For much of the last decade, scholars have highlighted both the role of AI and machine learning in reproducing inequality and the practical litigation difficulties such technologies create for proving intent. What the scholarship largely overlooks is the role of a handful of vendors who convert idiosyncratic biases into systematic, industry-wide barriers to employment opportunity. This capacity to scale discrimination in a way that may exclude certain groups from large segments of the labor market is a unique danger of AI vendors that warrants special scrutiny from Congress and regulators. Although underestimating the potential for harm, many scholars have also underestimated the effectiveness of current law in addressing these problems. Thus, the EEOC and Congress should focus enforcement energy on AI vendors—through commissioner charges in the short term and through new legislation and regulation designed to meet the evolving challenges of AI hiring tools in the long term.
INTRODUCTION

Amazon decided in 2014 to use its extensive technical expertise, vast troves of data, and functionally unlimited resources to automate the recruitment and evaluation of job applicants. A team of a dozen engineers built hundreds of machine learning models, each focused on a different set of job functions. In almost no time at all, the models were able to score candidate resumes based on more than 50,000 terms that showed up in the resumes of employees hired in the previous decade. Things seemed great until Amazon’s engineers realized, a year into the project, that the program

---

1 Jeffrey Dastin, Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women, REUTERS (Oct. 10, 2018, 7:04 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MKo8G [https://perma.cc/59PN-JDR3]. Amazon asserts that the tool was never put into use, but some sources within the company suggest that “recruiters looked at the recommendations generated by the tool when searching for new hires, but never relied solely on those rankings.” Id.

2 Id.

3 Id.
had “taught itself that male candidates were preferable.” The models prioritized candidates with resumes that used words disproportionately found on men's resumes (e.g., “executed” and “captured”) and deprioritized candidates with resumes that included the word “women’s” (e.g., “women’s chess club captain”) or featured all-women’s colleges.

In the year after Reuters broke the story about Amazon's failed attempt at automated hiring, dozens of think pieces debated the merits ad nauseum, with one concluding that “if a company like Amazon can’t pull [automated hiring] off without problems, it’s difficult to imagine that less sophisticated companies can.” And yet, every single day, companies far less resourced than Amazon rely on artificial intelligence (“AI”) to make hiring and firing decisions. Most of them, including a third of Fortune 100 companies, are aided in this endeavor by an ever-expanding industry of AI hiring vendors. Unfortunately, there is little to suggest that AI vendors have solved the problems that Amazon could not. Instead, each vendor sells its AI tool, including whatever biases may be baked into its models to dozens, hundreds,
or thousands of customers. In doing so, each vendor replicates and systematizes discrimination in an unprecedented manner.

Title VII was designed to “strike at the entire spectrum” of discriminatory conduct in employment, but it appears to be falling short in this context, as no plaintiff has successfully brought an employment discrimination suit based on the use of AI hiring software. Nonetheless, the technology has generated considerable attention from activists, government agencies, media, and scholars.

This Comment proceeds in four parts to analyze the normative problems, review the legal frameworks, and propose potential solutions to address AI-driven employment discrimination. First, because there are as many definitions of AI as there are law review articles discussing AI, Part I defines AI in the employment context before describing how AI tools are currently being used in the workplace.

The remaining parts address three key gaps in the algorithmic discrimination scholarship to date. First, existing antidiscrimination scholarship on AI in hiring focuses, explicitly or implicitly, on two chief harms of AI tools: (1) making employment decisions based on job-irrelevant correlations in a way that results in discriminatory outcomes and (2) creating new barriers for legal redress. These concerns are substantial and well-documented, but the scholarship overlooks a third significant harm of AI hiring: systematizing discrimination and thereby excluding people or communities from large swaths of the labor market. Although the

---

10 Meritor Savs. Bank v. Vinson, 477 U.S. 57, 64 (1986). Although this Comment focuses on Title VII as amended by the Civil Rights Act of 1991, much of the analysis would apply to discrimination based on age or disability, as prohibited by the Age Discrimination in Employment Act (ADEA) and the Americans with Disabilities Act (ADA).


15 See Pauline T. Kim, Data-Driven Discrimination at Work, 58 WM. & MARY L. REV. 857, 881 (2017) (articulating problems with AI discrimination, including both that “the algorithm is relying on a factor that has a discriminatory effect but is not actually connected to job performance” and that “[w]hen such a model is relied on to screen or rank applicants, it obscures the basis on which employers are making ultimate employment decisions”).
systematizing effect of AI has not been discussed at length in the employment discrimination literature, scholars have identified it as a distinctive concern of AI-driven decisionmaking in other contexts. In Part II, I seek to build on these scholars’ work by considering the normative problem of systematcity in the context of employment discrimination, where the concern is particularly pressing.

Second, most of the current proposals for reform focus on improvements to the technology or fundamental changes to antidiscrimination law. I suggest that this is because many scholars underestimate the potential effectiveness of existing disparate impact doctrines. Thus, in Part III, I examine the caselaw and argue that AI-driven discrimination can be effectively challenged under a disparate impact theory.

Third, scholarship to date has generally set aside the question of third-party liability. However, if one is concerned about systematicity, third-party...
liability becomes a central focus, because few actors have a bigger role in systematizing employment discrimination than the vendors who create and sell AI hiring tools to hundreds of employers. Thus, even if current disparate impact doctrines are adequate to hold employers accountable for AI-driven discrimination, the lack of liability for most third parties under Title VII and other nondiscrimination statutes presents a considerable problem. In Part IV, I propose strategies that the EEOC could pursue immediately, as well as long-term reforms Congress should enact, to address this problem.

Finally, a note on terminology. Resolving the thorny debates around the definitions of “bias” and “discrimination” is beyond my scope here. Nonetheless, for a working definition, I embrace Pauline Kim’s conception of “classification bias,” which she defines as occurring “when employers rely on classification schemes, such as data algorithms, to sort or score workers in ways that worsen inequality or disadvantage along the lines of race, sex, or other protected characteristics.”20 Similarly, I am fundamentally concerned with an antisubordination theory of discrimination, which flows intuitively from Kim’s use of classification bias.21 Accordingly, when I discuss biased data or technology throughout this Comment, I am not using that term in a technical sense to suggest that algorithms are statistically unsound; instead, I’m referring to a tendency to further the subordination of marginalized groups.22

I. ARTIFICIAL INTELLIGENCE AT WORK

The design and operation of artificial intelligence involves considerable complexity, which is undergirded by robust and technical scholarly literature, most of which is beyond the scope of this Comment. However, identifying basic definitions, building blocks, and human decision points is essential to understanding both how AI tools facilitate discrimination and how the law can address such harms. At the broadest level, Joshua Kroll’s straightforward maxim is compelling: “AI is just automation.”23 The AI tools used in hiring are generally built with a combination of algorithms, machine learning, and training datasets selected by employers and vendors to automate employment processes previously performed by humans.

20 Kim, supra note 15, at 866.
21 See Kim, supra note 15, at 891 (“[A]ntisubordination theory . . . aims to promote equality by redressing structures and practices that disadvantage historically subordinated groups, regardless of whether the employer expressly or intentionally relied on race or other categories . . . .”).
22 For a discussion of the varying conceptions of bias in AI, see Mayson, supra note 16, at 2231-33.
A. The Building Blocks of Machine Learning

First, algorithms are an essential element of all AI tools.\textsuperscript{24} An algorithm is a formula or set of rules dictating “procedures for transforming input data into a desired output, based on specified calculations.”\textsuperscript{25} Algorithms aren’t new and don’t necessarily require a computer; they include “simple point-based scoring systems [that] are used . . . to automate credit decisions, rate recidivism risk, and make clinical medical decisions such as prioritizing vaccine administration in the COVID-19 pandemic response.”\textsuperscript{26} In practice, most of the AI tools discussed here use big data and machine learning and thus involve considerably more complexity than point-based scoring systems.

