ALGORITHMIC BIAS AND ACCOUNTABILITY: THE DOUBLE B(L)IND FOR MARGINALIZED JOB APPLICANTS

CHRIS CHAMBERS GOODMAN[*]

INTRODUCTION

"Help Wanted. Experience Necessary. Apply Online." The one-click nature of online employment applications means that organizations receive an exponentially higher number of résumés than ever before.[1] While the wealth of résumés creates opportunities for a broader and deeper exploration of the available talent pools for each position, the sheer numbers can overwhelm any *human*, human-resources professional. The increasing abilities of artificial intelligence (AI) technologies, including machine learning and large language models, provide effective mechanisms for funneling applicants into tranches of "qualified" or "highly qualified" and separating out the un- or under-qualified. Because of the economic practicalities, employers increasingly use AI technologies to assist with or make hiring decisions. Employers expect these technologies to quickly and accurately assess candidates' merit.

In addition to the economic savings of using AI technologies, some employers believe that AI also promotes greater fairness and nondiscrimination in the hiring process. However, it is important for employers to understand that some biases are imbedded in AI technologies, and failing to account for those biases can result in large-scale hiring discrimination. In fact, the Equal Employment Opportunity Commission (EEOC) stepped in

---

    1. Alyssa Lankford, *Artificial Intelligence Use Continues to Rise in Employment*, CAL. EMP. L. LETTER, Jan. 26, 2024.

to issue guidance to all employers in May 2023,[2] which is discussed below.[3]

The Biden Administration was also concerned with biases lurking in AI technologies. On October 23, 2023, President Biden issued an executive order ("2023 AI EO"), which aimed at accelerating the federal government's efforts to establish standards for AI development, use, safety, and security.[4] In response, the Office of Management and Budget (OMB) issued Draft Guidance in March 2024 applicable to all federal government agencies using AI technologies that impact rights or safety.[5] The OMB Draft Guidance requires federal agencies to conduct an annual inventory of all AI use cases, identify which are rights- or safety-impacting, and detail practices to mitigate risk for those uses.[6] The OMB also issued a memorandum with additional details about rights- and safety-impacting AI uses by federal agencies and established new requirements for risk management and governance of AI technologies, including designating an AI Officer and developing plans for managing risk in the face of innovation.[7]

Another important concern is the clash between fairness and privacy. This Article uses the term "privacy," as defined by Gupta et al., as a "shorthand to refer to informational privacy, namely the right of individuals to have a meaningful say in the way data about them is collected, stored, and used."[8]

---

2. *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964*, U.S. EQUAL EMP. OPPORTUNITY COMM'N (May 18, 2023) [hereinafter *EEOC Guidance*], https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial [https://perma.cc/HJ65-ZB3Y].

3. *See infra* Section II.A.

4. Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Oct. 30, 2023) [hereinafter 2023 AI EO].

5. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, DRAFT GUIDANCE FOR 2024 AGENCY ARTIFICIAL INTELLIGENCE REPORTING PER EO 14110 (Mar. 28, 2024) [hereinafter AI DRAFT GUIDANCE], https://insideaipolicy.com/sites/insideaipolicy.com/files/documents/2024/mar/ai03272024_4.pdf [https://perma.cc/9ZFW-3EM4].

6. *Id.*

7. OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, OMB MEMORANDUM NO. M-24-10, ADVANCING GOVERNANCE, INNOVATION, AND RISK MANAGEMENT FOR AGENCY USE OF ARTIFICIAL INTELLIGENCE (Mar. 28, 2024) [hereinafter OMB MEMORANDUM NO. M-24-10].

8. Arushi Gupta et al., *The Privacy Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government*, 2023 PROC. OF THE ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 492.

Additionally, it recognizes that there are other aspects, like the right to be left alone.[9]

Privacy is (still) a right. Fair treatment is a right. Violations of these rights can produce emotional, physical, and economic harms. Harm can also result from conflicts among these rights. For instance, when employers use protected characteristics to screen out applicants, such as height and weight for prison guards, disparate impacts are clearer: employers see fewer successful female applicants. With machine learning processes, however, we do not know the basis for screening; algorithms analyze hundreds of thousands of data points when determining which résumés are passed along for further review. Compelling applicants to submit to AI screening invades privacy by requiring disclosures that, while appearing facially neutral, can actually have a substantial discriminatory impact.

Consider an example of an applicant screening tool designed to determine which employees are most likely to be productive and have exemplary attendance records. One way to predict future attendance could be to measure how often the employee stays away from work because of illness or other health reasons. In the United States, it could violate the Americans with Disabilities Act if an employer refused to hire individuals with illnesses, or it could violate the Pregnancy Discrimination Act on the grounds that those individuals would be more likely to use more sick days. Programmers can program the algorithm to avoid screening out based on characteristics that would violate these federal laws, but that does not mean that the algorithm will not use information about illness or potential pregnancy in making its determination. The machine learning process, after evaluating large amounts of data, is very likely to find proxy factors to consider that have the effect of screening people out based on their health or likelihood of pregnancy or pregnancy complications. These proxies for poor attendance or bad health may be indiscernible to the employer using the technology, yet result in real harm.

Instead, we might test for fairness of process, which we can measure at each stage—from recruiting, to application consideration, to interviewing and callback consideration—without disclosing private information about outcomes. Under this sort of "process-defect theory," processes that are not fair to

---

9.   *Id.* at 493–94.

all violate the fair treatment principle.[10] With AI technologies in the hiring arena, processes are opaque and often outright inscrutable, and thus it is virtually impossible to evaluate fairness in this way.

Implementing AI in hiring processes can inadvertently lead to issues related to discrimination and privacy. This Article explores the *double bind* that impacts people of color and others with nontraditional or nonstandard backgrounds in job-seeking spaces dominated by AI technologies. To alleviate concerns about fairness by permitting auditing of outcomes, this Article argues applicants must not only disclose but also agree to retention of private information. Without such disclosure and retention of data, the machines may continue to learn in ways that exacerbate, rather than alleviate, biases in hiring.

Based on remarks made by legal scholars at the 2024 Rothgerber Conference: AI and the Constitution at the University of Colorado Law School on Friday, April 19, 2024, during the AI and Privacy panel,[11] this Article proceeds as follows: Part I provides some background on how employers are using AI technologies and highlights concerns with AI-assisted employment processes. Part II describes efforts by the executive branch to regulate AI systems and some of the limits in the federal arena. Part III highlights recent state and local attempts to regulate AI for the first time and focuses on a 2023 New York City ordinance. From there, it explores ways to build upon and improve that start. Part IV concludes the Article with additional recommendations for *double-blinding* data to optimize the balance between privacy and fairness. For instance, organizations that use AI tools to sort and hire job seekers should consider conducting an ethical risk assessment followed by a bias risk assessment. These organizations should then

---

10.   *See generally* JOHN HART ELY, DEMOCRACY AND DISTRUST (1980).

11.   The panel description was as follows: "This panel explores the challenges of protecting the right to privacy in the context of the explosion of AI. The discussion will range from how privacy can and should be protected from a wide angle lens to more granular assessments. More specifically the panelists will consider: the challenge of defining and protecting 'sensitive' information; the need for data privacy protections tailored to marginalized groups to guard against exploitation, oversurveillance and political deception; . . . and the efficacy of groundbreaking local laws that require impact assessments for algorithms used in hiring decisions." The author is grateful to the engaging conversations with her fellow panelists: Paul Ohm, Spencer Overton, and Scott Skinner-Thompson. *2024 Rothgerber Conference: AI and the Constitution*, SILICON FLATIRONS, https://siliconflatirons.org/events/rothgerber-conference-ai-and-the-constitution [https://perma.cc/PL8C-LH46].

make any needed adjustment before deploying AI in their hiring processes. This assessment should be an iterative process to ensure that the AI tool does not perpetuate bias and prevent the organization from hiring diverse candidates.

I.   HOW USING AI IN HIRING CAN REINFORCE HUMAN BIASES

A.   How Is AI Being Used in Hiring?

The hiring process moves from sourcing to screening to interviewing to selection to evaluation.[12] Job descriptions, advertising, matching, and headhunting are all important parts of the widest part of the funnel: sourcing.[13] Some of the ways that AI and machine learning technologies are used in hiring include (1) creating position ads,[14] (2) disseminating ads,[15] (3) culling/screening résumés,[16] (4) evaluating personality and other screening tests,[17] and (5) video interview assessment and screening.[18] AI tools can help with evaluating the level of applicants' substantive knowledge, their adaptability, and their likelihood of accepting job offers or staying in a job for an acceptable period of time. In addition, there are tools that analyze the likelihood of the employee engaging in sexual harassment.[19] Employers like Amazon and Google use these AI tools through social media platforms such as LinkedIn, targeted ads based on interests and preferences, and profiles on other

---

12.   *See* MIRANDA BOGEN & AARON RIEKE, HELP WANTED: AN EXAMINATION OF HIRING ALGORITHMS, EQUITY, AND BIAS 13–14 (2018) (exploring "how new predictive hiring tools are being used in each stage, describing and analyzing illustrative products on the market today").

13.   *Id.*

14.   Lankford, *supra* note 1.

15.   *Id.*

16.   Gary D. Friedman & Thomas McCarthy, *Employment Law Red Flags in the Use of Artificial Intelligence in Hiring*, ABA BUS. L. TODAY (Oct. 1, 2020), https://www.americanbar.org/groups/business_law/resources/business-law-today/2020-october/employment-law-red-flags [https://perma.cc/UR8C-QMTV].

17.   *Id.*

18.   *Id.*

19.   *See generally* ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE 57–59, 60, 106 (2022).

social media.[20] Recruiters seek to ensure the candidates that they think would be good fits see their advertisements.[21]

It may not be surprising that job descriptions containing more masculine language or tone will produce more male applicants. Similarly, digital advertising that is targeted based on social media and other tools effectively forecloses some people from being informed about open positions, particularly those who are different in various ways from existing employees.[22] Recommendations, like Zip Recruiter's matching system, also play a role in reducing how many potential applicants enter the pipeline, thus further decreasing the funnel capacity.[23]

More and more employers use assessments during online application processes and then use technology to evaluate these assessments.[24] After the initial application, many companies deploy online assessment tools to score and rank candidates using their résumés and tests. At the next stage, employers use analytic tools to conduct an algorithmic evaluation of video interviews. Often, employers do not stop once the employee is hired and continue to use AI tools during the employment relationship—tools that impact retention, promotion, and thus salaries.[25] For instance, companies require workers to install tracking software on their cell phones to monitor their specific locations when working from home or off-site. Employers also have cameras take screenshots of the employees' computers to document whether the screen shows work-related material and photos of the employees' faces to confirm that they are at the computer.[26] Algorithms also track productivity (e.g., how long

20. Candi Castleberry et al., *How Amazon Leverages AI and to Enhance the Hiring Experience for Candidates* (June 5, 2023), https://www.aboutamazon.com/news/workplace/how-amazon-leverages-ai-and-ml-to-enhance-the-hiring-experience-for-candidates [https://perma.cc/4VDU-KWBC].

