CURBING WIDESPREAD DISCRIMINATION BY ARTIFICIAL INTELLIGENCE HIRING TOOLS: AN EX ANTE SOLUTION

BRADFIELD E. A. BIGGERS*

Abstract: Artificial intelligence (AI) acolytes insist AI will pave the way for idealistic workplace hiring. Their idea is that AI can circumvent common human frailties, such as prejudices and narrow-mindedness, to achieve social and economic equality. Yet inherent biases permeating into AI algorithms stifle society’s goal of promoting equality in the workplace. Despite a growing awareness of AI’s potential for prejudice, politicians are slow to support substantial legislation regulating AI. Nevertheless, the judiciary will inevitably be tasked with addressing discrimination by AI technologies through an unprecedented construction of Title VII of the Civil Rights Act of 1964. Concerned with how this ex post regulation of AI hiring tools will dispense the costs associated with AI discrimination, this Essay advocates for a state level ex ante solution that seeks to equalize and protect society from discrimination by AI algorithms.

INTRODUCTION

Artificial intelligence (AI) maintains its status as an international buzzword because of advocates’ fantastic promises and critics’ warnings.¹ For example, some commentators suggest that AI is humanity’s savior while others suggest that it is fourth horseman of the apocalypse.² Nevertheless, AI is likely less polarizing than commentators suggest, as many optimistic companies are doubling-down on AI technology.³ Organizations including Chase, Hilton, AT&T, and Amazon are rapidly implementing AI hiring tools to, among other things, sort thousands of applications quickly to review and interview candidates.⁴ Companies value AI hiring

---

³ See Dowd, supra note 1 (finding that, in spite of Elon Musk’s apocalyptic convictions about AI, companies continue to utilize AI technologies to improve their businesses).
tools not only for their efficiency, but also for their ostensible impartiality.\textsuperscript{5} Machine learning algorithms are personified as the ultimate unbiased deciders.\textsuperscript{6} Notwithstanding these idealistic praises and anticipated gains, companies and legal scholars are identifying that AI hiring tools are more susceptible to inherent bias and discrimination than anticipated.\textsuperscript{7}

The advent of AI hiring technologies prompted a wave of scholarly discussions as to how these tools will fit into the employment discrimination landscape of Title VII of the Civil Rights Act of 1964 (Title VII).\textsuperscript{8} Title VII provides, inter alia, judicial relief to job applicants who can prove that a prospective employer discriminated against him or her on the basis race, color, religion, sex, or national origin.\textsuperscript{9} Title VII’s disparate treatment\textsuperscript{10} and disparate impact\textsuperscript{11} inquiries will likely cover discriminatory automated hiring decisions, but victims of AI discrimination may have difficulty satisfying their burdens.\textsuperscript{12} Scholars anticipate that demonstrating Title VII requirements, such as an employer’s discriminatory intent and knowledge, may be difficult if an algorithm makes the hiring decisions.\textsuperscript{13}

Notwithstanding these uncertainties, the ex post nature of Title VII’s inquiries puts society in a reactive position to protect itself from harms concealed within a black box.\textsuperscript{14} Employers will discover that AI hiring tools can actually inflict

\textsuperscript{5} Cynthia Dwork & Deirdre K. Mulligan, It’s Not Privacy, and it’s Not Fair, 66 STAN. L. REV. ONLINE 35, 35 (2013).

\textsuperscript{6} Id.


\textsuperscript{8} See, e.g., Matthew Scherer, AI in HR: Civil Rights Implications of Employers’ Use of Artificial Intelligence and Big Data, 13 THE SCI-Tech LAWYER 12, 13 (2017) (discussing the difficulty of using Title VII to address AI discrimination).


\textsuperscript{11} “Disparate impact” occurs when a facially neutral employment practice systematically has an adverse effect on a protected class and is not supported by a legitimate business justification. § 2000e-2(a)(1); Alessandra, supra note 10, at 1756.

\textsuperscript{12} Solon Barocas & Andrew D. Selbst, Big Data’s Disparate Impact, 104 CALIF. L. REV. 671, 700–01 (2016) (discussing that the uncertainty of Title VII’s “knowledge” requirement, as applied to data mining, may make it difficult to hold employers liable under Title VII).

