October 2026

The Future of Disparate Impact

Report

Robert VerBruggen
Senior Fellow
Manhattan Institute

Executive Summary

Since the 1971 Supreme Court case *Griggs v. Duke Power*, American employers have faced liability for "disparate impact." If some demographic groups perform better than others on a test or other selection criteria, a company can be forced to defend the test in lengthy court proceedings, even if there is no evidence of any intent to discriminate. This report details two major developments related to disparate-impact liability and proposes a path forward.

First, the Trump administration's Office of Legal Counsel (OLC) has issued an opinion sharply limiting the executive branch's use of the doctrine. To bring the doctrine in line with the Constitution's guarantee of equal protection, the opinion reframes disparate impact as a way to ferret out covert intentional discrimination (in effect, cases where an employer is using a test with no connection to the job as a way of screening out undesired groups). The OLC opinion does not bind private plaintiffs and can be rescinded under a future administration, but it represents a compelling vision of reform that the Supreme Court or Congress could make the law of the land.

Second, the growing use of algorithms and artificial intelligence (AI) in hiring decisions poses challenges both to the disparate-impact doctrine itself and to reform efforts. Algorithms can have their own biases that do not fit neatly into the concept of "intentional discrimination" and are sometimes hard to detect. Race-based decision-making, even if implemented by algorithms, is clearly illegal under current law. But it is not clear how courts will apply existing laws and doctrines to the new situation, especially if the disparate-impact doctrine is scaled back or eliminated.

About Us

The Manhattan Institute is a community of scholars, journalists, activists, and civic leaders committed to advancing economic opportunity, individual liberty, and the rule of law in America and its great cities.

This report proposes that disparate-impact reforms should be accompanied by an explicit, targeted way to address discrimination that algorithms implement without being instructed to, in order to ensure that this discrimination is addressed and to prevent overreach that stifles promising new technologies. Specifically, someone challenging decisions made by an algorithm should be required to show that: (a) the algorithm relies on factors that have no reasonable relationship to the job; or (b) it warps specific job-related factors into proxies for protected characteristics.

To provide transparency and allow for implementation of this standard, algorithms used to make employment decisions should have to be auditable and explainable. They should provide human-understandable assessments of specific job-related characteristics, even if their inner workings are extremely complicated. Employers should also receive some protection from liability in exchange for auditing their systems ahead of time.

Introduction

Since the 1971 Supreme Court case *Griggs v. Duke Power*, employers have faced liability for "disparate impact." If members of some groups pass an employment test (or meet other selection criteria) at higher rates than others, a company can be forced to defend the test in lengthy court proceedings, even if there is no evidence of any intent to discriminate. Since demographic disparities in both academic and physical performance are very common, disparate-impact liability is an ever-present threat. Paradoxically, the doctrine pressures employers to design hiring policies around racial outcomes, which is both undesirable as policy and in tension with the Constitution's Equal Protection Clause.

This report details new developments in the realm of disparate impact and proposes a path forward. Like the *Griggs* case, it focuses on employment, but disparate impact has been applied in other areas as well—such as housing—and the broader points made here are applicable to those contexts, too.

The first new challenge to disparate impact is an opinion from the Trump administration's Office of Legal Counsel (OLC), which limits the executive branch's use of the doctrine, though it does not bind private plaintiffs and can be changed by a future administration. To address long-standing constitutional objections to the doctrine, the opinion reframes disparate impact as a way to ferret out intentional discrimination—essentially, cases where an employer is using criteria unrelated to the job to screen out applicants from unwanted groups.[1]

The second challenge—to disparate-impact doctrine and to reform efforts—is that hiring processes have become increasingly driven by AI and algorithms. The Left worries that algorithms can perpetuate biases that exist in their training data, in ways that are hard to detect or prove. The Right is concerned that new technology could be used to quietly implement DEI preferences—just when those practices have finally been losing in court. Both sides have a point.

Under the plain language of the 1964 Civil Rights Act, companies may not refuse to hire applicants (or make certain other key employment decisions) "because of" protected characteristics. This clearly prohibits the use of algorithms that make decisions based on such characteristics. It is an open question, however, as to exactly how courts will apply existing approaches to antidiscrimination law—which tend to divide violations into

"intentional" discrimination and disparate impact—to the new situation, especially if disparate-impact doctrine is curtailed or eliminated. In the meantime, some states and localities have imposed aggressive regulations to proactively identify disparate impacts in algorithmic decisions, highlighting the need for an approach that is both nationwide and narrow.

I propose that disparate-impact reforms should be accompanied by an explicit, targeted way to address discrimination that algorithms implement without being instructed to. Someone challenging decisions made by an algorithm should be required to show that: (a) the algorithm relies on factors that have no reasonable relationship to the job; or (b) it warps specific job-related factors into proxies for protected characteristics.