21. Lindsey Fuchs, Note, *Hired by a Machine: Can a New York City Law Enforce Algorithmic Fairness in Hiring Practices?*, 28 FORDHAM J. CORP. & FIN. L. 185, 193–94 (2023) (describing how employers use algorithmic tools for recruitment and hiring, and then addressing some of the privacy and discrimination issues that arise from the tools, gaps in enforcement, and how to fill some of those gaps).

22. BOGEN & RIEKE, *supra* note 12, at 17–19.

23. *Id.* at 19–20.

24. *Id.* at 30–36 (illustrating some of the ways that historical and social patterns may not act thoroughly or positively react to applicants of different backgrounds).

25. Fuchs, *supra* note 21, at 196.

26. Jodi Kantor & Arya Sundaram, *The Rise of the Worker Productivity Score*, N.Y. TIMES (Aug. 14, 2022), https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html [https://perma.cc/DBU8-SAZB].

does it take employees to read or edit a report, how long or how frequently do they take breaks, and how often they scroll back to re-read or double-check something) and provide "grades" on employee performance.[27]

B. *What are the Main Concerns About Using AI in Hiring?*

This Section discusses some of the primary areas of concern with AI employment decision-making. There are many reasons for employers and potential employees to carefully and conscientiously approach any consideration of AI technologies and to understand the effects of predictive tools on the hiring process.[28] While many workers are concerned about the proliferation of AI tools in the workplace and oppose their use for final hiring decisions, many also believe that AI is better than humans at evaluating applicants.[29] One consideration is that "[i]ncreasingly, workers are being called upon to exchange their privacy for the mere opportunity to be considered for employment."[30] In the past, employers commonly called references, conducted background checks, and sometimes reviewed credit reports. Current due diligence practices are far more intrusive. Potential employees are undermining their privacy rights by divulging a significant amount of sensitive personal information to employers who may be using AI to inappropriately screen them. Some divulged data may not even be particularly relevant to whether they can do the job—like their private social media posts, which may include information about everything from restaurant reviews, dating advice, health challenges, and travel locations.[31] These potential employees are sacrificing their right to be left alone in order to further economic outcomes, which is even more troubling when the

---

27. *Id.*

28. BOGEN & RIEKE, *supra* note 12, at 1–2 (providing a detailed overview of how employers use AI hiring tools, several cautions about bias-reducing and bias-enhancing aspects of AI use, and recommendations on transparency, government regulation, and process validation).

29. Lee Rainie et al., *AI in Hiring and Evaluating Workers: What Americans Think*, PEW RSCH. CTR. (Apr. 20, 2023), https://www.pewresearch.org/internet/2023/04/20/ai-in-hiring-and-evaluating-workers-what-americans-think [https://perma.cc/TU9A-DSXH].

30. Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 628 (2021) [hereinafter Ajunwa, *Auditing Imperative*] (citing Ifeoma Ajunwa et al., *Limitless Worker Surveillance*, 105 CALIF. L. REV. 735, 736 (2017)).

31. *See id.* at 681.

employer declines to hire them while retaining their personal data.

It is crucial to recognize that hiring is a "cumulative series of small decisions," each of which can be influenced by algorithmic tools.[32] Although AI systems can "reduce interpersonal bias," deleting data on characteristics (such as race or gender) does not necessarily alter its reliance on "institutional and systemic biases."[33] For instance, some AI companies offer "social media background checks" that analyze their social media activity to predict which candidates might be most at risk of becoming problem employees, like bullies or harassers.[34] Another tool provides a sliding scale of salaries, benefits like personal and sick time, retirement contributions, and stock options, as well as other perks, so employers can try to make the most attractive offer package as possible for each desired candidate.[35] Tools that predict the likelihood of a candidate accepting a job offer, particularly if they are based on salary parameters, can undermine recent laws that prohibit past salary information inquiries.[36]

Professor Ajunwa, an award-winning professor who focuses on the privacy and discrimination implications of automated decision-making technologies, identifies four problems with automated hiring systems:

(1) . . . culling systems that discreetly eliminate applicants from protected categories without retaining a record;

(2) automated hiring systems that allow for the deployment of proxies for protected categories . . .

(3) intellectual property law, specifically trade secret, protects automated hiring systems from outside scrutiny and allows discrimination to go undetected; and

(4) a worker's lack of control over the portability of applicant data captured by automated hiring systems increases the chance of repeated employment discrimination, thus

---

32. BOGEN & RIEKE, *supra* note 12, at 1.
33. *Id.*
34. *Id.* at 39–40.
35. *Id.* at 41.
36. *Id.*

raising the specter of an algorithmically permanently excluded class of job applicants.[37]

The fourth problem is implicated when data retention practices interfere with re-applications over time and the right (in European Union (EU) jurisdictions though not recognized in the United States) to have data erased, and in essence, forgotten.[38]

The next four Sections will expand on each of Professor Ajunwa's four problems with AI hiring systems in more depth, highlighting real-world examples and practical consequences of using AI in hiring.

1. Discrete Elimination from the Applicant Pool

The first problem Professor Ajunwa addresses—using AI to cull résumés of people in protected classes—is particularly harmful, because humans comparing the rejected applicants to those that remain in the process often cannot discern a basis for the AI decisions beyond facial characteristics like race and gender. For those applicants moving forward in the process, a human will usually be involved at some stage, such as interviewing. However, those who are rejected rarely get a human look. While the tools are usually an assistant to a human decision-maker in the "yes" cases, "they often automate rejections."[39] As cybersecurity expert David Gewirtz notes, the three V's of "volume, velocity, and variety" dwarf any bias that an individual could perpetuate.[40] For instance, online personality tests exacerbate disability-based discrimination, as they have a disparate impact on those with learning challenges,

---

37.   Ajunwa, *Auditing Imperative*, *supra* note 30, at 629–30.

38.   *See generally* Ben Wolford, *Everything You Need to Know About the "Right to be Forgotten"*, GDPR.EU, https://gdpr.eu/right-to-be-forgotten [https://perma.cc/7C5W-N8SP] (discussing the existence and application of the right to be forgotten under the General Data Protection Regulation relating to personal data).

39.   BOGEN & RIEKE, *supra* note 12, at 1.

40.   Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1679 (2020) [hereinafter Ajunwa, *Paradox*] (quoting David Gewirtz, *Volume, Velocity, and Variety: Understanding the Three V's of Big Data*, ZDNET, https://www.zdnet.com/article/volume-velocity-and-variety-understanding-the-three-vs-of-big-data [https://perma.cc/L3MU-WM95] (last updated Mar. 21, 2018, 7:47 AM)) ("[D]ue to the 'volume, velocity, and variety' of data used in automated hiring, any bias introduced in the system will be magnified and multiplied greatly dwarfing the impact of any prejudice held by any one human manager.").

mental disabilities, and those who are neurodiverse. Similarly, other tests can reveal age biases.[41]

On the issue of volume, consider an example from Amazon, which deployed algorithms to cull résumés. The AI was trained on applicant résumés submitted to Amazon over a ten-year period (2005–2015) and rated applicants with one to five stars, much like Amazon's products.[42] Based on the prior résumés, which were overwhelmingly male, the AI trained itself to downgrade references to women's colleges and extracurricular activities, like women's basketball or field hockey.[43] The AI also evaluated masculine language in résumés, such as the verb "executed," more positively than more feminine language like "collaborated."[44] In fact, the AI basically disregarded programming abilities even for engineering jobs, because all the résumés had programming experience.[45] As a result, male résumés advanced when otherwise qualified female résumés did not. In other words, because the volume of female résumés in the training data was lower, ultimate hiring recommendations by AI were skewed to prefer men. When (mis-)used in this way, AI undermines any expansion of the notion of who belongs, as conformity and assimilation are more heavily valued.[46]

Turning to velocity—the speed at which algorithmic decisions are made—AI hiring tools make it easy for employers to reject large numbers of applicants automatically. They can do so while disregarding—or even making decisions *because of*—their protected status. Because so many large companies use the same or similar résumé-screening systems, those systems may rely upon the same databases, especially since many companies do not have the resources to build up their own databases.[47]

As to the issue of variety, relying upon the same databases and underlying information decreases the variety of inputs and

---

41.   Fuchs, *supra* note 21, at 200.

42.   Jeffrey Dastin, *Insight—Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, REUTERS (Oct. 10, 2018, 6:50 PM), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG   [https://perma.cc/S3YJ-WFD7].

43.   *Id.*

44.   *Id.*

45.   *Id.*

46.   Ajunwa, *Paradox*, *supra* note 40, at 1714.

47.   Shomik Jain et al., *Algorithmic Pluralism: A Structural Approach to Equal Opportunity*, 2024 PROC. OF THE ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 197, 204.

increases the variety of circumstances under which discriminatory algorithms can operate unchecked.[48] As explained above, AI can review hundreds of thousands of data points, and those data points will be in different formats, such as emails, photos, videos, and PDFs.[49] Photos and videos can reveal race and gender more readily than emails and PDFs. It is this variety of data AI uses that opens up additional concerns where unchecked AI can rely upon race, gender, or other protected characteristics.

2. Protected Category Proxies

In her second category, Professor Ajunwa notes that speech and facial expression recognition AI programs trained on unrepresentative data may devalue Black, Indigenous, and other people of color.[50] For instance, speech patterns can vary across cultures, and preferred speech patterns—those that the AI will identify as a net "positive"—are likely to be more closely associated with the dominant group.[51] Speech recognition software may perform differently depending on what type of speech it is trained on, and algorithms cannot understand nuances in meaning, tone, and facial expression.[52] Scientific studies have not legitimized the use of differences in interpreting facial expressions in the employment context, making it a problematic practice.[53] Nevertheless, a properly trained algorithm can provide useful information in certain employment contexts.

Relying upon preferred speech and expression patterns can turn those patterns into substitutes, or proxies, for protected categories. These proxies can lead to automated hiring systems perpetuating the status quo. This perpetuation of the status quo operates through a "closed loop." A closed loop is where an algorithm creates advertisements, screens applicants, evaluates them, and determines who to hire and when it is successful. The

---

48. *Id.* (noting that "over 60% of Fortune 100 companies and 8 of the top 10 largest U.S. Federal agencies use the same résumé-screening service for hiring").

49. Gewirtz, *supra* note 40.

50. Ajunwa, *Auditing Imperative*, *supra* note 30, at 636–37 (discussing representation in data sets and concerns about automation bias as potential areas of misuse).

51. *Id.* at 637–40 (providing detail on automated video interviews and the analyses done by AI).

52. *Id.* at 637.