Big data and machine learning have fundamentally transformed algorithms from simple formulas into complex and opaque decisionmaking systems.\textsuperscript{27} Big data can refer to a large corpus of data or the process of data mining to gather such a corpus. Through machine learning, a software program uses that data to develop a predictive algorithm, which can then be applied to new data to predict a given outcome, like employment success.\textsuperscript{28}

\textsuperscript{24} For the sake of brevity, I refer simply to “AI hiring tools” and “AI-driven discrimination” throughout this Comment. But I note for clarity that the tools I describe here—i.e., the tools on the market—are almost all built on supervised machine learning, which relies on human labeling of training data to guide the program’s learning. See Jennifer Alsever, How AI Is Changing Your Job Hunt, FORTUNE (May 19, 2017, 6:30 AM), https://fortune.com/2017/05/19/ai-changing-jobs-hiring-recruiting [https://perma.cc/NUK7-C7Y7] (“Most of the software available today . . . [is] what’s called ‘supervised’ learning.”). For a simple example of supervised learning, developers might label certain resumes as belonging to “successful” employees and others as belonging to “average” or “marginal” employees, and then direct the program to learn how to predict future success by looking at applicants’ resumes. This is different both from unsupervised learning, which discerns patterns in very large, unlabeled datasets, and from a rules engine, which uses a set of if–then statements to direct an outcome. David F. Jacobs & Fleming E. Keefe, Attorney’s Guide to AI, JAG REP., Oct. 2021, at 2–3, https://www.jagreporter.af.mil/Portals/88/2021%20Articles/Documents/20211027%20Jacobs.pdf [https://perma.cc/G8BA-LYAY]. Although it is difficult to imagine how current unsupervised AI would be useful or legal in the employment context, the structure and auditability of a rules-based AI system means it likely presents fewer employment discrimination risks. Nonetheless, both are beyond the scope of this Comment.

\textsuperscript{25} Tarleton Gillespie, The Relevance of Algorithms, in MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 167 (Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot eds., 2014); see also Kroll, supra note 23 (describing credit rating algorithms as a set of “clear rules that can be applied . . . quickly” and a vaccine distribution algorithm as a “formula [that] assigned points”).

\textsuperscript{26} Kroll, supra note 23.

\textsuperscript{27} See Anya E.R. Prince & Daniel Schwartz, Proxy Discrimination in the Age of Artificial Intelligence and Big Data, 105 IOWA L. REV. 1257, 1273 (2020) (“The scale of such training data has increased dramatically in recent years . . . [as] firms have increasingly come to rely on data secured from a broad number of external sources.”).

In the end, AI tools developed through machine learning involve two algorithms. To create the first algorithm, developers give the machine learning program instructions about what kind of predictive algorithm to develop. These instructions include the outcome to be predicted (the “target criteria” or “target variable”), the input variables for the program to draw on, and the scope of training or baseline data.²⁹ Armed with these instructions, the machine learning program explores the training data and creates the second algorithm—the screening algorithm.³⁰ This screening algorithm consists of a set of rules that the machine has inferred from the patterns it observed in the training data: “they are, quite literally, rules learned by example.”³¹ And those rules attempt to predict target criteria based on (superficially unrelated) candidate attributes.³²

B. Human-Selected Data Points

Relying on a “veneer of high-tech objectivity,”³³ AI tools often “enjoy an undeserved assumption of fairness or objectivity.”³⁴ But such deference overlooks the key decisions humans make at several points in the development and implementation of AI tools. Solon Barocas and Andrew Selbst highlight three key data points in the AI screening process that are particularly prone to introducing or reproducing bias: target criteria, training data, and candidate input data.³⁵

1. Target Criteria

Developers must define the attribute or outcome the AI tool is trying to predict—the target criteria.³⁶ In the employment context, the target criteria could potentially include objective measures of efficiency, productivity, or accuracy; high ratings on customer satisfaction surveys or supervisor evaluations; or any other desirable workplace outcome. The fundamental objective is always “good workers,” but employers (or vendors) must identify quantifiable aspects of being a “good worker” to incorporate as target criteria into a predictive algorithm.

---

³⁰ Id. at 132.
³² Id.
³³ Mayson, supra note 16, at 2221.
³⁴ Kroll et al., supra note 21, at 680.
³⁵ Barocas & Selbst, supra note 14, at 675.
³⁶ See Kleinberg et al., supra note 29, at 139 (“The choice of outcome is non-trivial even in the simplest of examples.”).
2. Training Data

The training data includes exemplars from which the AI tool will build its predictive algorithms. Most commonly drawn from past workers, the training data can be thought of “like a spreadsheet, with each row being someone whom the firm hired in the past. One or more columns would be measures of how each worker performed on the job. The remaining columns might capture application data that can serve as candidate predictors.”

Errors and gaps in training data increase the risk of discrimination because machine learning “helps to discover patterns that organizations tend to treat as generalizable findings even though the analyzed data only includes a partial sample from a circumscribed period.” The Amazon case study highlights this risk. By drawing on resumes that had been submitted to the company over the prior decade, Amazon curated a training dataset that was dominated by men and systematically underscoring women. This comports with Barocas and Selbst’s observation that when a company builds an AI tool with disproportionate representation of a particular class (in Amazon’s case, men), the resulting predictive model “may skew in favor of or against the over- or underrepresented class.”

3. Candidate Inputs

Candidate input data encompass the many variables known about candidates, which the screening algorithm evaluates for a correlation with the outcome measure. Importantly, candidate inputs need to overlap with the training data. If, for instance, an employer uses resumes and supervisor evaluations of successful employees as the training data, only using interview transcripts as the candidate input would not be successful.

It is exceptionally difficult to disentangle the risk of bias generated by candidate inputs from underlying societal bias. To illustrate this, assume a best-case scenario where an employer has identified perfectly accurate target criteria. If the employer uses resumes of employees who score well on that target criteria, it is likely the AI tool will find correlations with elite

---

37 Barocas & Selbst, supra note 14, at 680.
38 Kleinberg et al., supra note 29, at 134.
39 Barocas & Selbst, supra note 14, at 686.
40 Dastin, supra note 1.
41 Barocas & Selbst, supra note 14, at 686. Vendor-curated datasets only increase the risk of discriminatory false correlations by obfuscating differences among industries, job functions, and geographic regions.
42 See Mayson, supra note 16, at 2224 (“[Any prediction] distills patterns in past data and interprets them as projections of the future. Algorithmic prediction produces a precise reflection of digital data.”).
education, unpaid internships, or other inequitable career markers. And target criteria that are less than perfect, such that they reproduce implicit prejudices of supervisors or customers, will only increase this effect.

C. The Trend of AI Hiring

AI proponents promise hiring managers that AI will “reduce the time you spend reviewing unqualified applicants,” “help[] you make more confident hiring decisions and overcome unconscious bias,” “achieve an increase in recruiting staff productivity between 10 and 50 percent,” and “make better hiring decisions, faster.” One recent survey of large companies found that more than forty percent of U.S. employers use AI tools to screen and assess candidates during recruitment. Another survey from 2019 found that seventy-eight percent of human resources departments predicted using AI tools for talent acquisition by 2021. Today, companies employ AI tools to select who will see a job advertisement, screen candidates ahead of interviews, assess behavioral traits during interviews, evaluate the


44 Beyond the AI context, this sort of discrimination is often permitted under a business necessity defense. See infra subsection III.B.3 (discussing the business necessity defense to disparate impact liability). Nonetheless, for reasons discussed in Part II, we still have reason for concern in the AI context.


52 See Heilweil, supra note 13 (“[R]ecruiters are increasingly using AI to make the first round of cuts . . . “).

53 See Will Knight, Job Screening Service Halts Facial Analysis of Applicants, WIRED (Jan. 12, 2021, 8:00 AM), https://www.wired.com/story/job-screening-service-halts-facial-analysis-applicants [https://perma.cc/GMY6-ATKK] (detailing HireVue’s discontinued service that assessed
performance of current employees, and set wages for gig workers. It is not an exaggeration to say that AI tools can govern every phase and facet of the employment relationship.

Three kinds of businesses develop AI tools in the employment context. First, vendors of AI technology build and sell a full range of AI tools to employers. This includes both established companies like Oracle and Workday, and independent startups like HireVue, Gem, and SeekOut. Regardless of size, these vendors license proprietary AI software to employers as a subscription service and are the primary focus of this comment. The second category consists of platforms like Facebook and LinkedIn, which leverage machine learning to provide sourcing and recruitment services to employers, but don’t license AI tools directly to employers. Although some of the largest platforms also sell machine learning models, this discussion focuses on how employers use AI tools to assess candidates’ “behavior, intonation, and speech” based on webcam footage and ultimately “assign[ed] certain traits and qualities” to the candidate.


56 To be fair, some of these “start-ups” have multi-billion-dollar valuations and hundreds of clients. See, e.g., Taylor Soper, Latest Unicorn Sighting in Seattle: SeekOut Lands $115M to Expand Recruiting Software, GEEKWIRE (Jan. 12, 2022, 6:00 AM), https://www.geekwire.com/2022/latest-unicorn-sighting-in-seattle-seekout-lands-115m-to-expand-recruiting-software [https://perma.cc/V386-B3MH] (profiling SeekOut, an AI vendor with more than 1,000 customers valued at over $1 billion); Christine Hall, Recruiting Platform Gem Gains Unicorn Status with $100M Raise to Change the Way Companies Hire, TECHCRUNCH (Sept. 28, 2021, 9:00 AM), https://techcrunch.com/2021/09/28/recruiting-platform-gem-gains-unicorn-status-with-100m-raise-to-change-the-way-companies-hire [https://perma.cc/M6BQ-BT4M] (profiling Gem, an AI start-up with more than 800 customers and a valuation of $1.2 billion).