\textsuperscript{13} Id.

\textsuperscript{14} See infra Section II (providing an overview of how AI hiring tools discriminate against classes of people protected by Title VII of the Civil Rights Act of 1964).
historic prejudices into the modern workforce, but lawmakers should not wait for public injury before protecting job applicants from discriminatory algorithms.¹⁵ States should instead proactively shield their workforces from discrimination through *ex ante* regulation of AI hiring tools.¹⁶ This Essay advocates for state level *ex ante* regulation of AI hiring tools.¹⁷ Part I provides an overview of AI hiring technology.¹⁸ Part II discusses how these tools discriminate.¹⁹ Part III includes an example of such discrimination at a large international corporation.²⁰ Finally, Part IV discusses what protections should be included in a state-level regulation.²¹

I. THE TECHNOLOGICAL UNDERPINNINGS: AI, DATA MINING, & MACHINE LEARNING

Broadly stated, three categories of technologies animate AI hiring tools: AI, data mining, and machine learning.²² These categories are subsets of data science, which interact much like a set of misshapen Russian nesting dolls.²³ Although the term “artificial intelligence” may conjure the mystique of an abstruse concept, AI is simply any technology designed to mimic and automate intellectual tasks usually performed by humans.²⁴ AI technologies do not necessarily include data mining

---

¹⁵ See infra Part III (using Amazon’s discriminatory AI hiring tool to emphasize the importance of AI oversight and regulation).
¹⁷ See infra Part III.
¹⁸ See infra Part I.
¹⁹ See infra Part II.
²⁰ See infra Part III.
²³ Heiler, supra note 22.
²⁴ CHOLLET, supra note 21, at 4.
and machine learning. For example, companies often use AI customer service hotlines or Internet chats that use the façade of human interaction and programmed responses to guide customers through a maze of automated responses.

In contrast, data mining and machine learning are more narrowly concerned with supporting the human decision-making process by organizing and characterizing a previously unintelligible quantity of data. Data mining is the process of using algorithms to identify statistical relationships in vast datasets (Big Data). These statistical relationships manifest as useful patterns that later decision making can rely on. A model, a group of statistical relationships found in a dataset, can automate the process of determining the value, status, or likely outcome of a situation, person, or entity.

Machine learning takes data mining’s identification of statistical relationships in Big Data one step further. Machine learning extrapolates lessons, or rules, from programmer selected examples about what data attributes can serve as proxies for certain metrics and desired outcomes. These examples train the algorithm to devise its own methods of achieving a specific goal by relying on lessons learned to navigate unprecedented data configurations. For instance, a machine learning algorithm can analyze vast datasets of call logs and end-user reports of spam callers to generate lessons. These lessons can then enable a call-blocker mobile phone app to identify and block likely incoming spam calls.

II. HOW AI HIRING TOOLS DISCRIMINATE

The complexities of data mining and machine learning algorithms make it difficult to trace intentional and unintentional prejudice through AI hiring tools. This Part simplifies four foundational aspects of these technologies and provides illustrations of how these mechanisms can contribute to the mistreatment of

---

25 See id. (providing an example of AI technology without machine learning, specifically an AI chess program that uses previously encoded chess move responses to appear human-like).
26 See id. (providing an example of an AI chess program that uses hardcoded rules written by programmers to react to human responses, analogous to how AI customer service hotlines and chats respond to calling customers).
28 Heiler, supra note 22.
29 Id.
30 Barocas & Selbst, supra note 12.
31 Id. at 678.
32 Id.
33 CHOLLET, supra note 22, at 5.
34 See id. at 6–7 (demonstrating how a machine learning algorithm navigates unprecedented data with an example of an AI program using lessons learned from input data images to classify new photographs).
35 See id. (demonstrating that machine learning algorithms can identify and classify unfamiliar data based on lessons learned from training data).
36 Barocas & Selbst, supra note 12, at 677 (noting that “[u]nlike more subjective forms of decision making, data mining’s ill effects are often not traceable to human bias….”).
protected classes.\textsuperscript{37} Specifically, Section A discusses target variables and class labels.\textsuperscript{38} Section B discusses the collecting and labeling of training data.\textsuperscript{39}