53. *Id.* at 637–38.

algorithm creates the advertisement, which it uses for résumé sorting, leading to automated evaluation of candidates.[54]

"Automated onboarding" becomes part of this loop when the only candidates who make it through each step of the automated recruiting process are those with the characteristics the algorithm was designed to find. The result, Professor Ajunwa cautions, is that "the use of machine learning algorithms in decision-making hiring processes represents a particularly sensitive legal issue because of the potential to create or exacerbate economic inequality."[55]

One particularly pernicious result of such a closed loop is the way that the machine learning process attempts to approximate which résumés demonstrate "cultural fit."[56] Cultural fit can become a strict algorithmic rule, rather than the "amorphous concept"[57] that it is when humans apply it. A related question, of course, is whether these hiring processes actually produce the best employees.[58] For example, employers might be better off with chameleons—those who can quickly adapt to the corporate culture or climate—changing as needed to keep up. Being flexible and able to rapidly adjust to new cultural environments while maintaining individuality may actually be a better proxy for long-term hiring.[59]

3. The Lack of Transparency

The third problem Professor Ajunwa identifies involves the lack of transparency, which hampers accountability for decision-making. Transparency refers to disclosures and openness about processes and how they operate; what data is used; and, to the extent possible, *how* particular decisions are

---

54. Ajunwa, *Paradox*, *supra* note 40, at 1695.

55. *Id.* at 1695–96.

56. *Id.* at 1713.

57. *Id.*

58. A peer-reviewed study concluded "that the capacity to change and flexibility—that is, high 'enculturability'—were more important than pre-existing cultural fit in regard to long-term success." *Id.* at 1719–20 (citing Amir Goldberg et al., *Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness*, 81 AM. SOCIO. REV. 1190 (2016)).

59. *Id.* at 1718 nn.255–256. The *Paradox* article concludes with a proposal to make direct discrimination *per se* actionable under Title VII, which would provide additional protections and lessen the burden of proof on discrimination plaintiffs. *Id.* at 1726–34. Professor Ajunwa models this after the doctrine of negligence *per se*. *Id.* The author explains that this concern is a legal, not a technical, problem. *Id.* at 1707–08.

made.[60] While awareness through transparency is important, data *retention*, more so than mere disclosure, is also necessary.[61] Nevertheless, transparency alone is not an adequate response because there still may not be enough information about *why* a particular decision was made.[62] People rejected for positions or admissions want to know why in large part because they want to make sure they were treated fairly in the process. They may ask themselves questions like: Was there a clerical error? Did a person ever even see my résumé? Was I sorted out for an arbitrary reason? Without transparency, these questions remain unanswered.[63] These concerns form the basis for procedural due process arguments, which are beyond the scope of this Article and have been addressed in this author's previous work.[64]

Some scholars caution that greater transparency about how algorithms operate will not solve problems with discrimination. For instance, Professor Pauline Kim advocates for using technological tools, rather than disclosures about them, to

---

60. Gianclaudio Malgieri & Frank A. Pasquale, *From Transparency to Justification: Toward Ex Ante Accountability for AI* 11 (Brussels Priv. Hub, Working Paper No. 33, 2022) ("As regards transparency justification, the data controller should prove that the algorithmic processing is legible in the sense that, at least, meaningful information about the logic, the significance and envisaged consequences of the decision-making are communicated to the subject at the beginning of the data processing and, upon request, after the processing has started.") (internal citations omitted).

61. Professor Ajunwa explains why awareness of this underlying information is important, noting that "there is a necessity for compulsory data retention by employers making use of automated hiring systems and that, furthermore, such data retention should facilitate both mandated and voluntary audits." Ajunwa, *Auditing Imperative*, *supra* note 30, at 646.

62. Professor Ajunwa cites Joshua Kroll for suggesting randomness as a way to test. *Id.* at 643.

63. Consider Professor Ajunwa's explanation that "this human need for 'intuitive understanding' is a desire for justice, rather than a quest for technical redress. There is both a human need to understand the factors under which one is judged (especially for access to livelihood) and a desire to see factors done away with that do not conform to principles of fairness." *Id.* Professor Ajunwa also proposes a new cause of action which she calls "*discrimination per se*, which takes into account the particular difficulties of proof presented when a plaintiff seeking to challenge an employer's use of an automated hiring system for employment discrimination," which would be a third cause of action under Title VII after disparate impact and intentional discrimination. *Id.* at 649.

64. *See generally* Chris Chambers Goodman, *AI, Can You Hear Me? Promoting Procedural Due Process in Government Use of Artificial Intelligence Technologies*, 28 RICH. J.L. & TECH. 700 (2022).

reduce algorithmic discrimination.[65] Her article *Auditing Algorithms for Discrimination* describes the limits of technological responses and available technical tools,[66] explaining that "[t]echnical tools alone cannot reliably prevent discriminatory outcomes because the causes of bias often lie not in the code, but in broader social processes."[67] Regardless of the level of transparency in the process and attention to diversifying training data sets, "technical tools cannot guarantee that algorithms will not discriminate because bias may result from social processes that lie outside the code."[68]

4. The Individual Risk of Permanent Algorithmic Exclusion

As Professor Ajunwa explains, her fourth problem is data retention. When AI generates decisions, such as which applicants to reject, it creates a record about that applicant and about that decision. It creates the record after the résumé culling. Recall that bias in résumé culling is the first problem discussed in Section I.B.1. The culling and data retention together implicate the issue of algorithmic blackballing.[69] If an applicant applies again, even for a different position years after gaining more relevant experience, that applicant still may be rejected based on the data about that past rejection.[70] For instance, a person who scores very low on a videogame-style test (a test that requires eye-hand coordination between viewing images or words on a computer screen on the one hand, and engaging in physical movements with a joystick, keyboard, mouse, or other device) would be rejected. Later, if that person practiced and gained skills to better perform on that style of test,

---

65. Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 190 (2017). This article challenges claims made in an earlier article in the same law review by Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017).

66. Kim, *supra* note 65, at 193–94.

67. *Id.* at 191.

68. *Id.* at 202.

69. Professor Ajunwa explains: "[w]hile an applicant may not be right for a specific job at a specific point in time, using the same information that underlies that determination and applying it to a different job, even if at the same company, is antithetical to the bedrock legal doctrine of equal opportunity for all job applicants." Ajunwa, *Auditing Imperative*, *supra* note 30, at 682.

70. *Id.* (citing Richard A. Bales & Katherine V.W. Stone, *The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace*, 41 BERKELEY J. EMP. & LAB. L. 1, 1 (2020)).

the previous low score might prevent them from even getting to the videogame round of interviews. Moreover, if that earlier rejection is now part of a database that other organizations can access, that applicant may have trouble getting access to the next round of testing for any job.[71]

Another way to think about this concern is "algorithmic monoculture," which can result in the same people being systematically excluded from a variety of potential employers and positions.[72] Thus, data retention further interferes with the right to be left alone. Validating the tests or assessments that are used for repeat job applications over time (or finding that they are not valid) might help increase fairness.

The next Section will discuss some recent actions by the Equal Employment Opportunity Commission (EEOC), the Office of the President, and other federal agencies.

II.   EXECUTIVE EFFORTS TO BROADLY ADDRESS ALGORITHMIC DISCRIMINATION AND BIAS

The Biden Administration slowly addressed issues of AI discrimination and bias. Agency guidance memos, policy reports, a presidential executive order, an accountability framework, and other agency actions have begun to provide some guidelines and cautions in the federal arena. The proceeding Sections discuss several of these actions. A brief analysis of federal constraints concludes this Part.

A.   The Equal Employment Opportunity Commission (EEOC) Guidance

One notable action highlighting potential algorithmic bias in the employment realm occurred in 2021, when the EEOC, a federal agency, began examining the implications of AI technologies for hiring and other employment actions.[73] In 2022,

---

71.   *Id.* at 683–84.

72.   Jain et al., *supra* note 47, at 200 (explaining that "[a]lgorithmic monoculture occurs when multiple decision-makers controlling access to a large quantity of valued goods rely on the same or similar datasets and/or models" and "can lead to a pattern of homogeneous outcomes in which the same people are subject to consistent errors or negative outcomes").

73.   *See Artificial Intelligence and Algorithmic Fairness Initiative*, U.S. EQUAL EMP. OPPORTUNITY COMM'N, https://www.eeoc.gov/ai [https://perma.cc/GY79-

the EEOC was considering potential Title VII violations,[74] such as when AI hiring tools routinely screen out applicants of a particular race or ethnicity. Subsequent guidance (of sorts) designed to explain how existing Title VII requirements can apply to AI assessment tools was issued in May 2023.[75] "Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964"[76] (hereinafter "EEOC Guidance") provided notice that employer liability can attach where the tool was developed or administered by an outside vendor.[77]

With a milquetoast caveat that the EEOC Guidance is "not meant to bind the public in any way,"[78] the EEOC Guidance admits its limitation "to the assessment of whether an employer's 'selection procedures'—the procedures it uses to make employment decisions such as hiring, promotion, and firing—have a disproportionately large negative effect on a basis that is prohibited by Title VII."[79] Focusing on disparate impact discrimination, the EEOC Guidance provides a handy measurement of selection rates. Selection rates refer to the percentage of applicants from a particular demographic group who are advanced to the next stage of the process or ultimately selected for employment. First, to determine the selection rate, the number of applicants in a particular category that advance to the next round is compared to the total number of applicants. For example, if there are ten Black applicants for a position and three Black applicants advance, the selection rate for Black applicants is 30 percent.

When different demographic groups have substantially different selection rates, there may be a prima facie case of disparate impact discrimination. If, in this example, there are ten White applicants and six White applicants advance, the

---

P5VK] ("In 2021, U.S. Equal Employment Opportunity Commission (EEOC) Chair Charlotte A. Burrows launched an agency-wide initiative to ensure that the use of software, including artificial intelligence (AI), machine learning, and other emerging technologies used in hiring and other employment decisions comply with the federal civil rights laws that the EEOC enforces.").

74. This EEOC action was being contemplated in 2022. Goodman, *supra* note 64, at 744.

75. *EEOC Guidance*, *supra* note 2.

76. *Id.*

77. *Id.*

78. The EEOC Guidance is not binding because it is not an agency rule. *Id.*

79. *Id.*

selection rate for White applicants is 60 percent. The EEOC Guidance explains how to compare selection rates for different demographic groups, and how to determine roughly what constitutes a substantially lower rate for purposes of disparate impact analysis.[80] In this example, the ratio of Black applicants to White applicants is thirty over sixty, or 50 percent.

The EEOC Guidance explains that under the "four-fifths rule" a ratio of less than four-fifths, that is, 80 percent, is considered substantially different for disparate impact claims.[81] Therefore, the selection rate for Black applicants is considered "substantially different" from the selection rate of White applicants in this example. Noting that the four-fifths rule is not an absolute cutoff, the EEOC explains that smaller differences can also be actionable, particularly if they involve large denominators.[82]

While nonbinding, the EEOC Guidance Question and Answer section notably recommends that employers critically examine potential negative impacts of their practices rather than wait until an applicant challenges them.[83] The report concludes with a reminder that it is *not* meant to bind the public.[84]

B. President Biden's AI Executive Order

A presidential executive order was another federal effort to address algorithmic discrimination. In October 2023, President Biden issued an executive order that sought to establish new standards for AI safety and security, as well as privacy protections ("2023 AI EO").[85] Some of its key components included requiring developers to share safety test results with the federal government. This requirement was enacted under

---

80.   *Id.*
81.   *Id.*
82.   *Id.*
83.   The EEOC Guidance "encourages employers to conduct self-analyses on an ongoing basis to determine whether their employment practices have a disproportionately large negative effect on a basis prohibited under Title VII or treat protected groups differently. *Generally, employers can proactively change the practice going forward.*" *Id.* (emphasis added).
84.   Instead, it is "intended only to provide clarity to the public regarding existing requirements under the law." *Id.*
85.   2023 AI EO, *supra* note 4.

the Defense Production Act and has been criticized by some in the industry as hampering innovation.[86]

The 2023 AI EO instructed the National Institute of Standards and Technology to set standards for developing safe, secure, and trustworthy AI systems. Goals of the standard include advancing American leadership abroad and promoting innovation and competition, all the while ensuring responsible and effective use of the technologies.[87] The 2023 AI EO also emphasized equity and civil rights, which helps promote fairness in the criminal justice system's use of AI technologies, and addressing algorithmic discrimination.[88]

The 2023 AI EO also sought to promulgate standards to reduce the risks of AI-enabled fraud and deception and to enhance cyber-security programs to address critical vulnerabilities. On the privacy front, it required evaluating how federal agencies collect and use consumer data, including commercially available data.