57 See, e.g., HIREVUE, EXPLAINABILITY STATEMENT 7, 12 (2022), https://webapi.hirevue.com/wp-content/uploads/2022/04/HV_AL_Short-Form_Explainability_ipager.pdf [https://perma.cc/NY8C-RPDF] (explaining how HireVue’s natural-language processing model is built on top of the model developed by Facebook and Google).
will keep the platform aspect of their businesses (i.e., when they distribute ads) separate from the vendor aspect of their businesses (i.e., when they sell software). Finally, employers that develop and deploy AI hiring tools in-house make up the third group of businesses.  

II. THE NORMATIVE PROBLEMS OF AI-DRIVEN DISCRIMINATION

Like the tools themselves, concerns about AI-aided hiring are not new, but both the use and criticisms of these tools are rapidly accelerating. In considering why AI-driven discrimination should prompt scholarly attention, policy change, and doctrinal reform, Jack Balkin's guiding questions from two decades ago remain instructive: “[w]hat features of human activity or of the human condition does a technological change foreground, emphasize, or problematize? And what are the consequences for human freedom of making this aspect more important, more pervasive, or more central than it was before?”

AI-driven discrimination presents a moral concern, warranting particular attention and doctrinal reform, on multiple fronts: it may increase the overall volume of discrimination, it may make discrimination harder to identify, it may limit remedies available to victims of discrimination, or it may systematize discrimination in a way that makes it uniquely harmful (or, of course, it may do a combination of these things).

---

60 See, e.g., Dastin, supra note 1 (discussing Amazon’s AI hiring tool, which, without having access to gender labels, “taught itself that male candidates were preferable” by relying on signals like activities and college enrollment).


62 A major repository of publicly reported AI incidents identifies more than four times as many incidents of AI bias and discrimination in 2020-2021 compared to 2015-2016. AIAAIC, AI, Algorithmic, & Automation Incident & Controversy Repository, GOOGLE SHEETS, https://docs.google.com/spreadsheets/d/1Bn53B4xzz1-Rgd88BBZlt0n_4rzLGxFADMIVWqPY/edit?usp=sharing (last visited Oct. 12, 2022). Additionally, mentions of AI and machine learning in the Congressional Record increased more than thirteen-fold between 2016 and 2020. LAURIE A. HARRIS, CONG. RSC. SERV., R46795, ARTIFICIAL INTELLIGENCE: BACKGROUND, SELECTED ISSUES, & POLICY CONSIDERATIONS 23 fg.3 (2021).

A. AI Can Make Discrimination More Common

Scholars debate whether AI increases the frequency of discrimination. A growing number of well-publicized incidents depict AI bias, but proponents of AI hiring tools point out that the key question is not whether AI tools discriminate, but how often they discriminate relative to human decisionmakers. Although some research suggests that such tools may discriminate less frequently than humans, the lack of robust or comprehensive data may owe, at least in part, to the transparency issues to be discussed in Section II.B. For the purposes of this Comment, I accept the arguments of AI proponents that the tools either have no effect on the frequency of discrimination or, perhaps, even a slightly ameliorative effect. But that should not be our only concern.

B. AI May Conceal Discrimination When It Occurs

Although some scholars have sought to minimize the importance of transparency, most agree that opacity is one of the chief dangers of AI-aided decisionmaking.

---

64 See AIAAIC, supra note 62 (documenting a four-fold increase in media reports of AI bias and discrimination between 2015-2016 and 2020-2021).


from AI vendors has exacerbated this problem.\textsuperscript{68} Still, this is the only harm to date that has resulted in a meaningful policy response from legislatures, with Maryland, Illinois, and New York City passing regulations to require greater transparency in the use of AI hiring tools.\textsuperscript{69}

C. AI May Limit Antidiscrimination Remedies

The automated nature of hiring decisions has significant potential to insulate employers from many disparate treatment claims,\textsuperscript{70} leaving victims of discrimination with either no viable claim or a disparate impact claim. Although I am more optimistic than some about the viability of disparate impact claims in the AI hiring context,\textsuperscript{71} the move from disparate treatment to disparate impact claims still creates difficulties for plaintiffs. First, the business necessity defense allowing for disparate impact claims is much more generous to defendants than the bona fide occupational qualification (BFOQ) defense available in the disparate treatment context, allowing discriminatory employers to fend off more claims.\textsuperscript{72} Second, even if a plaintiff prevails on the more difficult disparate impact claim, compensatory damages aren’t available.\textsuperscript{73}

D. AI Can Systematize Harm

Finally, AI-driven discrimination fundamentally inflicts more harm than human discrimination because its harms are uniquely systematic. Although

\textsuperscript{68} See Muñoz v. Orr, 200 F.3d 291, 307 (5th Cir. 2000) (upholding the district court’s decision to deny plaintiffs access to an algorithm used for promotions).


\textsuperscript{70} See Barocas & Selbst, \emph{supra} note 14, at 701 (concluding that, because of the doctrine’s focus on intent, “disparate treatment doctrine does not appear to do much to regulate” discriminatory AI).

\textsuperscript{71} See infra Section III.B.

\textsuperscript{72} I discuss the business necessity defense in more depth in Part III. Briefly, \emph{International Union v. Johnson Controls, Inc.} underscores the difference in available defenses. There, the district court had analyzed a Title VII claim under the disparate impact framework and found that the facts supported a business necessity defense. 499 U.S. at 183. The Supreme Court, however, determined that the case called for a disparate treatment analysis, which only allows for a BFOQ defense and not a business necessity defense, and ultimately held that the facts could not support the more stringent BFOQ defense. Id. at 200, 206.

\textsuperscript{73} See 42 U.S.C. § 1981a(a)(1) (excluding an “employment practice that is unlawful because of its disparate impact” from the statute’s authorization of punitive and compensatory damages).
not often discussed in the employment discrimination literature, law and technology scholars have noted that the standardization inherent to AI-aided decisionmaking serves to amplify previously idiosyncratic harms.\textsuperscript{74} In the employment context, this systematization can happen in two interlocking ways: through vendors that sell the same AI tool to many companies and through the reuse of training data and algorithmic models. Although several scholars have observed the distinctive ability of AI tools to replicate biases that exist in society, none have meaningfully considered the unique role of AI vendors in this process.\textsuperscript{75}

Unlike Amazon's home-brewed tool, which would have only impacted candidates applying to Amazon, many of the top AI vendors have hundreds of clients.\textsuperscript{76} HireVue alone deploys its AI tool for one million interviews in some months.\textsuperscript{77} Exacerbating the problem, AI vendors often rely on public training data or common machine learning models, so issues with these datasets and tools reoccur across different companies.\textsuperscript{78}

\textsuperscript{74} See Pascal D. König, Dissecting the Algorithmic Leviathan: On the Socio-Political Anatomy of Algorithmic Governance, 33 PHIL. & TECH. 467, 474 (2020), https://doi.org/10.1007/s13347-019-00363-w [https://perma.cc/35TM-5E3T] (coining the term "algorithmic Leviathan" to describe the aggressive standardization function to machine-learning-driven decisionmaking); Creel & Hellman, supra note 16, at 35 (discussing the potential systemic exclusion that would follow from standardizing arbitrary decisionmaking); Mayson, supra note 16, at 2280-81 (observing that, in the criminal legal context, the capacity for algorithmic assessments to inflict harm is on a much larger scale than individualized subjective assessments).

\textsuperscript{75} See, e.g., Ajunwa, The Paradox of Automation as Anti-Bias Intervention, supra note 19, at 1692 (acknowledging that there may be some distinctions between AI tools developed in-house versus those purchased from vendors, but ultimately concluding that “differences in hiring algorithms are less important than the fact that . . . AI in the form of machine learning algorithms has automated many work functions previously thought reserved for human judgment”).

\textsuperscript{76} For a discussion of the substantial client base of AI vendor start-ups, see notes 8 & 57, supra, and accompanying text(describing the substantial client base of AI vendor start-ups).