\subsection*{A. Difficulty in Defining Target Variables & Class Labels}

AI programmers must convert complex employer objectives into code that directs algorithms to sort qualitative and quantitative characteristics of applicants.\textsuperscript{40} Data mining and machine learning scholarship recognizes this process as defining target variables and class labels.\textsuperscript{41}

Target variables are the desired outcomes or values that programmers seek to uncover or predict from Big Data.\textsuperscript{42} Target variables can range from binary outcomes, like whether a phone call is spam or not, to more abstract concepts, such as how one defines a “good” employee.\textsuperscript{43} In contrast, class labels organize each possible degree of satisfying a target variable into discrete categories.\textsuperscript{44} Class labels may be as straightforward as classifying library books as checked-out or not, or as complex as categorizing employees based on an employee’s predicted sales revenue.\textsuperscript{45} Target variables tell algorithms what the desired outcome is, whereas class labels designate the degree at which someone or something meets that target variable.\textsuperscript{46}

The process of defining target variables and utilizing class labels to distinguish candidates may inadvertently penalize a protected class.\textsuperscript{47} For instance, haphazardly choosing a target variable that is a systematically prevalent or absent characteristic of a particular group may affect how the algorithm treats such candidates.\textsuperscript{48} Accordingly, setting an observable characteristic that is typically absent in Black individuals as your target variable may result in a model that negatively classifies Black candidates.\textsuperscript{49} Another illustration would be if an employer defined its target variable of loyalty by how long employees remained at their prior positions.\textsuperscript{50} There, the hiring system would likely classify women as less loyal applicants, because historical data trends suggest that women are more likely than men to leave the workforce temporarily to care for a child or family member.\textsuperscript{51}

\subsection*{B. Impressionable Training Data: Labeling Examples & Data Collection}

\begin{itemize}
  \item \textsuperscript{37} Id. at 677–87.
  \item \textsuperscript{38} See infra Part II.A.
  \item \textsuperscript{39} See infra Part II.B.
  \item \textsuperscript{40} Id. at 679.
  \item \textsuperscript{41} Id. at 677–78.
  \item \textsuperscript{42} Id.
  \item \textsuperscript{43} Id. at 679.
  \item \textsuperscript{44} Id.
  \item \textsuperscript{45} Id.
  \item \textsuperscript{46} Id. at 677.
  \item \textsuperscript{47} Id. at 680.
  \item \textsuperscript{48} Id.
  \item \textsuperscript{49} Id.
  \item \textsuperscript{50} Id.
\end{itemize}
A machine learning algorithm learns from the examples identified by data mining. These examples allow algorithms to develop a model to guide future decision making. Opportunities to inadvertently spread preexisting biases arise when compiling training data and labeling examples to train a machine learning algorithm.

Labeling values or outcomes in Big Data as examples and assigning them class labels sets a rigid model that an algorithm will use to interpret subsequent data. This model teaches an algorithm from examples of past outcomes or values, which is problematic when the assumptions underlying the examples change. More importantly, labeling examples from a decision maker who systematically introduced biases into his or her decisions will result in data mining that necessarily infers rules exhibiting the same prejudices. Without alleviating the inherent biases or accounting for the assumption underlying these examples, an AI hiring tool taught from these examples will continue to imbed prior inequities into future hiring decisions.

Data collection often falters when the information gathered is not representative of every subsection of race, class, and gender in proportions that reflect society. For example, Big Data comprised of past employee performance evaluations that do not include disabled employees may corrupt subsequent data mining efforts. If these evaluations include criteria that fail to accommodate disabled employees, otherwise qualified disabled applicants will be adversely affected. Likewise, the overrepresentation of a group will inflate its attributes. Similar to labeling examples, data collection might also be tainted with nearly imperceptible bias embedded in prior decisions. Employers who feed their own proprietary employment data into their AI hiring tools often face these issues of disproportionate representation and unrecognized inherent biases. Overlooked practices, such as an absence of minorities in prior hiring decisions, or a supervisor’s consistent.