C. Subsequent Federal Agency Actions

Spurred by such federal action as Biden's 2023 AI EO and the EEOC's Guidance before that, various branches of the federal government have begun acting on the use of AI in decision-making. For instance, a number of agencies have successfully undertaken the required actions from the 2023 AI EO in a timely manner. The Stanford Human Artificial Intelligence Lab commended the federal government for its impressive improvements on transparency and for making serious progress thus far on those items that were due to be completed in the first ninety days.[89] For instance, the White House AI Council convened in response to the mandates of the

---

86. *See generally* Mohar Chatterjee & Brendan Bordelon, *The Campaign to Take Down the Biden AI Executive Order*, POLITICO (Jan. 26, 2024, 5:00 AM), https://www.politico.com/news/2024/01/25/conservatives-prepare-attack-on-bidens-ai-order-00137935 [https://perma.cc/W8AX-RJXP].

87. 2023 AI EO, *supra* note 4, § 2(b).

88. *Id.* § 7.

89. Caroline Meinhardt et al., *Transparency of AI EO Implementation: An Assessment 90 Days In*, STAN. UNIV.: HUMAN-CENTERED A.I. (Feb. 24, 2024), https://hai.stanford.edu/news/transparency-ai-eo-implementation-assessment-90-days [https://perma.cc/8H3M-GTWZ] (commending the "serious progress," and "admirable" transparency, while also noting areas for improvements, such as evening out agency variations in the level of detail and accessibility of information reported).

2023 AI EO. This council includes top officials from numerous federal agencies to evaluate their progress and cooperation in implementing the 2023 AI EO.[90] In addition, an AI talent search called the National AI Talent Surge[91] began to locate AI professionals to be hired across and throughout the federal government.[92] The 2023 AI EO mandates the Department of Education develop an AI toolkit and a task force at the Department of Health and Human Services as part of its education AI initiative.[93] The Commerce Department has prepared a proposed rule regarding U.S. cloud companies that provide foreign AI training.[94] Also, risk assessments have been conducted in each of the critical infrastructure sector categories delineated in the 2023 AI EO.[95]

Although critics claim that the 2023 AI EO was an "executive overreach" on the grounds that there was no emergency that would necessitate the use of the Defense Production Act,[96] that Act gives the federal government broad powers over private companies. Under the Act, President Biden authorized the Commerce Department to collect reports from technology companies, set guidelines for these companies, and establish timelines for some of the measures set out in the 2023

---

90. *Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden's Landmark Executive Order*, THE WHITE HOUSE (Jan. 29, 2024) [hereinafter *Fact Sheet* (1/29/24)], https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order [https://perma.cc/R6TW-GN97].

91. *Join the National AI Talent Surge*, AI.GOV, https://ai.gov/apply [https://perma.cc/FRU2-P7GF].

92. *Fact Sheet* (1/29/24), *supra* note 90.

93. "The executive order mandates the U.S. Department of Education to develop an 'AI toolkit' to aid education leaders in applying recommendations for the use of artificial intelligence within classrooms." Kayla Kelly & Thomas Rodgers, *Biden's Executive Order on Artificial Intelligence and Education*, WHITEBOARD ADVISORS, https://whiteboardadvisors.com/bidens-executive-order-on-artificial-intelligence-and-education [https://perma.cc/Y2S7-37FD].

94. *Fact Sheet* (1/29/24), *supra* note 90.

95. *Id.* Critical infrastructure includes communications systems like satellite and wireless internet, utilities, and certain manufacturing processes, as well as commercial facilities and financial services. *Critical Infrastructure Sectors*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY, https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors [https://perma.cc/D59P-WEMP].

96. Chatterjee & Bordelon, *supra* note 86.

AI EO.[97] Agencies have since used the Defense Production Act to compel developers to share risk data.[98]

Another common criticism of the 2023 AI EO was that its regulations would stifle innovation by creating new barriers to entry into the AI market. Critics explain that openness fosters innovation, and the barriers of new regulation can make it more difficult for new entrants with potentially fresh ideas to compete with established presences in the AI market.[99] Some critics expressed concern about entrenching the incumbents and stifling the start-ups on the grounds that regulating technologies rather than sectors, or processes rather than performance, could hamper start-ups.[100]

A policy report and accountability framework are additional examples of federal action on the AI front. In March of 2024, the National Telecommunications and Information Administration's *AI Accountability Policy Report* was published, which included an AI accountability framework submitted by the General Accounting Office.[101] The four pillars of this accountability framework are Governance, Data, Performance, and Monitoring, or "GDPM."[102] The Office of Management and Budgeting (OMB) framework reiterates the federal government's responsibility to maintain oversight of AI technologies that may impact people's rights and safety. It also

---

97. *See* NAT'L TELECOMMS. & INFO. ADMIN., U.S. DEP'T OF COM., NTIA ARTIFICIAL INTELLIGENCE ACCOUNTABILITY POLICY REPORT (2024), https://www.ntia.gov/sites/default/files/ntia-ai-report-final.pdf [https://perma.cc/K77D-F8TZ].

98. *Id.*

99. *See generally* Alden Abbott, *Should the Federal Government Regulate Artificial Intelligence?*, FORBES (May 20, 2024, 11:22 AM), https://www.forbes.com/sites/aldenabbott/2024/05/20/should-the-federal-government-regulate-artificial-intelligence [https://perma.cc/JS5Q-MCNH].

100. *See generally* Daniel Castro, *Ten Principles for Regulation That Does Not Harm AI Innovation*, CTR. FOR DATA INNOVATION (Feb. 8, 2023), https://itif.org/publications/2023/02/08/ten-principles-for-regulation-that-does-not-harm-ai-innovation [https://perma.cc/Y3ZN-BGU6].

101. NAT'L TELECOMMS. & INFO. ADMIN., *supra* note 97, at 38. *See also Fact Sheet: Vice President Harris Announces OMB Policy to Advance Governance, Innovation, and Risk Management in Federal Agencies' Use of Artificial Intelligence*, THE WHITE HOUSE (Mar. 28, 2024) [hereinafter *Fact Sheet* (3/28/24)], https://www.whitehouse.gov/briefing-room/statements-releases/2024/03/28/fact-sheet-vice-president-harris-announces-omb-policy-to-advance-governance-innovation-and-risk-management-in-federal-agencies-use-of-artificial-intelligence [https://perma.cc/P5FM-D5TA].

102. NAT'L TELECOMMS. & INFO. ADMIN., *supra* note 97, at 38.

addresses innovation, expanding the AI workforce, and strengthening AI governance.[103]

The OMB framework requires an inventory of all AI use cases for all agencies.[104] It provides a definition of AI and what it means by a "use case," which includes AI that assists the agency in doing its work.[105] In addition, the OMB framework promotes transparency by requiring agencies to release an inventory of their uses of AI technologies[106] and publish risk assessments of and rationales for each use.[107] Agencies must also report metrics from those uses, explain any exemptions or waivers, and release any government-owned code.[108]

Helpfully, the OMB framework provides a list of which AI uses are included and which are excluded from the inventory. Notable exclusions include the use of AI for a single task a single time; however, the framework recognizes that if AI is used to "accomplish a single task repeatedly, or to carry out a group of closely related standalone tasks," the agency "must inventory it as a use case."[109] Some of the other exclusions include freely available products that have not been modified for the government or are off-the-shelf, research and development technologies, and technologies used as a part of a national security system or by the intelligence community.[110]

The OMB framework also includes detailed information about the type of questions that need to be asked and answered in the required disclosures depending on the type of use. Each agency must consider data that were used to train, retrain, and evaluate the technology, as well as any demographic variables

---

103.   *See id.* at 3–5.

104.   *Id.*

105.   *Id.* at 2 ("[A]ny application of AI designed, developed, acquired, or used specifically to advance the execution of agencies' missions and their delivery of programs and services, enhance decision making, or provide the public with a particular benefit.").

106.   *Id.*

107.   Brian Fung & Sam Fossum, *VP Harris Announces New Requirements for How Federal Agencies Use AI Technology*, CNN (Mar. 28, 2024), https://www.cnn.com/2024/03/28/tech/vp-kamala-harris-agencies-ai-technology /index.html [https://perma.cc/ZT3X-NGKE]. The OMB framework requires each agency to provide "online a complete list of the AI systems it uses and their reasons for using them, along with a risk assessment of those systems." *Id.*

108.   *Fact Sheet* (3/28/24), *supra* note 101; OMB MEMORANDUM NO. M-24-10, *supra* note 7.

109.   AI DRAFT GUIDANCE, *supra* note 5, at 3.

110.   *Id.*

that are used.[111] These disclosures could be made in a multiple-choice or multiple-select format or an actual text input on the disclosure forms.[112] The OMB framework also describes specific requirements for agencies seeking waivers.

In addition, each agency must explain its compliance with risk management practices to identify if the AI use is "rights-impacting, safety-impacting, both, or neither."[113] Each agency must also identify the key risks, determine whether an independent evaluation has been conducted, and determine whether there is "a process to monitor performance of the AI system's functionality and changes to its impact on rights or safety as part of the post-deployment plan for the AI use case."[114] Agencies must address the timeliness and reasonableness of notice for people interacting with AI. Each agency also must evaluate the impact on particular groups and individuals and make efforts to reduce significant disparities across groups. To accomplish these tasks, agencies should incorporate feedback from affected groups.

Federal agencies must have implemented concrete safeguards whenever the AI technologies they use impact rights or safety by December 1, 2024. Agencies unable to adopt appropriate safeguards where their use of AI impacts these rights "*must cease using the AI system,* unless agency leadership justifies why doing so would increase risks to safety or rights overall or would create an unacceptable impediment to critical agency operations."[115]

While the press release touts this U.S. framework as international leadership,[116] the European Union already approved a sweepingly broad AI law by early 2024.[117] As put by some journalists, such a move resulted in the European Union "once again leapfrogging the United States on regulating a critical and disruptive technology."[118]

---

111. *Id.* at 5.
112. *Id.*
113. *Id.*
114. *Id.* at 6.
115. *Fact Sheet* (3/28/24), *supra* note 101; *see also* OMB MEMORANDUM NO. M-24-10, *supra* note 7, at 21 ("Consistent with applicable law, cease use of the AI for agency decision-making if the agency is unable to adequately mitigate any associated risk of unlawful discrimination against protected classes.").
116. Fung & Fossum, *supra* note 107; *Fact Sheet* (3/28/24), *supra* note 101.
117. Artificial Intelligence Act, 2024 O.J. (L 1689).
118. Fung & Fossum, *supra* note 107.