\textsuperscript{78} See HireVue, EXPLAINABILITY STATEMENT, supra note 59, at 7, 12 (explaining how HireVue’s natural-language processing model is built on top of the model developed by Facebook and Google); Bernard Koch, Emily Denton, Alex Hanna & Jacob G. Foster, Reduced, Reused & Recycled: The Life of a Dataset in Machine Learning Research 7 (2021) (unpublished manuscript), https://arxiv.org/pdf/2112.01716.pdf [https://perma.cc/BNG3-KVSJ] (finding that datasets created by just a dozen institutions are used for more than fifty percent of machine learning projects). Some studies have found pervasive errors in the commonly used natural-language processing models that are essential elements of hiring tools focused on social media scraping, resume screening, and interview analysis. See Khari Johnson, AI Researchers Create Testing Tool to Find Bugs in NLP from Amazon, Google, and Microsoft, VENTUREBEAT: THE MACHINE (July 9, 2020, 9:10 AM), https://venturebeat.com/2020/07/09/ai-researchers-create-testing-tool-to-find-bugs-in-nlp-from-amazon-google-and-microsoft [https://perma.cc/E55R-NP3G].
As a result, the scale of harm inflicted by these vendors is several orders of magnitude greater than any individual company. And, importantly, because vendors homogenize harm by using the same data and models across hundreds or thousands of customers, vendor-provided tools produce harmful effects above and beyond any harm inherent in the AI tools themselves.⁷⁹

To illustrate this, consider a variation on the Amazon story: if hiring managers in the Amazon Prime engineering department have internalized misogyny such that they deprioritize women with certain backgrounds, like attending women’s colleges or playing on women’s team sports, that creates a moral and legal problem for the department and the candidates who apply. It is a problem that employment discrimination law should address, but in the meantime, those candidates may still have opportunities elsewhere within the company and at other companies. Importantly, those other departments need not lack any bias for this group of candidates to have an opportunity—they only need to have a different bias than the Amazon Prime hiring managers.

Then, imagine Amazon encodes the prejudices of those hiring managers into an in-house AI hiring tool by relying on its existing workforce for training data. Now, the program shuts out women who went to women’s colleges or played on women’s sports teams from jobs across the company. AI has magnified the harm by systematizing what were idiosyncratic biases and replicating them at a larger scale.

In a final scenario, instead of Amazon building the tool based on the entrenched biases of the Prime engineering team, imagine HireVue, Oracle, or another major AI vendor builds the tool. In that case, the harm is multiplied again—it is no longer contained to just one company but is instead reaching thousands. Not only does that mean more women are likely to be harmed by discrimination, it also means that the women who are harmed by it will have dramatically fewer job opportunities than when it was just one department at Amazon.

III. THE LEGAL SOLUTIONS & LIMITS

Once we understand the functioning of AI hiring tools and the harms such tools represent, the next question arises regarding the legal strategies available to address such harms. Title VII, the most robust and well-developed of the federal employment discrimination laws, prohibits two

---

⁷⁹ See Creel & Hellman, supra note 16, at 27 (“If the same algorithms produced by the same companies are uniformly applied across wide swathes of a single domain—hiring or lending, for example, thereby homogenizing decision outcomes—a person could be irrationally excluded from a significant number of important opportunities.”).
relevant types of conduct: disparate treatment and disparate impact.  

Under a disparate treatment theory, plaintiffs allege that an employer, union, or referral agency intentionally treated them differently based on a protected trait and that they were harmed by such differential treatment.  

Alternatively, under a disparate impact theory, plaintiffs are not required to establish motive; they need only prove that a policy or practice has an unjustifiable and disproportionate adverse impact based on a protected trait.  

Under both theories of liability, plaintiffs can sue employment agencies and labor unions, in addition to employers, but not other third parties.  

Although some scholars have argued that plaintiffs could use the disparate treatment framework to address AI-driven discrimination, for the reasons discussed below, disparate impact is likely to be a more effective tool for addressing the unique harms of AI vendors.

### A. Disparate Treatment Liability

The disparate treatment framework extends broadly, encompassing explicitly prejudiced conduct, facially neutral policies shown to be motivated by discriminatory intent, and classifications based on a protected trait, even if enacted without animus.  

This last type of disparate treatment is often called “rational” or “statistical” discrimination because it is not seen as driven by any kind of prejudice, but instead involves reliance on assumptions or true-in-the-aggregate statistics about the productivity or effectiveness of workers based on protected characteristics like race, gender, and age.  

Importantly, the law does not care whether such assumptions or statistics are correct—even

---

80 42 U.S.C. § 2000e-2; see also EEOC v. Abercrombie & Fitch Stores, Inc., 575 U.S. 768, 771 (2014) (stating that disparate treatment and disparate impact are the only causes of action under Title VII).


83 42 U.S.C. § 2000e-2(a)–(c) (describing employers, employment agencies, and labor organizations as the only actors liable under Title VII); see also 42 U.S.C. § 12111(2) (same for the ADA); 29 U.S.C. § 623(a)–(c) (same for the ADEA).

84 See Bornstein, supra note 17, at 526 (“[S]ome algorithmic discrimination may be challenged as disparate treatment using Title VII’s stereotype theory of liability.”).


when courts accept or assume the truth of employer claims that women, mothers, or elders are less productive, they still hold policies discriminating against such groups to be unlawful.87 Indeed, because rational discrimination is categorized as facially discriminatory disparate treatment, little opportunity exists for defendants to assert a legitimate nondiscriminatory reason for the practice, making a full range of remedies typically available to successful plaintiffs.88

The Fifth Circuit’s decision in *Diaz v. Pan American World Airways, Inc.* provides a helpful illustration of the rational discrimination doctrine in action and the problems AI presents in enforcing it.89 There, the airline explicitly preferred to hire women as flight attendants because it concluded that its customers strongly preferred women in such roles.90 The court accepted this conclusion as true, yet still found the practice to be unlawful disparate treatment, concluding, “[w]hile we recognize that the public’s expectation of finding one sex in a particular role may cause some initial difficulty, it would be totally anomalous if we were to allow the preferences and prejudices of the customers to determine whether the sex discrimination was valid.”91 Thus, even if a business’s customers will be dissatisfied without having women in public-facing customer service roles, employers generally cannot refuse to hire men for that reason.

---

87 See, e.g., City of L.A. Dep’t of Water & Power v. Manhart, 435 U.S. 702, 707-08 (1978) (finding an employer’s policy that required women to make larger monthly pension contributions to be unlawful despite accepting the fact that the cost of pensions was higher for women and the employer, therefore, only “treated its women employees differently from its men employees because the two classes are in fact different”); Phillips v. Martin Marietta Corp., 400 U.S. 542, 544 (1971) (holding that Title VII prohibits employers from having “one hiring policy for women and another for men—each having pre-school-age children”); Young v. United Parcel Serv., Inc., 575 U.S. 206, 233 (2015) (Alito, J., concurring) (“[t] does not matter whether the employer’s ground for the unfavorable treatment is reasonable; all that matters is the employer’s actual intent.”); 29 C.F.R. § 1604.2(a)(1)(iii) (2021) (providing that “the preferences of coworkers, the employer, clients or customers” cannot be used to justify disparate treatment). BFOQs exist as a relatively narrow statutory exception, which permits employment decisions based on religion, sex, national origin, or age when hiring based on the trait is “reasonably necessary to the normal operation of [a] particular business.” 42 U.S.C. § 2000e-2(e) (religion, sex, and national origin discrimination); accord 29 U.S.C. § 623(f)(1) (age discrimination). Courts read this BFOQ exception narrowly. See *Int’l Union*, 499 U.S. at 203 (“[I]n order to qualify as a BFOQ, a job qualification must relate to the ‘essence,’ or to the ‘central mission of the employer’s business.’” (citations omitted)).


89 *Diaz v. Pan Am. World Airways, Inc.*, 442 F.2d 385 (5th Cir. 1971). I’m hopeful that this case reflects anachronistic customer preferences, but it is not difficult to imagine circumstances in which implicit or explicit customer biases persist today.

90 *Id.* at 387 (recalling the trial court’s finding that a mixed gender hiring policy would result in a lesser average performance for passengers).

91 *Id.* at 389.
To demonstrate the limitations of disparate treatment liability in the AI hiring context, consider a similar situation where a manager is assisted by an AI hiring tool. Imagine that the tool had access to information about gender in the training data, or perhaps relied on names or other common gender signals in resumes. Either way, during training, the tool identifies that women receive particularly high customer satisfaction ratings—the specified target criteria. Accordingly, the tool prioritizes women applicants, predicting accurately that they are the most likely, in the aggregate, to receive high customer satisfaction ratings. The process and result almost identically mirror what Diaz expressly prohibited: the systematic classification of men as less desirable employees. The challenge is that this is no longer an intentional, facial policy of the employer, but the mens rea-free result of algorithmic decisionmaking. Through the alchemy of algorithms, an employer has transmuted clear-cut disparate treatment liability into a challenging disparate impact claim. This alchemical process is why most scholars hold pessimistic views about the potential success of disparate treatment claims against AI-driven discrimination.