---

52 See Amanda Levendowski, How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem, 93 WASH. L. REV. 579, 590–93 (2018) (discussing how a supervised AI algorithm learns by describing the process of teaching an algorithm to decode frenetically scribbled addresses). These examples represent possible values or outcomes in Big Data. Id.
53 Barocas & Selbst, supra note 12, at 680.
54 See Levendowski, supra note 52, at 591–92 (explaining how bias can skew the results of machine learning algorithms).
56 Id. at 12.
57 Barocas & Selbst, supra note 12, at 682.
58 Id.
59 See id. at 684 (noting that “[e]ven a dataset . . . of consistently high quality can suffer from statistical biases that fail to represent different groups in accurate proportions.”).
60 Pauline T. Kim, Data-Driven Discrimination at Work, 58 WM. & MARY L. REV. 857, 876–78 (2017) (capturing less data about a particular group renders a model less accurate when applied to that group).
61 Id.
62 Barocas & Selbst, supra note 12, at 687.
63 Id. at 682.
64 Kim, supra note 60, at 876–78.
hyercritical reviews of female employees, may negatively impact an organization’s earnest goal of promoting diversity.\textsuperscript{65}

III. THE GRAVITAS OF AI DISCRIMINATION: AMAZON’S MONSTER

The mechanisms of AI hiring tool technologies, together with their susceptibility to prejudice, can cause a company to further ingrain discrimination into the fabric of the global workforce.\textsuperscript{66} This is a particularly endemic issue for large corporations that may use biased AI hiring tools because of the scale of their applicant pool.\textsuperscript{67} To grasp the gravitas of large corporations utilizing biased algorithms, this Part discusses how an AI hiring tool at one of the world’s largest technological corporations contributed to the industry’s historic male bias.\textsuperscript{68}

In 2014, Amazon developed an experimental AI candidate scoring and ranking system poised as the “holy grail” of AI hiring tools.\textsuperscript{69} Amazon hoped this tool would fully automate its hiring process and drastically reduce its candidate search costs.\textsuperscript{70} To build its grail, the e-commerce juggernaut trained an algorithm using submitted resumes during a ten-year period.\textsuperscript{71} The goal was to develop a model that would classify and rank resumes.\textsuperscript{72} The AI hiring tool algorithm would then use this model to compare patterns present in the historical resume dataset to attributes in a new resume.\textsuperscript{73} The algorithm would then give that new resume a class label of one to five stars.\textsuperscript{74} Resumes with a one star class label were considered the least desirable, whereas resumes with a five star class label were considered the most desirable.\textsuperscript{75} This algorithm’s desired outcome, or target variable, was to identify five star resumes.\textsuperscript{76}

By 2015, Amazon aborted its mission of autonomous AI hiring because its tool developed discriminatory preferences that discounted female resumes.\textsuperscript{77} Amazon’s male-dominant workforce, coupled with the technology industry’s overall male bias, resulted in a disproportionate number of male resumes entering the

\textsuperscript{65} See id. (highlighting how large datasets tend to underrepresent, and thereby disadvantage, minority groups).

\textsuperscript{66} See supra Part I & Part II (discussing how AI hiring tools can reinforce the intentional and unintentional bias present in prior hiring decisions).

\textsuperscript{67} See Lauren Weber, Your Résumé vs. Oblivion, WALL ST. J. (Jan. 24, 2012), https://www.wsj.com/articles/SB10001424052970204624204577178941034941330 (highlighting the fact that 7.6 million job applications were submitted to Starbucks in 2011).

\textsuperscript{68} Dastin, supra note 21.

\textsuperscript{69} Id.

\textsuperscript{70} See id. (highlighting Amazon’s aspirations to completely automate its hiring process).


\textsuperscript{72} Id.

\textsuperscript{73} Id.

\textsuperscript{74} Dastin, supra note 21.

\textsuperscript{75} Id.