Indeed, the EU Parliament's March 2024 AI law is historic legislation. This legislation took a risk-based approach to ensure that AI products comply with the law before they are made available to the public by asking the big tech companies to explain how they plan to mitigate the risks of generative AI technologies. The European Union was concerned about monopolistic abuse by large tech companies and the scale of the risks. The law categorizes high-, medium-, and low-risk AI systems, requiring different protections depending upon the applicable level. The EU AI Act is lengthy. Generally, it highly regulates and even prohibits some high-risk systems. It subjects lower-risk systems to lower levels of regulation, such as requiring developers and deployers to alert users when they are interacting with AI. Minimal risk, such as the use of AI in video games, is not regulated.[119]

The EU AI Act has a broad reach. For instance, high-risk developers or providers who will be using the systems in the European Union, regardless of where the company is based, need to comply with this guidance. Providers and developers of AI have higher obligations than users and deployers.

The EU AI guidance gives a detailed list of the kinds of systems that are prohibited, including an AI system that:

(1) Deploys subliminal, "purposefully manipulative or deceptive techniques" with the purpose or effect of distorting behavior,[120]

(2) "[E]xploits any of the vulnerabilities" of people or groups related to "age, disability, or a specific social or economic situation" with the purpose or effect of "materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause" significant harm,[121]

(3) Engages in social scoring, based on people's "social behaviour or known, inferred or predicted personal or personality characteristics,"[122]

---

119.   *See* Artificial Intelligence Act, 2024 O.J. (L 1689), art. 6, at 3.
120.   *Id.* art. 5, at 1(a).
121.   *Id.* at 1(b).
122.   *Id.* at 1(c).

(4) Assesses or predicts "the risk of a natural person committing a criminal offence, based solely on the profiling" of that person, or "their personality traits and characteristics" (with a few exceptions),[123]

(5) Creates or "expand[s] facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage,"[124] and

(6) Uses "'real-time remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement," with some exceptions.[125]

D. Constraints on Federal Action in AI Bias-Reduction Goals

While the previous Sections have highlighted the U.S. federal government's concrete actions, addressing hiring bias in AI tools remains a challenge, in part because of the unintended consequences of two federal statutes.[126] One statute, the Privacy Act, enhances protections for personal privacy vis-à-vis government agencies, officers, and actors.[127] The Privacy Act mandates that governmental agencies should collect personally identifiable information only to the extent that it: (1) is "minimally necessary to carry out their statutory mission," (2) can only be used for that particular purpose, and (3) should not be shared or linked with other agencies.[128] The Paperwork Reduction Act prevents agencies from using additional methods of data collection without going through the administrative notice and comment period.[129] While algorithmic hiring eschews

---

123. *Id.* at 1(d).

124. *Id.* at 1(e).

125. *Id.* at 1(h).

126. ARUSHI GUPTA ET AL., STAN. UNIV. HUMAN-CENTERED A.I., THE PRIVACY-BIAS TRADE-OFF (2023) (highlighting the key takeaways from a recent paper by the same authors). *See generally* Privacy Act, 5 U.S.C. § 552a (1974); Paperwork Reduction Act, 44 U.S.C. §§ 3501–21 (1980).

127. 5 U.S.C. § 552a(e).

128. GUPTA ET AL., *supra* note 126, at 2 (summarizing the data minimization principle found in the Privacy Act).

129. *Id.* The Paperwork Reduction Act, 44 U.S.C. § 3517(a), notes that "[i]n developing information resources management policies, plans, rules, regulations, procedures, and guidelines and in reviewing collections of information, the Director shall provide interested agencies and persons early and meaningful opportunity to

paperwork, this federal legislation impacts testing for bias. By compartmentalizing and minimizing data collection, the Act protects data but inhibits comprehensive bias assessment.

A 2023 policy brief and supporting research article (hereinafter "ACM Conference") identify a number of concerns with balancing privacy and fairness issues—ensuring results are accurate across groups and that minority groups are represented in the data sets requires different levels of privacy for majority versus minority groups.[130] Those differential levels may mean that the minority group information is at greater risk of inadvertent disclosure than the information of majority groups.

For instance, to determine whether a federal employer treated an applicant fairly in considering their résumé for an open position, information about other applicants and the outcomes of their consideration is needed. To learn whether multiple government agencies are discriminating against applicants from particular groups, it would be necessary to link data from multiple agencies. Yet federal law curtails both of these practices. Providing that information impacts the privacy of applicants regardless of whether they were selected for employment. The lack of information about the outcomes for other participants means one cannot "test" for fairness in outcomes in any meaningful way. Salient testing for fairness would thus require disclosing or disseminating additional private information in violation of the Privacy Act.

Another consequence of data minimization—related to the tension discussed in the immediately preceding paragraph—prohibits some agencies from collecting or linking demographic information. Thus, when the information is needed to perform bias assessments, the data is not there. The ACM Conference authors cautioned that "[d]ata minimization should not function as a license for blindness to disparities."[131] While the authors recognized the social construction of race and the arguments that the government should not classify by race, on the issue of algorithmic bias, their work focuses on "fairness through

comment," and § 3508 explains that "[b]efore making a determination the Director may give the agency and other interested persons an opportunity to be heard or to submit statements in writing."

130.   Gupta et al., *supra* note 8, at 492.

131.   *Id.* at 493.

awareness,"[132] not only because "race and ethnicity have measurable disparate impacts on individuals," but also "across intersections of demographic characteristics."[133]

There is a dearth of demographically linked data for many federal agencies, which hampers assessments of algorithmic biases. For instance, laws inhibit most, if not all of the ten agencies studied.[134] Three of the agencies that do collect demographically linked data do so inconsistently and in ways that could be unreliable. For example, some agencies engage in a "visual assessment," which requires staff to guess the race or ethnicity of an individual by merely visually inspecting them. However, two agencies—the Social Security Administration and Health and Human Services—have been successful in collecting, linking, and sharing data in a consistent and reliable way.[135]

These researchers identified five categories of barriers to obtaining linked data:

(1) Legal restrictions that prevent data collection, such as ethnic data that exceeds what is "minimally necessary."

(2) "Fragmented or outdated technical infrastructure" that inhibits data retention or contains glitches that prevent users from inputting certain fields.

(3) Resistance by the actual data collectors who, for instance, may not want to ask constituents about their race or marital status.

(4) Concern that asking for demographic data will "increase non-response rates" given that the percentage of people who stay online or on a telephone call to complete a survey often has an inverse relationship with the number of questions on the survey instrument, such that the

132.   *Id.* (citing Cynthia Dwork et al., *Fairness Through Awareness*, 3 PROC. OF THE 3RD INNOVATIONS IN THEORETICAL COMPUT. SCI. CONF. 214–26 (2012)).

133.   *Id.*

134.   *See id.* at 494, 497 ("[L]egal barriers prevent data collection for some or all programs.").

135.   *Id.* at 495 (noting that these agencies have a "systematic approach to demographic data collection," and SSA "collaborates with other agencies, primarily Census, to link records from four different population surveys to determine race," and also combines with HHS).

greater the numbers of questions the lower the completed response rate.

(5) A lack of "dedicated financial and personnel resources"[136] focused on tasks such as the purchase, cleaning, sorting, and storing of data in a way that is relevant and useful for the agency's operations.

The ACM Conference researchers proposed some concrete solutions to the problems associated with the inability to meaningfully conduct disparity assessments without demographic data. For instance, while the Privacy Act contains some exceptions, "bias is not explicitly acknowledged as a valid exception" and does not easily fit into any of the categories of exceptions, such as for statistical research, "need to know," or simple routine use.[137] It is rather surprising that the current process for collecting data that the EEOC uses for evaluating potential wage discrimination claims is not the same as what the Consumer Protection Bureau uses for evaluating potential discrimination in non-mortgage lending. For example, for food stamps and farm subsidies administered by the United States Department of Agriculture (USDA), methods of data collection are limited to visual observation and voluntary disclosures on forms.[138] Visual observation without an assertion of racial or ethnic identity is not sufficient in EEOC claims.

Addressing pay equity is challenging because in order to understand how different racial or gender groups are being treated, it is important to have the data of people from different groups and their salaries. But to the extent that some groups only have a small number of data points, their salaries would end up being revealed. Let us now say that our hypothetical federal agency used salary information to demonstrate that its managers of color were not being underpaid compared to White managers. For example, salary data shows that all of the managers make between $100,000 and $200,000, annually. For managers of color, the data shows that the salary range is $105,000 to $125,000. If there are only two managers of color within that agency, then their individual salaries—which are

---

136. *Id.*
137. Gupta et al., *supra* note 8, at 496.
138. *See id.* at 497 tbl.2.

within a narrower range rather than the overall salary range—would obviously be $105,000 and $125,000.

Even if the agency reports an average for the managers of color at $115,000, it would be clear that one person of color earns no less than $100,000 (given the overall range), and therefore the other earns no more than $130,000 (to meet the average of $115,000). In any case, disclosing salary data by race results in a greater infringement on the privacy rights of people of color. This disclosure would be important and necessary to determine whether the AI hiring, evaluation, and assessment processes are creating discriminatory outcomes.

In contrast, if there are one hundred White managers, the same impact of disclosing private salary information would not occur because any one of them could be earning any amount between $100,000–$200,000. When the group is a much larger one, merely disclosing the average of $150,000 does not identify where any particular person falls on the salary scale.[139] Thus, collecting data from groups to identify potential bias has the contrary effect of creating a higher risk of personally identifying individuals from minority backgrounds. The data retention policies make the information more easily identified with minority individuals than when that same information is collected about White people, or anyone whose demographic characteristic is in the majority.

Another challenge is that an individual's privacy risks should not be outweighed by a lack of understanding of the "collective benefits of demographic data collection."[140] Consider the example of the anticipated effect of adding a citizenship question to the 2020 Census and the substantial concern that it would diminish response rates.[141] The concern about requests for private information reducing response rates may be valid, but the "effect of demographic questions on response rates to government surveys is difficult to predict and likely varies by context and demographic characteristic of interest."[142]

---

139. *Id.* at 499.
140. *Id.*
141. *See generally* J. David Brown et al., *Predicting the Effect of Adding a Citizenship Question to the 2020 Census*, 56 DEMOGRAPHY 1173 (2019).
142. Gupta et al., *supra* note 8, at 499.

E. Recommendations on the Federal Front

Gupta et al. make several recommendations for increasing transparency in federal bias assessments. The first is to allow linking records among agencies to facilitate bias assessments.[143] This linkage would provide a greater number of data points, particularly for members of underrepresented groups. A larger amount of data can enhance the accuracy of the assessment in determining whether certain groups are being systemically excluded or simply have a smaller representative population. Privacy concerns must remain at the forefront, but they recommend adding new exceptions[144] or re-interpreting existing statistical research and routine use exceptions to include bias assessments.