B. Disparate Impact Liability

The disparate impact framework of employment discrimination functions as a safety net under disparate treatment liability, catching harmful and unjustifiable instances of discrimination that slip through for various reasons. Disparate impact liability allows plaintiffs to prevail when challenging employment practices without evidence of intent if (1) they demonstrate that a particular selection procedure “causes a disparate impact on the basis of” a

92 See Barocas & Selbst, supra note 14, at 695 (“The irony is that the use of protected class as an input is usually irrelevant to the outcome in terms of discriminatory effect. . . . Given a rich enough set of features, the chance that such membership is redundantly encoded approaches certainty.”); see also Dastin, supra note 1 (discussing Amazon’s AI hiring tool, which, without having access to gender labels, “taught itself that male candidates were preferable” by relying on signals like activities and colleges).

93 Diaz, 442 F.2d at 389.

94 See Charles A. Sullivan, Employing AI, 63 VILL. L. REV. 395, 405 (2018) (concluding that because AI “doesn’t have any ‘motives’ . . . [it] can’t be said to violate Title VII’s disparate treatment prohibition” (footnotes omitted)); Barocas & Selbst, supra note 14, at 701 (concluding that, except for when AI is used for the purpose of concealing discrimination, “disparate treatment doctrine does not appear to do much to regulate” discriminatory AI). But see Bornstein, supra note 17, at 525-26 (arguing that harmful AI tools could be challenged under an antistereotyping theory of disparate treatment).
protected trait, and (2) the employer fails to prove that the procedure is “job related for the position in question and consistent with business necessity.”

Much of the scholarship arguing that disparate impact liability poorly addresses AI-driven discrimination focuses on the business necessity defense and the correlation-reliant nature of both the defense and AI tools. However, such analyses tend to overlook the development of the disparate impact framework, the difficulties employers face in identifying legitimate business objectives, and the potential pitfalls of validation efforts. In light of these factors, I argue that the disparate impact framework holds significant potential to address AI-driven discrimination.

1. The Logic of Disparate Impact

Thirty years before codification by Congress, the Supreme Court articulated the disparate impact analysis in the second-ever Supreme Court case interpreting Title VII, Griggs v. Duke Power Co. The disparate impact theory developed in the context of seniority systems and written tests imposed just after Title VII came into effect, when “the issue was the perpetuation of past intentional but lawful discrimination that would contravene the purposes of the legislation.”

Although novel in some ways, AI tools comport neatly with the logic of holding employers liable for adopting seniority systems and written examinations. Structurally, the inputs and outcomes of AI tools are measurable and reproducible in the same way as written tests, which makes them more amenable than most hiring practices to the quantitative validation practices sought by the Uniform Guidelines on Employee Selection Procedures (“the Uniform Guidelines”). Perhaps more importantly, just as

---

97 See, e.g., Prince & Schwarck, supra note 27, at 1305 (“[F]irms using AIs that proxy discriminate will typically have little problem showing that this practice is consistent with business necessity . . . . because, by definition, proxy discrimination helps the AI predict a legitimate objective: the target variable it is programmed to optimize . . . .”); Kim, supra note 15, at 866 (“[T]o ask whether the model is ‘job related’ in the sense of ‘statistically correlated’ is tautological.”); Ajunwa, The Paradox of Automation as Anti-Bias Intervention, supra note 19, at 1726–27 (proposing a “discrimination per se” standard to replace the business necessity defense and make disparate impact claims against AI-driven discrimination more viable).
98 See 401 U.S. 424, 430–31 (1971) (holding that “practices, procedures, or tests neutral on their face, and even neutral in terms of intent” can still violate Title VII if they are “discriminatory in operation” and “cannot be shown to be related to job performance”).
99 Michael Selmi, Was the Disparate Impact Theory a Mistake?, 53 UCLA L. REV. 701, 708, 715. Although the disparate impact framework is framed broadly and has had some success in, for instance, physical test and English-only cases, Selmi’s empirical review of cases between 1983 and 2002 shows declining success rates on disparate impact claims. Id. at 738–43.
100 The Uniform Guidelines, codified at 29 C.F.R. § 1607, were jointly promulgated by the EEOC, Department of Justice, Department of Labor, and Civil Service Commission (succeeded by
written tests and seniority systems had the effect of locking in past patterns of intentional discrimination, an AI tool “holds a mirror to the past. It distills patterns in past data and interprets them as projections of the future. Algorithmic prediction produces a precise reflection of digital data.”

Put another way, algorithmic prediction “has the potential to perpetuate or amplify social inequality, all while maintaining the veneer of high-tech objectivity.” This potential mirrors the role of written tests and seniority systems in perpetuating “past intentional but lawful discrimination that would contravene the purposes” of Title VII.

*United Papermakers* illustrates the reasoning that could be brought to bear in AI-driven discrimination cases. There, the employer argued that it had stopped discriminating against Black employees once Title VII became effective and “[t]he fact that the system continues to prefer whites over previously hired Negroes in filling certain vacancies does not in itself show racial discrimination.” The court rejected this argument, concluding that when “an employer or union has discriminated in the past and when its present policies renew or exaggerate discriminatory effects, those policies must yield, unless there is an overriding legitimate, non-racial business purpose.”

Consider again Amazon’s experiment with AI tools. Had the company put the tool into practice, why should the logic of *United Papermakers*—and the cases and regulations that followed—not apply? The AI tool disfavored women because Amazon’s workforce and training data were historically dominated by men. Thus, to the extent that a tool replicates existing discrimination and is...

the Office of Personnel Management and Merit System Protection Board). 29 C.F.R. § 1607.1. Although not entitled to *Chevron* deference, the Uniform Guidelines have often proven persuasive in federal courts, particularly in analyzing disparate impact claims. See Albermarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975) (“[T]his Court has heretofore noted, [the Guidelines] do constitute ‘[t]he administrative interpretation of the Act by the enforcing agency,’ and consequently they are ‘entitled to great deference.’” (quoting *Griggs*, 401 U.S. at 433-34)); see also *Gulino* v. N.Y. State Educ. Dept, 460 F.3d 361, 383 (2d Cir. 2006) (“Although it has been stressed repeatedly that courts are by no means bound by the EEOC Guidelines, courts rely on them because following the Guidelines promotes consistency in the enforcement of anti-discrimination law.”). But see EEOC v. Arabian Am. Oil Co., 499 U.S. 244, 257 (1991) (noting that Congress “did not confer upon the EEOC authority to promulgate rules or regulations” and, accordingly, the deference afforded to agencies guidance will be limited to “all those factors which give it power to persuade”).

101 Mayson, supra note 16, at 2224.
102 Id. at 2221.
103 Selmi, supra note 99, at 715.
104 Local 189, United Papermakers v. United States, 416 F.2d 980 (5th Cir. 1969), abrogated on other grounds by *Bernard* v. *Gulf Oil Corp.*, 841 F.2d 547, 555 (5th Cir. 1988).
105 Id. at 986.
106 Id. at 989 (emphasis added).
107 See *Dastin*, supra note 1 (concluding that Amazon’s tool demonstrated gender bias “because Amazon’s computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period” and, with men making up sixty percent of Amazon’s workforce, “[m]ost came from men”); see also 29 C.F.R. § 1607.14(B)(4) (2021) (“[T]he sample
amenable to validation, a disparate impact analysis is appropriate. The following sections illustrate how such a claim could succeed by considering the elements of a disparate impact claim in light of the features of an AI hiring tool discussed in Part I.

2. Prima Facie Case of Disparate Impact

Plaintiffs have a straightforward initial burden in disparate impact cases: they “must begin by identifying the specific employment practice that is challenged.”

Next, plaintiffs must “offer statistical evidence of a kind and degree sufficient to show that the practice in question has caused the exclusion of applicants for jobs or promotions because of their membership in a protected group.” To demonstrate a statistical disparate impact, the Uniform Guidelines instruct that a selection rate for one sex or racial group “which is less than four-fifths . . . of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact.” Although the Uniform Guidelines presuppose that actual applicants make up the comparator group, courts have frequently held that the applicant pool should not be the comparator group if “otherwise qualified people might be discouraged from applying because of a self-recognized inability to meet the very standards challenged as being discriminatory.”

subjects [training data] should . . . be representative of the candidates normally available in the relevant labor market for the job or group of jobs in question, and should insofar as feasible include the races, sexes, and ethnic groups normally available in the relevant job market.”.

108 Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 994 (1988). The Court’s requirement in Watson that plaintiffs identify a discrete and specific employment practice was abrogated by the Civil Rights Act of 1991, which now requires plaintiffs to “demonstrate that each particular challenged employment practice causes a disparate impact, except that if the complaining party can demonstrate to the court that the elements of a respondent’s decisionmaking process are not capable of separation for analysis, the decisionmaking process may be analyzed as one employment practice.” 42 U.S.C. § 2000e-2(k)(B)(i) (emphasis added).

