - and save duplication of efforts across multiple companies.

A RAIC in each industry can be the working model that business leaders need to address algorithmic discrimination in a manner much faster than a U.S. legal system attempting to catch up to the exponentially evolving world of AI. New Governance approaches consider decision-making as a collaborative, rather than an adversarial process. Indeed, the RAIC concept suggested by this article is emblematic of the New Governance movement, as it represents a systematic process that uses "innovative, pragmatic, information-based, iterative, and dialogic mechanisms to gather, distill, and leverage industry learning (Ford, 2008, p. 5)" for the purpose of addressing complex corporate challenges. Finally, the RAIC concept is also responsive to recent guidance from the Biden administration, namely that "[w]hile automated systems have the capacity to drive extraordinary advances and innovations, algorithmic discrimination protections should be built into their design, deployment, and ongoing use (The White House, 2022)." For these reasons, industry specific RAICs may present a timely and impactful mechanism through which to address algorithmic discrimination.

REFERENCES

Burris, K. (2022). A Title VII dead end? Machine learning and employee monitoring. *Wm. & Mary L. Rev. Online*, 63, 91.

Civil Rights Act of 1964 § 7, 42 U.S.C. § 2000e et seq (1964).

Epstein, E. M. (2007). The good company: Rhetoric or reality-corporate social responsibility and business ethics redux. *Am. Bus. LJ*, *44*, 207.

Ford, C. L. (2008). New governance, compliance, and principles-based securities regulation. *Am. Bus. LJ*, *45*, 1.

Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

Jones v. City of Boston, 845 F.3d 28 (2016).

Lederer, F. I. (2020). Here there be dragons: The likely interaction of judges with the artificial intelligence ecosystem. *Judges J.*, *59*, 12.

McLellan, C. (2016, December 1). Inside the black box: Understanding AI decision-making. *ZDNet.* Retrieved from https://www.zdnet.com/article/inside-the-black-box-understanding-ai-decision-making.

Metz, C. (2018, March 6). Google researchers are learning how machines learn. *New York Times*, *16*. Retrieved from https://www.nytimes.com/2018/03/06/technology/google-artificial-intelligence.html.

National Fair Housing Alliance (2014, August 27). *Zip code inequality: Discrimination by banks in the maintenance of homes in neighborhoods of color*. Retrieved from https://nationalfairhousing.org/wpcontent/uploads/2022/02/2014_08_27_NFHA_REO_report.pdf

Ricci v. DeStefano 557 U.S. 557 (2009).

Runnels, M. B. (2011). Dispute resolution and new governance: Role of the corporate apology. *Seattle UL Rev.*, 34, 481.

Runnels, M. B., Burton, A. M. (2012). The Foreign Corrupt Practices Act and new governance: Incentivizing ethical foreign direct investments in China and other emerging economies. *Cardozo L. Rev.*, 34, 295.

Runnels, M. B., Giampetro-Meyer, A. (2011). Cooperative NRDA & new governance: Getting to restoration in the Hudson River, the Gulf of Mexico, and beyond. *Brook. L. Rev.*, 77, 107.

Runnels, M. B., Kennedy, E. J., Brown, T. B. (2010). Corporate social responsibility and the new governance: In search of Epstein's good company in the employment context. *Akron L. Rev.*, 43, 501.

The White House. (2022). *Blueprint for an AI bill of rights*. Retrieved from https://www.whitehouse.gov/ostp/ai-bill-of-rights/

United States v. Brennan, 650 F.3d 65 (2011)

Zakrzewski, C. (2023, April 25). Federal regulators call AI discrimination a 'New Civil Rights Frontier.' *Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2023/04/25/artificial-intelligence-bias-eeoc/