\textsuperscript{76} See id. (noting that “[t]hey literally wanted it to be an engine where I’m going to give you 100 resumes, it will spit out the top five, and we’ll hire those.”).

\textsuperscript{77} Id.
training data. Accordingly, this training dataset severely underrepresented desirable female resumes, producing a male-biased model that systematically penalized females. For example, resumes with words like “executed” and other words often used by men were elevated, whereas the term “women,” such as “Women in Law Empowerment Forum,” was penalized. And, graduates from at least two all-women’s colleges were automatically downgraded. Importantly, this algorithm used the results of its own prior classifications to improve the consistency of its decision making. Thus, this reinforcement of the algorithm’s male preference virtually guaranteed that its discrimination would be magnified and ingrained into the review of future candidates.

Fortunately, by regular auditing and monitoring, Amazon identified its algorithm’s male bias and removed the AI hiring tool from its hiring practices. Without Amazon’s oversight, its AI hiring tool would likely have marginalized countless qualified female applicants, further promoting male dominance in a disproportionately male industry. The message of Amazon’s monster is clear: companies need to be held accountable for monitoring and auditing AI hiring tools to prevent discrimination.

IV. REDISTRIBUTION OF AI HIRING BENEFITS: AN EX ANTE SOLUTION

In a strictly ex post Title VII regime, the balance of costs and benefits between employers utilizing AI hiring tools and the candidates being reviewed is fundamentally lopsided. As employers enjoy the efficacy of these tools, this ex post regime foists the costs of monitoring AI algorithms and identifying their biases regarding victims of discrimination. Even so, commentators are concerned that Title VII may not be an adequate redressive vehicle for victims of AI discrimination because of the lack of transparency in AI hiring algorithms. This lack of transparency makes it difficult for victims to demonstrate discriminatory intent, knowledge, and other elements of Title VII’s disparate treatment and disparate impact

---

79 Lavanchy, supra note 72.
80 Dastin, supra note 21; Goodman, supra note 78.
81 Dastin, supra note 21.
82 Lavanchy, supra note 72.
83 Id.
84 Dastin, supra note 21.
86 See Goodman, supra note 78 (suggesting that Amazon’s AI hiring tool would have likely discriminated against countless female applicants if Amazon did not monitor its algorithm).
87 See Barocas & Selbst, supra note 12, at 694–714 (describing the uncertain Title VII liability landscape that demands plaintiffs uncover evidence of discrimination from unfamiliar, opaque algorithms).
88 Id.
89 See id. (highlighting the insufficiency of Title VII’s disparate impact inquiry in the context of AI discrimination).
inquiries.\textsuperscript{90} Considering the uncertainty of whether Title VII will be able to protect job applicants from discrimination adequately, companies have little financial incentive to install substantial preventative measures.\textsuperscript{91} To redistribute the costs and benefits of AI hiring technologies fairly, states should seek to impose \textit{ex ante} regulations on employers.

Despite a call for technology-specific regulations at the federal level, policy makers have made little progress.\textsuperscript{93} Nonetheless, state legislatures have been successful.\textsuperscript{94} Among the states instituting \textit{ex ante} technology regulations, Illinois is leading the regulations of AI hiring tools with its Artificial Intelligence Video Interview (AIVI) Act.\textsuperscript{95} The AIVI Act protects job applicants by requiring employers to provide video interviewees notice of the AI software and explain how the AI is scoring interviews.\textsuperscript{96} In addition, the Act limits what employers can do with the resulting video data.\textsuperscript{97} Ultimately, The AIVI Act indicates that states are willing to implement workforce protections from AI hiring tools, without imposing overly burdensome regulations that stiffle innovation.\textsuperscript{98}