Secondly, the ACM Conference suggests that agencies use shared demographic data solely for conducting their own disparity assessments.[145] One way to implement this restriction would be to have agencies make demographic data available only to agency employees conducting the disparity assessments and bias audits.[146] Conducting these assessments with more data increases accountability.

Thirdly, the ACM Conference recommends "streamlin[ing] the process of gathering demographic data on forms and surveys."[147] For instance, when the Census's definitions or listings for "race" or "ethnicity" options are updated, other agencies do not and cannot automatically adopt the new definitions or lists. Without common categories or sufficient personnel to evaluate which new categories align with previous ones, propose revised agency rules, and navigate the notice and comment period to enact those changes, agencies cannot effectively share information. Amending the Paperwork

---

143.   GUPTA ET AL., *supra* note 126, at 4.

144.   Gupta et al., *supra* note 8, at 500 (recommending "adding an exception to the Privacy Act that permits inter-agency record linkage specifically for bias assessment subject to" additional protections which the authors identify).

145.   GUPTA ET AL., *supra* note 126, at 5 ("[A]gencies should adopt strict institutional protections so that demographic data is used exclusively for equity assessments."); *see also* Ali Hasan et al., *Algorithmic Bias and Risk Assessments: Lessons from Practice*, DIGIT. SOC'Y, Aug. 2022, at 5–6 (articulating that an equity assessment is an internal evaluation of policies and practices to determine whether and where discrepancies lie).

146.   Gupta et al., *supra* note 8, at 501 (recommending having an "internal separation of functions" within the agencies).

147.   GUPTA ET AL., *supra* note 126, at 5.

Reduction Act with this issue in mind could streamline the data-gathering process[148] and enhance accountability across agencies.

Fourth, the ACM Conference suggests that Congress should enact new measures that promote sharing in addition to the gathering of data noted above.[149] For example, Congress proposed a bill for a National Secure Data Service Act that would create a clearinghouse for collecting data, disseminating data, and facilitating interagency access to that data.[150] Increased data sharing can also increase accountability as patterns become clearer over time and across agencies.

These four suggestions could counteract the resistance that agency staff often have toward using interagency systems because the data and systems do not match up easily[151] and can dramatically affect accountability. Measures impacting efficiency are already being explored through President Biden's 2023 AI EO, but the pushback on linking records is likely to remain strong; congressional action resulting in actual legislation is unlikely at best.

III. STATE AND LOCAL EFFORTS

In the past few years, many state (and some local) governments have been considering AI regulations. In 2023, four states enacted AI statutes to address bias, and in the 2024 legislative session, forty-five states plus some territories introduced bills aimed at AI regulation.[152] The "AI Legislative

---

148. Gupta et al., *supra* note 8, at 501 (recommending a "streamlined process for capturing demographic data on federal forms or running auxiliary surveys," as an amendment to the Paperwork Reduction Act). "Privacy enhancing technologies such as differential privacy and secure multi-party computation are of course quite important here for enabling secure and private inter-agency data sharing, but we emphasize that they are unlikely a complete solution." *Id.*

149. These new measures should "advance privacy-protective sharing of administrative data." GUPTA ET AL., *supra* note 126, at 5.

150. H.R. 3133, 117th Cong. §2(c) (2021) ("[T]he Director shall engage with State and federal agencies to collect, acquire, analyze, report, and disseminate statistical data in the United States and other nations to support governmentwide evidence-building activities.").

151. Gupta et al., *supra* note 8, at 501 ("[N]umerous instances where the apparent agency resistance stemmed from lack of technical resources to incorporate demographic data into agency systems.").

152. *Artificial Intelligence 2024 Legislation*, NAT'L CONF. OF STATE LEGISLATORS, https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation [https://perma.cc/RUK2-6GK3] (last updated Sept. 9, 2024).

Scorecard," (hereinafter "Scorecard") launched in June 2024, provides a "structured tool" for evaluating proposed AI bills making their way through legislative bodies.[153] The Scorecard is a rubric for commercial uses of AI, setting out minimum standards and identifying "key provisions that effective AI legislation should contain."[154] The Scorecard focuses on ten categories of provisions, recognizing that not all need to be present depending on the purpose and scope of the bill.[155] These categories of provisions include definitions, baseline requirements, antidiscrimination, data minimization, security advancement, and rulemaking, among others.

This Part will begin with a brief description of New York City's local law, which is one of the first in the nation to attempt to address algorithmic discrimination in the private employment realm. The local law is a bold attempt to address the lack of transparency, accountability, and fairness in "automated employment decisions tools" (AEDT). Second, it will evaluate the New York City local law using the recently launched Scorecard rubric. This Part will conclude with an analysis of how the New York City law has been operating over the past year.

A. New York City Local Law 144

In a 2022 article addressing the procedural due process implications of AI technologies, this Author previewed the pending New York City Local Law 144 ("NYC law"). The law would prohibit the use of automated employment technology to screen job applicants unless an independent auditor had performed a bias audit of that technology prior to its use.[156] The final rule implementing the law took effect in July 2023, mandating four basic requirements: that employers (1) provide notice to potential job seekers through a link or website posting with information in the online application portal when AI systems are being used to make employment decisions,[157] (2)

---

153. *EPIC Releases Its AI Legislation Scorecard*, ELEC. PRIV. INFO. CTR. (June 25, 2024), https://epic.org/epic-releases-its-ai-legislation-scorecard [https://perma.cc/Q2PT-VZGE].

154. Kara Williams, *AI Legislation Scorecard*, ELEC. PRIV. INFO. CTR., at i (June 25, 2024), https://epic.org/wp-content/uploads/2024/06/EPIC-AI-Legislation-Scorecard-June2024.pdf [https://perma.cc/NU7G-SHN5].

155. *Id.* at 1.

156. Goodman, *supra* note 64, at 731–32; *see* N.Y.C., N.Y., ADMIN. CODE §§ 20-870 to -874. (2021).

157. N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-304 (2024).

give job-seekers a chance to opt out of automated decision-making and request an alternative process in the consideration of their individual employment application,[158] (3) obtain a third-party independent bias audit report,[159] and (4) publicize a summary of that audit report on the employer's website.[160]

Advantages of the law include fines for failing to provide notice, audit results, and opt-out provisions when AI is used in hiring or promotion decisions. Interestingly, the law also holds vendors liable if they supply technology that perpetuates or contributes to these biases.[161]

B. Applying the Scorecard and How the NYC Law Measures Up

Using the newly launched AI Legislative Scorecard ("Scorecard"), this Section analyzes the NYC law.[162]

In the first category, strong definitions, the Scorecard recommends particular definitions of "algorithmic discrimination"[163] and "protected classes."[164] The Scorecard also advises that definitions focus on the "function of the system,"[165] including the consequences of using that AI. For instance, the definitions should consider whether the consequences are high risk, high stakes, or low risk, low stakes, in terms of meaningfully impacting a "person's safety or well-being."[166] Moreover, the definitions should "facilitate

---

158.   *Id.* § 5-304(a).

159.   *Id.* § 5-301.

160.   *Id.* § 5-303.

161.   One student note addressed the New York City law in the early stages of its implementation. *See* Fuchs, *supra* note 21, at 205 (describing how employers use algorithmic tools for recruitment and hiring, and then addressing some of the privacy and discrimination issues that arise, gaps in enforcement, and how to fill some of those gaps).

162.   Williams, *supra* note 154.

163.   The Scorecard defines algorithmic discrimination as the "use of an AI system in a manner that discriminates, in treatment or effect, or otherwise makes unavailable the equal enjoyment of goods, services, or opportunities on the basis of a protected class (with exceptions for use to identify/prevent discrimination or increase diversity and inclusion)." *Id.* at 1.

164.   The Scorecard defines protected classes as including at a minimum "race, color, ethnicity, national origin, religion, sex, status as pregnant, gender identity, sexual orientation, familial status, disability, biometric or genetic information, income source or income level, or any other classification protected by law." *Id.*

165.   *Id.*

166.   *Id.* at 2.

decisions with legal or similarly significant effects on any person's civil rights, civil liberties, privacy, or equal opportunities."[167] In addition, consequences include when there is an "impact [on] any person's access to or significant change in the price of critical benefits, resources, or services."[168]

The NYC law does not fully satisfy the Scorecard's first category requirements. Definitions of algorithmic discrimination, protected classes, or consequences are absent from New York City's law. It does provide definitions of the terms "automated employment decisions tool" (AEDT) and "bias audit."[169]

The NYC law satisfies the second Scorecard category. This second category addresses baseline requirements, which include disclosing when the AI system is making a consequential decision as defined above, and not simply when AI is "the 'sole' or 'controlling' factor in consequential decisions."[170] The NYC law meets this second standard as it requires employers to disclose whether AEDTs are used to screen the candidate, as well as whether they will be used in assessing and evaluating a candidate for initial hiring or promotion.[171] The law and accompanying rule define the phrase "substantially assist or replace."[172]

The NYC law fails to meet the third and fourth Scorecard categories. Category three prohibits algorithmic discrimination, and category four prohibits particularly harmful uses and requires providing a mechanism for banning AI systems that are subsequently discovered to be harmful, either in the abstract or as applied in the real world.[173] Some of the harmful uses listed include, at a minimum, "emotion or attribute recognition, social scoring, one-to-many facial recognition, and nonconsensual deepfakes."[174]

Interestingly, the NYC law does not contain a prohibition against algorithmic discrimination in an AEDT; it requires only disclosure and audit.[175] Thus, posting notice that an AEDT *does*

---

167.   *Id.*
168.   *Id.*
169.   N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-300 (2024).
170.   Williams, *supra* note 154, at 3.
171.   N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-303 (2024).
172.   *Id.* § 5-300; N.Y.C., N.Y., ADMIN. CODE § 20-870 (2023).
173.   Williams, *supra* note 154, at 3.
174.   *Id.*
175.   N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-303 (2024).

discriminate seems to suffice under the local law. The NYC law does not prohibit any particular uses either, regardless of the level of harm. It focused solely on biased results, not the harm resulting from biased processes.

The NYC law also fails to address the limitations in the Scorecard's fifth category. The fifth category is data minimization, and the Scorecard recommends prohibitions on "collecting, processing, retaining, or transferring personal data" unless "necessary and proportionate to develop, train, or maintain a specific product or service," as long as the person gives affirmative consent.[176] The NYC law does not address data minimization, and it in fact requires the use of historical data.[177] This mandate indirectly *requires* data retention, and does not include any clear limitations.