109 Sherer, King & Mrkonich, supra note 18, at 495.

110 Watson, 487 U.S. at 994.

111 29 C.F.R. § 1607.4(D) (2021). This is merely the default rule, and courts frequently embrace other tests of statistical significance. See, e.g., Isabel v. City of Memphis, 404 F.3d 404, 412 (6th Cir. 2005) (holding T- and Z-tests of statistical significance are more appropriate to find an adverse impact in some circumstances); Anderson v. Zubieta, 180 F.3d 329, 339 (D.C. Cir. 1999) (holding that a disparity of just under two standard deviations under a “two-tailed test of statistical significance” was sufficient to establish a prima facie case).

112 Dothard v. Rawlinson, 433 U.S. 321, 330 (1977); see also EEOC v. Joint Apprenticeship Comm. of Joint Indus. Bd. of Elec. Indus., 186 F.3d 110, 120 (2d Cir. 1999) (“[A] statistical showing of disparate impact need not, and in [some] instances . . . should not, be premised on an analysis of
Thus, for a job applicant facing discrimination, the use of AI hiring tools is a readily identifiable employment practice that applicants can challenge under Title VII. And, when it comes to establishing a disparate impact, AI’s large scale may make it easier than normal to establish statistical significance.

3. The Business Necessity Defense

Once a plaintiff has shown a disproportionate adverse impact, a defendant can still avoid liability by showing that “any given requirement [has] a manifest relationship to the employment in question.” This business necessity defense hinges on three questions. First, is the target criteria (a desired skill or trait) identified by the employer actually important for the successful performance of the job? Second, is the selection procedure an adequate proxy for that target criteria? Finally, if the first two are true, can the plaintiff show that a less discriminatory alternative exists?

The characteristics of actual applicants.

See Scherer, King & Mrkonich, supra note 18, at 495 (“When disparate impacts arise, algorithmic selection procedures give potential plaintiffs an obvious target.”). The plaintiff’s task of identifying the use of AI hiring tools would certainly be aided by greater transparency in employer’s use of such tools. See supra notes 67–69 and accompanying text.

See Scherer, King & Mrkonich, supra note 18, at 484 (arguing that “[i]f an employer uses an algorithmic tool to assess hundreds or thousands of candidates, rejected candidates who sue may find that the bar for making out a prima facie case . . . is remarkably low” because the threshold for statistical significance is lower for very large data sets). Pauline Kim has also argued persuasively that targets of AI-driven discrimination should have the option of identifying disparate impact not only relative to the applicant or labor pool, but also to the training data. Kim, supra note 15, at 919.


It is helpful analytically to separate out these questions, but courts and litigants are often not so tidy in their analysis and frequently discuss target criteria and proxy validation in an integrated manner.

29 C.F.R. § 1607.14(A) (2021); see, e.g., Ernst v. City of Chicago, 837 F.3d 788, 799 (7th Cir. 2016) (noting that selection procedures must “significantly correlate[] with important job-performance elements”).

29 C.F.R. § 1607.14(B) (2021); see, e.g., Dothard v. Rawlinson, 433 U.S. 321, 331-32 (1977) (holding that height and weight are not sufficiently related to strength to support a business necessity defense).

42 U.S.C. § 2000e-2(k)(1)(A); see, e.g., Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) (“If an employer does then meet the burden of proving that its tests are ‘job related,’ it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer’s legitimate interest in ‘efficient and trustworthy workmanship.’ Such a showing would be evidence that the employer was using its tests merely as a ‘pretext’ for discrimination.” (citations omitted)).
a. Target Criteria

The Uniform Guidelines emphasize that screening criteria should “represent important or critical work behavior(s) or work outcomes,” which should be identified through a formal job analysis or in another manner that allows the employer to “show the importance of the criteria to the particular employment context.” Relying on this guidance, appellate courts have affirmed essential knowledge, physical skills, and concrete work outcomes as appropriate target criteria.

On the other hand, courts typically disfavor target criteria tied to inconsistent or highly subjective performance appraisals. Multiple circuits have also emphasized that target criteria cannot be validated by correlation with a second test or measure that is itself not properly validated.

Accordingly, an employer that has quantified and measured important work-related knowledge, skills, or outcomes could use such target criteria to train an AI tool and assert a business necessity defense. However, if the employer relies on unstructured supervisorial rankings or other subjective criteria to train the AI tool, it will have considerable difficulty prevailing on a business necessity defense.

Recall the hypothetical airline using an AI hiring tool to select flight attendants. If the airline uses a simple one-to-five customer satisfaction rating as the target variable for which the AI tool is trained to optimize results, the tool is likely to disproportionately favor women for such roles if, as Pan

---

120 29 C.F.R. § 1607.14(B)(3), (C)(4) (2021). Certain criteria, such as “production rate, error rate, tardiness, absenteeism, and length of service” do not require a formal job analysis. 29 C.F.R. § 1607.14(B)(3) (2021).

121 The First Circuit has approved of a written test for police sergeant candidates that was developed by “survey[ing] police officers in thirty-four jurisdictions nationwide” to identify areas of knowledge “critical to the performance of a police sergeant’s responsibilities.” Lopez v. City of Lawrence, 823 F.3d 102, 109 (1st Cir. 2016).

122 Employers may screen applicants based on actions that must be carried out on the job (or proxies for those actions), but not for generalized fitness standards unmoored from actual responsibilities. Ernst, 837 F.3d at 800.

123 For instance, the Third Circuit has accepted paratransit passenger safety as a critical work outcome when presented with evidence that people with disabilities are particularly vulnerable to violent crimes and that “employees of transportation providers commit a disproportionate share of those crimes.” El v. Se. Pa. Transp. Auth., 479 F.3d 232, 247 (3d Cir. 2007).

124 See Albemarle Paper Co. v. Moody, 422 U.S. 405, 432–33 (1975) (rejecting a screening test that was tied to “subjective supervisory rankings” where “supervisors were asked to rank employees by a ‘standard’ that was extremely vague and fatally open to divergent interpretations”).

125 See, e.g., Ernst, 837 F.3d at 802 (such circular validation attempts are not permissible because they constitute little more than “a statistical form of self-affirmation”); Guardians Ass’n of the N.Y.C. Police Dep’t, Inc. v. Civil Serv. Comm’n of N.Y. (“Guardians II”), 633 F.2d 232, 244 (2d Cir. 1980), aff’d on other grounds, 465 U.S. 582 (1983) (rejecting claims of validation based on “a high correlation between the results of two separate testing practices, neither of which by itself has been validated”).

126 See infra subsection I.B.1.
American airlines concluded in the 1970s, customers prefer to be waited on by women. The airline’s assertion of a business necessity defense should fail in this situation because it is impossible to determine “precisely what criteria of job performance the [customers] were considering, whether each of the [customers] was considering the same criteria or whether . . . the criteria actually considered were sufficiently related to the Company’s legitimate interest.”127

The result would be even starker if the airline relied on an out-of-the-box AI tool provided by a vendor that failed to incorporate the kind of individualized job analysis envisioned by the Uniform Guidelines, but instead optimized for generic customer service attributes.128 More generally, hiring tools that evaluate candidate interviews or social media accounts for broad characteristics like “dependability,” “emotional intelligence,” or “respectfulness,” are vulnerable to a disparate impact claim because such screening mechanisms “attempt[] to measure general qualities such as intelligence or commonsense, which are no more relevant to the job in question than to any other job,”129 and are thus “often biased in favor of a person’s familiarity with the dominant culture”130 and likely to “perpetuate the effects of prior discrimination.”131

Thus, an employer’s use of vague or unreliable target criteria in the development of an AI tool will likely prove fatal to its business necessity defense.

b. Screening Proxy

Even when target criteria are appropriate, the screening mechanism often falls short if it cannot be validated as related to such criteria.132 The Uniform

127 Cf. Albemarle Paper Co., 422 U.S. at 433 (holding that such a criterion, in the context of supervisor evaluations, is an impermissibly vague and unreliable basis on which to defend a charge of disparate impact).
128 See 29 C.F.R. § 1607.14(B)(2) (discussing the importance of determining employment criteria based on specific “review of job information”); see also Walston v. Cnty. Sch. Bd. of Nansemond Cnty., 492 F.2d 919, 926 (4th Cir. 1974) (“A job analysis for one . . . position (and the appropriate test for it) would not necessarily be suitable for another.”); M.O.C.H.A. Soc’y, Inc. v. City of Buffalo, 689 F.3d 263, 278 (2d Cir. 2012) (“[A] proper job analysis . . . include[s] a thorough survey of the relative importance of the various skills involved in the job in question and the degree of competency required in regard to each skill . . . .” (quoting Guardians II, 633 F.2d at 242) (internal quotation marks omitted)).
129 Guardians Ass’n of the N.Y.C. Police Dep’t, Inc. v. Civil Serv. Comm’n of N.Y. (“Guardians I”), 630 F.2d 79, 93 (2d Cir. 1980).
130 Id.
131 Id.; see also Gulino v. N.Y. State Educ. Dep’t, 460 F.3d 361, 385 (2d Cir. 2006) (confirming that “Guardians is still the law” in light of the 1991 Civil Rights Act and recent Supreme Court precedents).
132 The classic case of a bad screening proxy is Dothard v. Rawlinson. There, the Court rejects height and weight requirements imposed on correctional officers as a proxy for strength and holds that if strength is a business necessity, then the employer must “adopt[] and validat[e] a test for applicants that measures strength directly.” 433 U.S. 321, 332 (1977).
Guidelines envision three approaches validating the relationship between selection procedures and the job-related target criteria, but I will focus on criterion validation, as it is the one most likely to be used to defend AI hiring tools.