As Amazon demonstrates, the corporate monitoring and auditing of AI hiring algorithms is a practical way to help prevent AI hiring tools from reinforcing biases in society.\textsuperscript{99} Businesses utilizing these tools should have internal corporate mechanisms to protect employee candidates from the intentional or unintentional prejudice of their algorithms.\textsuperscript{100} State-level \textit{ex ante} protections should require companies using AI hiring tools to retain uninterested committees to audit and monitor their algorithms for the disparate treatment of or impact on candidates.\textsuperscript{101} In doing so, these preventative measures redistribute costs of AI discrimination back to employers—the stakeholders that benefit the most from the technology.\textsuperscript{102} Further,

\begin{itemize}
  \item \textsuperscript{90} See id. at 700–01.
  \item \textsuperscript{91} See id. (finding that a lack of transparency in AI hiring tools may make it difficult for victims of discrimination to hold employers accountable for their algorithms).
  \item \textsuperscript{92} See Stephanie Bornstein, \textit{Antidiscriminatory Algorithms}, 70 ALA. L. REV. 519, 524 (2018) (highlighting how \textit{ex ante} regulations will cause employers to help prevent discriminatory injuries).
  \item \textsuperscript{93} See generally David McCabe, \textit{Congress and Trump Agreed They Want a National Privacy Law. It’s Nowhere in Sight.}, N.Y. TIMES (Oct. 1, 2019), https://www.nytimes.com/2019/10/01/technology/national-privacy-law.html (reporting that the concept of a federal privacy law that regulates, \textit{inter alia}, technology companies draws bipartisan support, but that no such legislation is in sight).
  \item \textsuperscript{96} Id.
  \item \textsuperscript{97} Id.
  \item \textsuperscript{98} See \textit{id.} (demonstrating that—at the state level—elected officials, companies, and employees can collaborate on \textit{ex ante} legislation to anticipate and prevent the uncertain harms of AI technologies).
  \item \textsuperscript{99} See supra Part III (discussing Amazon’s identification of its discriminatory AI hiring tool).
  \item \textsuperscript{100} See Kimberly A. Houser, \textit{Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making}, 22 STAN. TECH. L. REV. 290, 339 (2019) (acknowledging that auditing an AI hiring algorithm can delete or mitigate bias from its results).
  \item \textsuperscript{101} See \textit{id.} at 327 (recognizing that auditing AI algorithms and other monitoring measures can help prevent exposing job applicants to discriminatory algorithms).
  \item \textsuperscript{102} See Claire Glenn, \textit{Upholding Civil Rights in Environmental Law: The Case for Ex Ante Title VI Regulation and Enforcement}, 41 N.Y.U. REV. L. & SOC. CHANGE 45, 74 (2017) (discussing how \textit{ex}}
encouraging companies to confront their algorithms reduces the likelihood of discriminating against candidates and helps remove biases in the local workforce. In fact, by auditing and monitoring AI tools, employers can promote workplace diversity, a factor shown to boost balance sheets, increase workforce morale, and strengthen connectivity.

CONCLUSION

AI hiring technology is becoming indispensable in increasingly competitive and interconnected marketplaces. This technology not only expands applicant pools while decreasing search and interview costs, but also presents an opportunity for employers to reap the benefits of workplace diversity if the AI is appropriately monitored. In contrast, employers blindly following AI hiring decisions may inadvertently magnify intentional and unintentional biases. To optimize the benefits of AI hiring tools while combatting its biases, state legislatures should demand that employers adopt internal auditing and other preventative measures.


ante environmental regulations shift the burden of monitoring the production of pollutants from affected communities to the companies polluting). By requiring employers to assume bias prevention costs, employees are less likely to be discriminated against, and, thus, employers will likely spend fewer resources on litigating discrimination actions. See Houser, supra note 100, at 339 (arguing that auditing algorithms will allow programmers to identify and reduce its biases).

See Houser, supra note 100, at 344 (discussing how the monitoring of AI hiring tools can reduce the discriminatory effects of biased training data and produce a more equitable workforce).

See id. at 294–96 (discussing how AI tools can generate more effective, efficient, and diversity-conscious hiring decisions).

See id. at 295 (emphasizing that employers using properly monitored AI hiring tools often hire more women); see also Daniela M. de la Piedra, Diversity Initiatives in the Workplace: The Importance of Furthering the Efforts of Title VII, 4 THE MODERN AM. 43, 45 (2008) (describing the financial and social benefits of workplace diversity, which include a reduction in employee turnover, stronger client relations, and greater profits).