Crucially, (b) does not merely require a correlation between the job qualification and the protected characteristic. Plaintiffs must identify one or more specific qualifications that are functioning as proxies for a protected characteristic and explain how the algorithm departs from a meritocratic assessment of those qualifications. A textbook example might be an algorithm that, in assessing applicants' education, assigns oddly low scores to degrees from historically black colleges, relative to other schools of objectively similar academic quality and selectivity. To defend itself, a business should be able to give a reasonable explanation for the algorithm making the challenged distinctions—for example, that the colleges in question have especially poor programs in areas of particular importance to the job.

Transparency regulations should ensure that algorithms are explainable and auditable to enable this process. Employers should be able to explain to a court (and job applicants should be told during the hiring process itself) what traits were measured, why they are relevant to the job, what information about the candidates was used to evaluate them, and what scores were given—even if the calculations "under the hood" are complicated. Employers should also receive some protection from liability in exchange for auditing their systems ahead of time.

This framework clarifies that algorithms are not allowed to do programmatically what humans have long been prohibited to do themselves, and it explicitly spells out the conditions under which a violation of the rule can be inferred. It zeroes in on the most problematic behaviors documented in real-world algorithms, without allowing endless second-guessing of the factors that a system does and does not consider, mere technical imperfections in how traits are measured, and the weights assigned to different criteria. Nor does it encourage plaintiffs to introduce far-flung legal theories or propose their own novel tests for algorithmic bias (some of which can be mutually exclusive, meaning that to pass one alleged bias test, an algorithm must fail another).[2]

The only questions are whether the factors considered are reasonably job-related and whether the plaintiffs can produce scenarios where equal qualifications—two equivalent roles at previous employers, two hobbies, etc.—are evaluated differently based on the demographic attributes that they signal. The main demand on employers is to use systems that they can explain to job applicants and courts—which, admittedly, will trade some amount of efficiency and accuracy for transparency and ease of enforcement.

The coming months and years will profoundly shape the long-term future of disparate-impact liability. The Trump OLC opinion may be a temporary measure undone by a future administration and soon forgotten. Or it could spur a full reimagining of the doctrine. This report hopes to contribute to the latter project.

Why Disparate Impact Needs Reform

In *Griggs*, the Court ruled that an employment test (or other selection method) could constitute illegal discrimination if members of different demographic groups pass the test at different rates, even if the employer and test designer have no discriminatory intent. Ultimately, courts settled on a three-step framework to decide when employers would be liable.

First step: those challenging a selection method must show that pass rates differ significantly. (A rule of thumb is that a gap is substantial enough if one group's selection rate falls below four-fifths of another's, though statistical-significance tests are also common.)[3] Second step: the business can defend the test on the grounds that it is "job-related and consistent with business necessity," which generally requires detailed expert testimony about the exam's technical aspects. Third step: even if the business wins at step 2, the challenger can still win by showing that an alternative selection method that the business has declined to adopt would serve the business's purpose with less disparate impact.[4]

Griggs was dubious as an interpretation of the Civil Rights Act of 1964—which explicitly protected employment tests so long as they were not designed, intended, or used to discriminate.[5] Amendments to the act in 1991 acknowledged the doctrine,[6] but disparate impact remains extremely controversial, both as a policy and constitutional matter.

As I explained in a previous Manhattan Institute report,[7] a positive gloss on the rule is that—at least in theory—it merely requires businesses to use lower-disparity hiring methods when such methods are available and at least equally effective. The reality of lawsuits to enforce the doctrine, however, is a different story.

The first step of the test is an ever-present threat to employers, because (as Gail Heriot has famously argued) virtually any selection method will have a disparate impact on some demographic group or another. For example, sex differences in physical abilities and racial gaps in academic performance are present among school-age children, long before they are measured on employment tests. Sex and race gaps on physical and written tests are thus to be expected. But such disparities are adequate to get a lawsuit underway.[8]

The second step—which allows an employer to show that a selection method is "consistent with business necessity"—involves a subjective question that typically leads to a "battle of the experts" in court. The issue is not simply whether a test measures obviously job-related skills and has been developed and administered in good faith; instead, the matter hinges on such minutiae as the justifications for cutoff scores and the strength of any statistical correlations found when the test was validated. Determining the appropriateness of a proposed alternative is similarly intricate.[9]

Ultimately, even when businesses design reasonable hiring practices to find the most qualified applicants, they still risk lawsuits, which can cost money and damage reputations even when unsuccessful. The doctrine also creates pressure to avoid liability at step 1 by keeping hiring racially balanced—so long as one can do so proactively and quietly, as businesses are not permitted to throw out the results of a test that has already been administered simply because some groups did better than others, and they are not allowed to use quotas or set different passing thresholds based on protected characteristics.