Criterion validation seeks to confirm that a selection procedure is “predictive of or significantly correlated with” important job performance characteristics. Criterion validation is used when employers cannot test candidates on the actual content of the job (such as by administering a typing test to prospective typists) and so must administer an assessment that correlates with the job. For instance, in hiring paramedics, candidates generally cannot actually rescue people, but employers could administer a skills test and correlate candidates’ scores with the scores of incumbent paramedics with high job performance ratings. Key to the success of a criterion-related validation is that “a study’s sample population should, as far as possible, be representative of the candidates normally available in the relevant labor market for the job” in terms of sex, race, and ethnicity.

Employers using AI hiring tools trained on data demographically distinct from the larger labor market are unlikely to prevail on a business necessity defense. This was fundamentally the flaw in Amazon’s experiment: by training its AI tool on its current employees, Amazon’s tool was designed to validate against a pool of almost entirely men. The more homogenous the training pool, the more obvious the discrimination, and the more difficult a business necessity defense, will be. This is because the number of false correlations, such as between men’s athletics and successful engineers, “may increase if the training examples tend to come from individuals from the same demographic group or groups, and who therefore share non-job-related attributes in the data.” Similarly:

133 See 29 C.F.R. § 1607.5(B) (2021) (identifying content, construct, and criterion-related validity tests). The guidelines have not been updated in more than forty years, even though the test design literature and practice appear to have evolved considerably, largely abandoning this criterion-content-construct trichotomy. See Scherer, King & Mrkonich, supra note 18, at 483 (“The Guidelines refer to construct validity as ‘a relatively new and developing procedure in the employment field,’ . . . today, far from an undeveloped and novel theory, construct validity is generally recognized as the overarching validity concept.” (footnotes omitted)).

134 Content validation, which seeks to confirm that the “content of the selection procedure is representative” of important job responsibilities, 29 C.F.R. § 1607.5(B) (2021), could be used to validate AI hiring tools that administer a task that mirrors a task that workers perform, but would not be appropriate for tools that seek to discern nonobvious correlations between applicant traits and successful workers.

135 29 C.F.R. § 1607.5(B) (2021) (emphasis added).

136 See Ernst v. City of Chicago, 837 F.3d 788, 796 (7th Cir. 2016) (discussing a criterion-related validity study of the city’s screening procedures for paramedic applicants).

137 Id. at 800 (quoting 29 C.F.R. § 1607.14(B)(4)).

138 See Dastin, supra note 1 (concluding the gender gap resulted in “Amazon’s system [teaching] itself that male candidates were preferable”).

139 Scherer, King & Mrkonich, supra note 18, at 488.
[If] musical tastes differ by race, and the best incumbent job performers for a particular position are predominantly from a given race, then a high correlation between musical taste and job performance may exist—but only due to demographics, and not because musical taste is an accurate and generalizable predictor of job performance. The less representative the training data are of the population at large, the higher the risk that a deep learning model will identify and create a model that relies upon such demographics-dependent correlations.\footnote{Id.}

Thus, although many scholars correctly state that AI tools are built to identify correlations, which would seem to make them uniquely effective at securing criterion validation, such correlations are only defensible if they stem from a comparator pool that reflects the labor market, which may not always be the case.

C. The Limits of Affirmative Action

In addition to creating a cause of action for employment discrimination, Title VII also creates a barrier to private sector affirmative action. Calls from scholars for “algorithmic affirmative action”\footnote{See Anupam Chander, The Racist Algorithm?, 115 MICH. L. REV. 1023, 1025 (2017) (book review) (“[I]f we believe that the real-world facts, on which algorithms are trained and operate, are deeply suffused with invidious discrimination, then our prescription to the problem of racist or sexist algorithms is algorithmic affirmative action.” (footnotes omitted)).} face technical and legal challenges in the employment context.\footnote{Regarding the technical difficulties, see Mayson, supra note 16, 2271-72, which considers proposals for algorithmic affirmative action in the criminal justice context and concludes that, even if such efforts are possible, they are “likely to have a substantial cost in accuracy, which means more incorrect predictions . . . [that] may fall disproportionately on black communities.”} When Congress codified disparate impact liability into Title VII, it also declared that employers may not “adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests” on the basis of protected characteristics.\footnote{42 U.S.C. § 2000e-2(l).}

Unlawful adjustments include after-the-fact alterations of criteria weighting\footnote{See Ricci v. DeStefano, 557 U.S. 557, 589-90 (2009) (rejecting an employer’s proposed re-weighting of criteria in the absence of evidence that the original weighting was arbitrary because it would likely violate Section 2000e-2(l)).} and “race norming,” meaning to statistically equate scores based on race.\footnote{See Chicago Firefighters Loc. 2 v. City of Chicago, 249 F.3d 649, 656 (7th Cir. 2001) (Posner, J.) (“[I]f banding were adopted in order to make lower black scores seem higher, it would indeed be a form of race norming, and therefore forbidden.” (emphasis added)).} Title VII permits such after-the-fact adjustments only if an employer has “a strong basis in evidence to believe it will be subject to disparate-impact liability if it fails to take the race-conscious, discriminatory
action.” 146 Without such a basis, employers likely face disparate treatment liability for “algorithmic affirmative action” or other efforts to make after-the-fact adjustments to the data. 147

Still, this restriction has a timing element. The Supreme Court has emphasized that designing an employment process to ensure all applicants have a fair shot is permissible, but “once that process has been established and employers have made clear their selection criteria, they may not then invalidate the test results.” 148 Thus, Title VII likely prohibits race-conscious efforts to equalize employment outcomes, but an employer can take proactive steps in the design phase of a selection procedure to avoid a disparate impact. 149

IV. ROUTES TO ROBUST ENFORCEMENT

On one hand, I have concluded that AI hiring vendors create the potential for discriminatory harm on a much larger scale than the scholarly literature has previously acknowledged. On the other, I am more optimistic about the ability of disparate impact doctrines to meaningfully address such harms when they’re identified. With that in mind, I propose two paths toward more robust enforcement of antidiscrimination laws: first, today, the EEOC could bring targeted commissioner charges with a focus on enjoining the use of unvalidated vendor tools. Second, in the long run, Congress should amend Title VII to better enable the EEOC and private plaintiffs to challenge the harms caused by AI hiring vendors.

A. Enjoin the Use of Unvalidated Vendor Tools

Although Title VII doesn’t allow for suits against third-party AI vendors, the EEOC can indirectly challenge such vendors by bringing enforcement actions against their customers. The EEOC has the power to both litigate charges filed by victims of discrimination and to proactively issue a “commissioner charge” under Title VII alleging that an employer has engaged in an unlawful employment practice. 150 Once a charge is issued, the agency

---

146 Ricci, 557 U.S. at 585.
147 See Mayson, supra note 16, at 2268 (discussing strategies for using statistical methods to “cancel out” racial distortions).
148 Ricci, 557 U.S. at 584.
149 Such proactive steps still have limits. See Scherer, King & Mrkonich, supra note 18, at 477 (“Using quotas or granting bonus points on the basis of protected class status, for instance, surely would not survive a disparate treatment challenge, even if an employer adds those features as part of initial test design.”).
150 42 U.S.C. § 2000e-5(b); 29 C.F.R. §§ 1601.6(a), 1601.11(a) (2021). The same power exists under other antidiscrimination laws. See 29 U.S.C. § 626 (empowering the EEOC to “make investigations” under the ADEA); 29 U.S.C. § 211(a) (authorizing the EEOC to “investigate and gather data” under the Equal Pay Act).
has the authority to investigate as it would any other charge of employment discrimination, including issuing and enforcing subpoenas for the testimony and production of relevant evidence.\textsuperscript{151} Although the EEOC cannot bring commissioner charges against third-party vendors,\textsuperscript{152} they can be used as an indirect enforcement mechanism against AI vendors by enjoining those vendors’ customers.\textsuperscript{153} Thus, the EEOC has the power to (1) bring a commissioner charge against an employer using AI hiring tools with a likely discriminatory effect, (2) issue the employer and other relevant parties subpoenas to further the investigation of discrimination, and (3) secure a
injunction prohibiting the use of that tool going forward if a court concludes that the employers’ use of AI hiring tools created an unlawful disparate impact based on a protected trait.