Category six focuses on transparency and accountability, and the NYC law partially meets the Scorecard requirements. Analyzing this category requires some additional details. The Scorecard provides some guidelines on what should be contained in those pre- and post-deployment impact assessments and audits.[178] It also requires testing, posting impact assessment results prior to deployment, and post-deployment audits and impact assessments.[179] In addition, there should be consequences when bias or discrimination are discovered, as well as when an organization fails an audit.[180]

Before addressing the specifics of the NYC law some additional background information is useful here. An audit requires independent evaluation, is intended to serve the public or some outside actor, and is based on binary outputs like "meets

---

176.   Williams, *supra* note 154, at 4.

177.   N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-302 (2024).

178.   Williams, *supra* note 154, at 4–5 (identifying minimum standards for inclusion in audits and assessments as the following: "provenance in quality of training data and inputs," how errors "are measured and limited," "inputs and logic on which the AI system operates," how it was "developed and tested," "intended uses and foreseeable misuses," "process and results of regular validation studies," "types of outputs generated," unintended downstream uses, the "results of any bias audits or testing," "data management policies and procedures," "procedures for human review or redetermination, and results of risk-benefit analyses").

179.   *Id.* at 5–6.

180.   *Id.* at 6 (noting that "deployers must pause [the] use of the AI system until that bias or algorithmic discrimination can be mitigated—or decommission the AI system if the bias or algorithmic discrimination cannot be addressed," and any failed audit must be reported to government regulators as well as the downstream deployers).

standards" or "does not meet standards."[181] Assurances include a grading system from A to F or even a qualitative assessment of excellent to poor.[182] Assessments, in contrast, are generally internal and "intended as a service to the organization," "aimed at providing feedback and usually at building recommendations and advising clients on how to perform better with respect to some legal or ethical standard."[183]

Some researchers recommend both ethical risk and algorithmic bias assessments in a recent article that examines key factors for conducting these assessments and concludes that focusing on their interdependence is a key component of minimizing the risk of harm.[184] In explaining terminology, these authors use the term "ethical risk assessment," intending that "the focus of such an assessment is the broader socio-technical context of the algorithm, how the algorithm is employed to serve certain purposes of an organization and how it affects the rights and interests of stakeholders—including whether it is unfair or biased in some way."[185]

This "ethical risk assessment" contrasts with technical evaluations referred to as "algorithmic bias assessments," which focus solely on identifying bias.[186] The ethical risks assessed are primarily related to the effects on the "central interests, well-being, and moral rights of any individuals, groups, or institutions."[187] While they often focus on the negative risks, they can also highlight positive benefits, such as increased access to education and job opportunities, as well as "improved health and well-being to users and to society"[188] as a whole.

181. Hasan et al., *supra* note 145, at 4.

182. *Id.*

183. *Id.* at 4.

184. *Id.* at 3–4. The researchers highlight that they have actually performed algorithmic assessments rather than merely studying and writing about them and provide useful insight in their article. *Id.* at 1.

185. *Id.* at 2 (citing Andrew D. Selbst et al., *Fairness and Abstraction in Sociotechnical Systems*, 2019 PROC. OF THE ACM CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 59; Shea Brown et al., *The Algorithm Audit: Scoring the Algorithms that Score Us*, BIG DATA AND SOC'Y, Jan.–July 2021, at 1). Ethical risk assessments refer to assessing the risk that "the use of the algorithm negatively impacts the rights and interests of stakeholders, with a corresponding identification of situations of the context and/or features of the algorithm which give rise or contribute to these negative impacts." *Id.* at 5.

186. *Id.* at 5.

187. *Id.* at 6.

188. *Id.* at 5.

While the NYC law mandates pre-deployment bias audits, which must be updated annually, the law does not address any consequences for negative impact assessments or failed audits. However, the NYC law provides penalties for noncompliance,[189] such as fines for failing to provide notice or opt-outs and failing to conduct the bias audit annually.

The seventh Scorecard category addresses data security, which requires measures to protect data from being hacked, stolen, and disseminated improperly. The eighth category prohibits unfair practices, specifically retaliation against consumers, adverse action against whistleblowers, and using "manipulative design or dark patterns to subvert individuals' decision-making."[190] The NYC law addresses neither data security nor unfair practices and thus fails to satisfy either of these Scorecard categories.

The NYC law performs well in category nine. Category nine focuses on individual rights, including notice requirements to job seekers, the right of access to information, and the right to correct or complete said information, for all affected individuals for whom AI is being used in a consequential way. It notes that an individual should also have an opt-out provision and an opportunity to request human review.[191]

The NYC law requires notice of the use of AEDTs, and an opt-out provision whereby applicants can "request an alternative selection process or accommodation."[192] Because it is limited to employment decisions on screening initial applicants and for promoting current employees,[193] the NYC law can be presumed to always involve "consequential decisions." Thus, it does a good job of meeting the category nine requirements.

The NYC law also satisfies the tenth and final category. The tenth category focuses on enforcement, rulemaking, mandating that statutory damages be available, and that some agency or government official with adequate funding has both "investigative and enforcement authority," including the

---

189.   N.Y.C., N.Y., ADMIN. CODE § 20-872 (2024) (providing for fines up to $500 for the first violation and others on that same day and ranging from $500 to $1500 per day for subsequent violations).
190.   Williams, *supra* note 154, at 8.
191.   *Id.* at 9.
192.   N.Y.C., N.Y., ADMIN. CODE § 20-871(b)(1) (2024).
193.   *Id.* § 20-870 (definitions).

authority to grant injunctive relief, order disgorgement, and permit consumers and the government to bring suit.[194]

The New York Department of Consumer Workers' Protection promulgated draft rules relating to the NYC law and made them available for public comment, revised the draft, and then finalized the rules in 2023.[195] They contain an enforcement provision that includes mandating compliance.[196]

On paper, the NYC law does *not* seem to satisfy most of the Scorecard's minimum standards. But how is it working in practice? The next Section responds to this question.

C. How Effective is the New York City Law in Reality?

To test the NYC law's effectiveness, shortly after its enforcement date, Cornell used student researchers to spend no more than thirty minutes per employer trying to find the required information. The researchers began with the company websites and followed up by telephone and email to gauge the level of employer compliance with the NYC law. The researchers found that only 5 percent of those employers they checked had audit reports, and only 4 percent had notices about automated decision-making tools.[197] The students expressed their frustration with finding audit information notices, which were often not located in a consistent or intuitive spot on employer websites.[198]

The researchers also expressed concern that the law was left open to interpretation by employers. For instance, the law requires disclosure only when AI tools are "used to 'substantially assist or replace discretionary decision making,'" thus permitting employers themselves to determine what is "substantial assistance" and whether it is replacing rather than

---

194. Williams, *supra* note 154, at 10.

195. N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-302 (2024).

196. N.Y.C., N.Y., ADMIN. CODE § 20-873 (2024).

197. Grace Gedye, *New Research: NYC Algorithmic Transparency Law Is Falling Short of Its Goals*, CONSUMER REPS. (Feb. 8, 2024), https://innovation.consumerreports.org/new-research-nyc-algorithmic-transparency-law-is-falling-short-of-its-goals [https://perma.cc/5AVZ-E639]. The article explains that one aspect of the law is that companies are required to post a notice to jobseekers when automated technologies are used to make employment decisions. *Id.* A second aspect of the law is to provide a link to an independent audit addressing potential biases in the technology. *Id.* This article analyzes the findings of Cornell University researchers and Consumer Reports on the usefulness of the law. *Id.*

198. *Id.* The notice should provide information about the use of AEDT, how to opt out, and the results of the annual bias audit of those AEDT tools.

supplementing discretionary decision-making.[199] However, the accompanying rule in 20-870 et seq. defines "substantial assistance" as relying upon a simplified input (such as a binary go-no-go), where one response is relied upon more heavily than any other factor, or when it is used to overrule decisions based on other factors, such as human decision-making.[200]

The complete research report[201] identifies substantive concerns with the NYC law, including that it "does not require that systems meet any discrimination threshold, including the four-fifths rule," "[n]or . . . provide any guidance for remediation of systems when audits disclose disparate impact," while recognizing that general employment discrimination laws still apply.[202] This absence of a standard can lead to a situation where an employer complying with the NYC law posts its audit reports, which then makes it potentially subject to liability for disparate impact discrimination.[203] The New York State Department of Consumer and Workers' Protection "has demanded transparency about an activity that another regulator (the EEOC) has jurisdiction over but would not be able to observe in the usual course of business."[204]

Another critique of this null compliance approach is that it undermines enforcement abilities as well as the transparency goal, particularly because "employers . . . are excused from reporting impact ratios for groups that they have hired the least."[205] Nevertheless, "[d]espite its significant flaws," the law has already had some success in promoting change, such as "employers who claim to have stopped using these tools in New York City."[206]

The research report identified four goal areas and criticized implementation of the New York City law in each of those areas. The biggest critique was the inability to determine whether a company was complying (or attempting to comply) with the

---

199. *Id.* Notwithstanding their findings that it was "falling far short of its goals, and that, in practice, job-seekers can't be expected to find or make use of the notices and audits companies are required to disclose." *Id.*

200. N.Y.C., N.Y., RULES OF THE CITY tit. 6, § 5-300 (2024).

201. Lucas Wright et al., *Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability*, 2024 PROC. OF THE ACM CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1701 (2024).

202. *Id.* at 1703.

203. *Id.*

204. *Id.*

205. *Id.* at 1709.

206. *Id.* at 1709–10.

ordinance due to the notion of "null compliance." The term "[n]ull compliance describes a state in which the absence of evidence of compliance cannot be ascertained as noncompliance because the investigator lacks the information to determine if the regulated party's actions or products are in the scope of the regulation." Accordingly, subtracting a compliance percentage from one hundred does not result in an actual noncompliance rate.[207]

Several situations may fit within the null compliance category, such as when an employer does not use any automated employment decision system, or if they do, the employer claims that their system is "outside of the scope of the law." Other employers whose uses of AEDTs are within the scope of the law may simply be in the process of seeking an auditor, and thus have no audit to post (yet).[208] Those employers who have conducted an audit may have decided not to publicly post the results, but rather to provide notice to individual applicants. Even those employers who have conducted an audit and are posting the results may be doing so in a way that is difficult for applicants to find or access, thereby thwarting the transparency goal.[209]

Successful notice and consent to algorithmic decision-making "depends on informed decision-making, which relies on the accessibility, comprehensibility, and usability of information."[210] In New York City, much of the information was inaccessible, incomprehensible, or in an unusable format.[211]

These researchers explained that it was difficult for job seekers to reap the benefits of the NYC law, and easy for employers to defeat the purpose of the law—which is to be transparent with job seekers and the public about uses of AEDTs.[212] They concluded that "the law is not helping job-seekers or improving overall algorithm transparency

---

207. *Id.* at 1704.

208. *See id.*

209. *Id.* at 1705.

210. *Id.*

211. *Id.* at 1707–08 (discussing the difficulty in locating disclosures and, once found, the ambiguity within them).