Indeed, EEOC Commissioner Sonderling has recently indicated that he thinks the agency should do exactly this.\textsuperscript{154} Although this could turn into a high-stakes game of whack-a-mole, with the agency bringing charges against each customer one at a time, it is more likely that facing an EEOC investigation, subpoena, and injunction will encourage customers to abandon shoddy AI tools. This will, in turn, encourage vendors to carefully validate AI tools before putting them on the market. Despite this strategy’s roundabout nature, it is immediately available to challenge AI-driven discrimination under current law.

B. Amend Title VII to Address Unique Problems of AI Vendors

The long-term strategy is for Congress to amend employment discrimination statutes to empower the EEOC to directly charge AI vendors who sell products that they know or should know contribute to employment discrimination. Such a reform should have three elements: (1) establish direct

\textsuperscript{151} 29 C.F.R. § 1601.16(a), (c) (2021). The EEOC’s subpoena power is not unlimited. If seeking to enforce such a subpoena in court, the agency has the burden of showing that it has a “realistic expectation rather than an idle’ hope that the information requested will advance its investigation . . . .” EEOC v. Konica Minolta Bus. Sols. U.S.A., Inc., 639 F.3d 366, 369 (7th Cir. 2011) (quoting EEOC v. United Air Lines, Inc., 287 F.3d 643, 652-53 (7th Cir. 2002)); EEOC v. Tricore Reference Lab’ys, 849 F.3d 929, 937 (10th Cir. 2017). The EEOC must further show a link between the EEOC’s investigatory power, the charges of discrimination, and the relevance of the subpoenaed information. EEOC v. Shell Oil Co., 466 U.S. 54, 65 (1984). Nonetheless, this is a far more potent prelitigation information-gathering tool than is available to private plaintiffs.

\textsuperscript{152} Commissioner charges are limited to the same parties as are covered by the rest of Title VII, including “employer[s], employment agenc[ies], [and] labor organization[s].” 42 U.S.C. 2000e-5(b).

\textsuperscript{153} See 42 U.S.C. § 2000e-5(g) (authorizing a court, if it finds an employer has discriminated, to “enjoin the [employer] from engaging in such unlawful employment practice, and order . . . any other equitable relief as the court deems appropriate”).

liability for third parties; (2) establish record retention, auditing, and transparency requirements; and (3) empower the EEOC to promulgate regulations with the force of law to address AI-driven discrimination.

First, under Section 707 of Title VII, the EEOC already has the power to seek injunctions against any persons "engaged in a pattern or practice of resistance" to Title VII rights if that "pattern or practice is of such a nature and is intended to deny the full exercise of the rights." This "any persons" language of Section 707 is broader than the rest of Title VII and is currently the only authority to bring suit against third parties other than employers, employment agencies, and labor unions.

Although the "any persons" language is broad, courts interpret the "pattern or practice" language narrowly to apply only to parties engaged in systemic disparate treatment, where "racial discrimination [is] the company's standard operating procedure." Congress should expand the EEOC's authority to bring Section 707 charges not just against those who are intentionally denying the full exercise of Title VII rights, but also against those (a) who sell a product or service to employers; (b) which the seller knows or should know is likely to cause an adverse disparate impact on the basis of race, color, religion, sex, national origin, disability, or age; (c) if the product or service does cause such a disparate impact; and (d) the seller is engaged in the business of selling such a product or service. This framework borrows from the Restatement (Second) of Torts, which is animated by the related principle that people and companies who sell a product should bear the costs of the foreseeable harms that the product causes. This amendment to Title VII would allow robust enforcement against the most dangerous third parties—AI vendors—without requiring a more comprehensive overhaul of the employment discrimination enforcement scheme.

---


156 Section 707 has allowed the EEOC (and the Department of Justice before it) to seek injunctions against the Ku Klux Klan and other non-employers intentionally interfering with Title VII rights. See United States v. Original Knights of the Ku Klux Klan, 250 F. Supp. 330, 349 (E.D. La. 1965) (holding that the Civil Rights Act of 1957 allows the Attorney General protect civil rights by seeking injunctive relief against "any person, public or private"); United States v. Board of Educ., 911 F.2d 882, 892 (3d Cir. 1990) ("One need not be the employer of the employees whose Title VII rights are endangered in order to be liable under this section . . . .").


158 RESTATEMENT (SECOND) OF TORTS § 402A(1) (AM. L. INST. 1965); see also RESTATEMENT (THIRD) OF TORTS: PRODUCT LIABILITY § 2 cmt. a (AM. L. INST. 1997) (asserting that imposing liability on manufacturers for harm caused by their products "encourages greater investment in product safety," and because "manufacturers invest in quality control at consciously chosen levels, their knowledge that a predictable number of flawed products will enter the marketplace entails an element of deliberation about the amount of injury that will result from their activity").
Second, Congress should nationalize the transparency legislation percolating at the state and local levels. Illinois, 159 Maryland, 160 and New York City 161 have passed bills that focus on disclosure and auditing of AI tools used in employment. While lacking the kind of disparate impact enforcement mechanisms needed to address the potential systematizing of harms by AI vendors, the bills create much-needed transparency about the use of AI tools in hiring and alert workers to situations in which pursuing an EEOC claim might be beneficial. This should be the national minimum. Congress should require all employers to both inform candidates when employers subject them to automated decisionmaking and to inform the EEOC of AI hiring tools used via disclosures for large employers. 162

Third, Congress should authorize the EEOC to promulgate regulations governing the use of AI hiring tools. Although the agency could immediately work to update the Uniform Guidelines—and that wouldn’t be a wasted effort—courts have repeatedly emphasized that such guidelines do not have the force of law. 163 For the EEOC to robustly enforce antidiscrimination laws and prevent AI vendors from systematizing discrimination, it must have the authority to regulate. Indeed, this is an area well-suited to agency regulation, as it is narrow, evolving much faster than Congress can legislate, and would benefit from the expertise of agency staff.

CONCLUSION

Scholars, watchdogs, and even some policymakers have been sounding the alarm for several years about the potential dangers of discriminatory

---

159 The Illinois Artificial Intelligence Video Interview Act (AIVIA) requires companies to notify applicants when AI tools will be used for candidate screening, explain how the technology works, obtain the consent of candidates, and keep such videos private. 820 ILL. COMP. STAT. 42 (2019). AIVIA was amended in 2021 by H.B. 53, 102nd Gen. Assemb., Reg. Sess. (Ill. 2021), to require employers who use AI video screening tools to collect and report certain demographic information.


161 The New York City Council passed a local law that requires employers to inform candidates about the use of “automated employment decision tool[s]” and to conduct a “bias audit” on such tools prior to their use. N.Y.C., N.Y. ADMIN. CODE, tit. 20, ch. 5, subch. 25 (2022).

162 Notably, the EEOC is likely authorized under current law to require, at minimum, that the agency be informed of AI hiring tools in use, though such a regulation would be a departure from historical practice. See 42 U.S.C. § 2000e-8(c) (requiring the EEOC to establish “reasonable, necessary, or appropriate” regulations requiring covered entities to make, keep, preserve, and report information necessary to enforce Title VII); see also 29 C.F.R. § 1602.7 (2021) (requiring every employer with 100 or more employees to complete the EEO-1 report of demographic data).

163 See, e.g., EEOC v. Arabian Am. Oil Co., 499 U.S. 244, 257 (1991) (Congress “did not confer upon the EEOC authority to promulgate rules or regulations” so the Uniform Guidelines only receive deference based on “all those factors which give it power to persuade”).
algorithms. These leaders have rightly drawn attention to the pernicious aspects of AI hiring tools, their distinctive ability to avoid redress under current antidiscrimination regimes, and their tendency to dignify cultural stereotypes with a thin veneer of science. Nonetheless, the conversation so far has overlooked the unique role of AI vendors in scaling and systematizing employment discrimination across entire industries or regions. As more employers adopt such tools and as these vendors continue to grow in size, the threat to equal employment opportunity becomes more urgent. However, significant opportunities exist to begin holding these vendors accountable immediately and to enact focused reforms in employment discrimination law to strategically confront the escalating threat of AI-driven discrimination.