212. Lucas Wright et al., *Studying How Employers Comply with NYC's New Hiring Algorithm Law*, CAT LAB, https://citizensandtech.org/research/2024-algorithm-transparency-law [https://perma.cc/WDS8-AQFW] (noting that it was "practically impossible for job-seekers to learn about their rights or exercise them under Local Law 144," and expressing strong doubt that "all employers are complying with the law's transparency requirements").

because it gives employers extreme discretion over compliance and strong incentives to avoid transparency."[213]

There are several additional criticisms of the NYC law that could help guide the drafting future laws to be more effective. The New York Civil Liberties Union noted one particularly salient critique—calling the law "too weak to make much of a difference."[214] Another notable critique was the delegation of enforcement authority without any assignment of rulemaking authority.[215] Although the New York Department of Consumer and Workers' Protection did craft some rules, they were not adequate for the task. In addition, the auditing requirement itself is perhaps not specific enough and should include more details about what is required in the audit.[216] Increasing transparency of payments and timing to ensure auditor independence could also strengthen the effect of the statute.[217] As discussed above, transparency is not enough. So, what should regulators and employers do? The next Section explores options.

IV. Concluding Thoughts: What Should be Done about Algorithmic Discrimination in the Employment Realm

A. Begin with an Ethical Risk Assessment

Organizations who wish to use AI technologies in the hiring space should first undertake two different assessments—one for ethical risk and another for bias risk—to make sure their AI tool is not perpetuating bias. The ethical risk assessment should

---

213.   *Id.*

214.   Simon McCormack & Daniel Schwarz, *Biased Algorithms Are Deciding Who Gets Hired. We're Not Doing Enough to Stop Them*, ACLU OF N.Y. (Oct. 20, 2023), https://www.nyclu.org/en/news/biased-algorithms-are-deciding-who-gets-hired-were-not-doing-enough-stop-them [https://perma.cc/3JX6-PCQK] (noting that while the law is advertised as requiring AI technologies that do not include negative biases against women or people of color, there are "many ways that an automated tool declared to be bias-free by an audit could nevertheless be discriminatory").

215.   Fuchs, *supra* note 21, at 188–89.

216.   *See id.* at 208 (addressing the symbiotic relationship between auditors who are relying upon clients to get work and clients who are relying upon reports that their algorithms are not demonstrating inappropriate biases could be problematic, and clarifying the auditing requirement and providing more detail is one strategy that the note suggests as well as prepayment for the audits so outcome does not impact auditor compensation).

217.   *Id.* at 213–15.

include two distinct stages: (1) the identification stage, and (2) the prioritization stage.

In the first stage, organizations should identify the list of potential harms without making any judgment as to which are more or less likely to occur or which are significantly worse or better than others. For example, potential harms include inaccurate classifications and classifications based on inappropriate data. While the first stage is most likely to focus on "potential harms for end users, or members of society on whom algorithms are deployed," it is still important to "be on the lookout for independent risks to other stakeholders."[218]

Next, in the prioritization stage, the aim is to link each ethical risk with the "underlying features of the product or technology that is the primary driver for that risk" and then prioritize how important that feature is in terms of the overall use of the product.[219] Employers should ensure that the risk of harm is not evaluated in a vacuum but rather in comparison or contrast to alternatives. For instance, when AI technologies are employed to screen employment applicants, one potential harm is that activating facial recognition software may impair its ability to accurately recognize and interpret facial expressions of individuals with darker skin. The harm of discrimination against people of color may be significant; if, however, the facial recognition component is not necessary or plays only a small role in the algorithmic evaluation of the candidate, then the assessment may recommend deactivating the facial recognition software. If the technology can function effectively and improve outcomes for applicants of color, this adjustment could represent an ethical choice.

Employers should conduct bias assessments only *after* performing an ethical risk analysis. This delay allows the process to become iterative as it "allows for a feedback loop between the ethical risk assessment and the bias assessment."[220] Then, the bias assessment "in turn informs the ultimate assessment of the main ethical risks and helps guide proposed recommendations"[221] to the employers or clients. The iterative process builds on itself over time to work more efficiently and effectively.

---

218.   *See* Hasan et al., *supra* note 145, at 7.
219.   *Id.*
220.   *Id.* at 8.
221.   *Id.* at 11.

B. Then Perform the Bias Risk Assessment or Audit

Next, before performing the bias risk assessment, organizations should ensure that testing data will:

> (1) mimic, to the extent possible, the conditions under which the algorithm is deployed, (2) be labeled by self-identified race, gender, and other protected attributes of interest in order to construct intersectional groups, and (3) contain sufficient number of datapoints in the regions of the parameter space that are ethically salient, to obtain statistically significant test results.[222]

Having job seekers self-identify their own demographic labels is crucial because "individuals are in the best position to make these sorts of determinations for themselves."[223] This concern also includes the testing metrics and how to measure fairness given that it is not a mathematical concept.[224] Thus, in the context of résumé screening, it is important that the tool be trained on the résumés of real people rather than those that have been created for the purpose of providing additional training data.[225]

C. Make Adjustments After Evaluating Data from the Risk Assessment or Audit

So how can an employer appropriately respond to substantial racial disparities in testing results? Organizations should know that post-audit adjustments are legally permissible and do not violate *Ricci*'s "reverse discrimination" rule.[226] In *Ricci,* the U.S. Supreme Court addressed invalidating an employer's job-related testing data after it discovered a substantial racial disparity in those results. The U.S. Supreme Court held that by invalidating the test results after the fact, the

---

222.   *Id.* at 8–9.
223.   *Id.* at 9. "[W]e cannot tell how good the algorithm is at detecting the disease *across different racial groups* without having an independent way of determining what racial group the people in the testing data belong to." *Id.*
224.   *Id.*
225.   Researchers caution that "the testing data should include real résumés of individuals that are likely to apply to such jobs, and not fabricated or artificial résumés that could, in subtle ways, introduce artificial elements or yield an unrepresentative data set." *Id.* at 10.
226.   *See* Ricci v. DeStefano, 557 U.S. 557, 562–63 (2009).

employer was discriminating against the White applicants who had scored well on the previously articulated metrics.[227] Some scholars such as Kroll et al. argue that using auditing "to revise processes to eliminate implicit or unintended biases," such as with employment algorithms that disproportionately reject those from minority groups, would violate federal law based upon the *Ricci* case.[228]

Arguments that revising algorithms that produce racially disparate impacts violates Title VII misinterpret *Ricci*.[229] As Professor Kim explains, *Ricci* applied to "legitimate expectations." One legitimate expectation is that those who did well on the test would be promoted within the New Haven fire department, whereas applicants to general employment positions "have not suffered an adverse action because of their race merely because the employer decided to change its hiring algorithm."[230] Because there is no legitimate expectation that a person will get a job simply by applying, "if the employer chose to revise the algorithm to eliminate unintended biases, no legitimate expectations would be disrupted and nothing in *Ricci* would prevent the employer from making the change prospectively."[231]

Professor Kim cites to both Justice Kennedy's majority opinion and the dissenting opinion of the four Justices, which states that "Title VII does not prohibit an employer from considering, before administering a test or practice, how to design that test or practice in order to provide a fair opportunity for all individuals, regardless of their race."[232] She surmises that redesigning a test or practice could also be covered under this aspect of the U.S. Supreme Court's ruling and thus "employers are permitted to audit automated decision processes and change them prospectively in order to eliminate identified biases."[233] She cites a subsequent Second Circuit case, which made the distinction that where the "problem was with the test itself, rather than with a particular set of results," changing the

227. *Id.*
228. Kim, *supra* note 65, at 197 (citing Kroll et al., *supra* note 65).
229. *Id.* at 199.
230. *Id.*
231. *Id.*
232. *Id.* at 200 (quoting *Ricci*, 557 U.S. at 585).
233. *Id.*

test did not violate Title VII.[234] Thus, Professor Kim reasons that prospectively changing a test or decision-making process does not result in an adverse action under Title VII.[235]

Based on this analysis, Professor Kim concludes that auditing and adjusting are constitutionally valid remedies.[236] Professor Ajunwa agrees[237] that bias assessments are appropriate legal ways to "ensure that any benefits of automated hiring are not negated by (un)intended outcomes, such as unlawful discrimination on the basis of protected characteristics."[238] Professor Ajunwa reasons "not just that the law allows for audits, but that the spirit of antidiscrimination law *requires* it."[239] She explains that "not only is the nature of prediction problematic (particularly given historical employment discrimination), but also, the manner in which such prediction is accomplished further creates opportunities for unlawful discrimination and exclusion."[240]

This rationale was acknowledged as valid by the U.S. Supreme Court as recently as 2023 in *Students for Fair Admissions, Inc. v. President of Harvard College.* The majority opinion reasoned that such remedies do not automatically violate the Equal Protection Clause.[241] In addition, the 2023 EEOC Guidance explains that "[g]enerally, employers can *proactively* change the practice going forward."[242]

---

234. *Id.* at 200 n.57 (quoting Maraschiello v. City of Buffalo Police Dep't, 709 F.3d 87, 95 (2d Cir. 2013)).

235. *Id.* at 201.

236. *Id.* at 202–03 ("[A]uditing and correcting for bias is not only legally permissible, it also represents the type of voluntary compliance effort that Supreme Court precedents have long endorsed.").

237. Ajunwa, *Auditing Imperative, supra* note 30, at 652 ("[A]ny affirmative duty of care imposed on an employer should carry also an auditing imperative for automated hiring systems."). The author relies upon David Oppenheimer's work on negligent discrimination, applying the notion of duty to employers to prevent discrimination and bring tort law into employment sphere. *Id.* at 652–55 (citing David Benjamin Oppenheimer, *Negligent Discrimination*, 141 U. PA. L. REV. 899 (1993)).

238. *Id.* at 624.

239. *Id.* at 625 (emphasis in original).

240. *Id.* at 629.

241. Students for Fair Admissions, Inc. v. President of Harv. Coll., 600 U.S. 181, 215 (2023) (citing Franks v. Bowman Transp. Co., 424 U.S. 747, 763 (1976)) ("When it comes to workplace discrimination, courts can ask whether a race-based benefit makes members of the discriminated class 'whole for [the] injuries [they] suffered.'").

242. *EEOC Guidance, supra* note 2 (emphasis added).

D. Final Recommendations

Employers using AEDTs should be aware that they must assess the tools in context—considering the surrounding circumstances, including programming, people, and products.[243] Users from marginalized groups, such as people of color, should be aware that they are under additional pressure to provide data and permit retention because doing so can provide the only mechanism, though imperfect, for addressing fairness between racial, ethnic, and other groups.

So, what can employers do?

(1) Open the AEDT development and deployment processes to diverse creators, testers, and end-users.

(2) Watch for discriminatory impacts on unrepresented and underrepresented groups.

(3) Train your AEDTs. Audit them. Adjust them. Then, deploy them.

Potential employees should be aware of the *double bind*—the risks and rewards of guarding personal data. Applicants should ask whether a potential employer is using AEDTs, and if so, ask how and in what context. They should question employers about bias and ethical risk assessments, audits, and their outcomes. Applicants should inquire about opt-out provisions and how to request the return or destruction of their individual data after the job search process ends. Data submitted to one potential employer may be stored and even shared with any and every future employer, unless and until there are more guardrails in place to curtail data abuses. But applicants should also recognize that blinding their data can hinder antidiscrimination goals.

Happy (Job) Hunting.

---

243. *See* Hasan et al., *supra* note 145, at 17 ("Just as algorithms should not be assessed independently of their context and use, so assessments should not focus on the code, but on the interplay between the code, its outputs, and the users.").