In addition, the doctrine raises constitutional issues. The doctrine may be in tension with the guarantee of equal protection because disparate-impact protections may apply, or not, to a given individual based on his race. Further, as the late Justice Antonin Scalia put it in his concurrence in the 2009 case *Ricci v. DeStefano*, "Title VII's disparate-impact provisions place a racial thumb on the scales, often requiring employers to evaluate the racial outcomes of their policies, and to make decisions based on (because of) those racial outcomes."[10]

More recently, the Pacific Legal Foundation's Alison Somin has offered two other constitutional arguments. First, the wide discretion given to executive agencies to pursue disparate-impact cases—given that, again, virtually any hiring method is susceptible to some sort of challenge—could be an unconstitutional delegation of Congress's power to make law. Second, at least in some applications, the doctrine could violate the "major questions" doctrine. This rule holds that, when executive agencies claim an ability to make especially consequential decisions, they must be able to point to crystal-clear statutory language authorizing them to do so.[11]

Challenges from the Trump Administration

Disparate-impact investigations and lawsuits can run through several channels, often starting with a complaint to the Equal Employment Opportunity Commission (EEOC). When it receives a complaint, EEOC works with the complaining party and employer to resolve the issue. If that fails, EEOC can bring suit against a private employer itself, refer cases involving state and local governments to the Department of Justice (DOJ), which can also initiate investigations on its own,[12] or return the matter to the complainant, who may then sue privately if he chooses.[13]

The executive branch, which includes both DOJ and EEOC, plays a large role in enforcing disparate-impact liability, but the Trump administration has minimized the doctrine's use. A recent opinion from DOJ's OLC laid out very narrow circumstances in which agencies may pursue disparate-impact cases.[14] While this document can be rescinded under a future president and does not stop private plaintiffs from bringing suit, it represents a generally attractive vision of how disparate impact might ultimately work in a reformed legal regime.

Echoing the logic of Scalia's *Ricci* concurrence, the opinion finds EEOC's prior guidelines unconstitutional "because they contemplate liability based on disparate effects alone, without regard to an employer's likely intent, and pressure employers to engage in race-based decisionmaking." To bring disparate impact in line with the Constitution, it reframes the doctrine as a way to identify intentional discrimination, rather than a way to hold businesses liable regardless of intent.

While the opinion is a departure from both *Griggs* and the 1991 statute, the executive branch has a duty not to pursue cases that rely on unconstitutional doctrines. Moreover, suspicions of intentional discrimination have long undergirded the most defensible uses of the disparate-impact doctrine. As OLC noted,[15] even in *Griggs* itself, the defendant,

Duke Power, had discriminated by race until the law forced it to stop in 1965, and it enacted new testing policies the very day the law went into effect—although the Court also noted the company's "special efforts to help the undereducated employees" since.[16]

How would disparate-impact doctrine look if it were redesigned to address intentional discrimination? The biggest changes are to steps 2 and 3, after a disparate impact has been shown. Under OLC's framework, employers at step 2 no longer need to jump through elaborate technical hoops to prove "business necessity"; instead, they merely need to show that "the challenged practice rationally serves a valid business purpose." Even if the business fails to do so, the challenger needs to propose a superior selection method at step 3. By giving employers an extra chance to avoid step 3, the OLC framework inverts how the step has traditionally worked, which was to give an extra bite at the apple to *plaintiffs* who lost at step 2, not defendants. (The Supreme Court case *Inclusive Communities*, regarding disparate impact in housing, however, notably appears to have misstated the traditional rule.)[17]

"Properly understood," the OLC opinion contends, "disparate-impact liability proscribes only those practices that reflect a significant likelihood of intentional discrimination." Notably, though, the framework does not require direct proof of discriminatory intent—leaving it somewhat unclear as to what might happen if a hiring algorithm fails the three-step test but in a way that is difficult, if not impossible by definition, to interpret as "intent." (Similar questions bedevil the long-standing *McDonnell Douglas* framework, another test used to infer intentional discrimination in the absence of direct proof.)[18]

We'll return to algorithms shortly, but setting that wrinkle aside, OLC's approach is attractive, in terms of addressing potential constitutional issues and in terms of creating a fair policy that encourages colorblind hiring. The framework ends the obvious injustice of holding a company liable for discrimination when it did nothing to actually discriminate. It also would prevent plaintiffs from bringing a claim based on a raw disparity and then getting to nitpick the test to death in court. Under the OLC framework, employers simply have to test abilities that are reasonably connected to the job. Meanwhile, the framework continues to recognize that it *is* discrimination for an employer to use job-irrelevant criteria that exclude undesired groups.

Because the OLC opinion binds only the executive branch and can be undone by a future administration, further action would be needed to make it the law of the land. That could happen in two ways: one is for Congress to write it into the law, though this would require significant bipartisan support to overcome a filibuster in the Senate; the other, far more likely if still far from certain, is for the Supreme Court to adopt OLC's approach. In an appropriate case, the Court could, as Scalia suggested doing two decades ago, overrule the disparate-impact doctrine, except in a narrow set of cases where disparate impact likely reflects genuine discrimination.

Both approaches would provide the opportunity for critics of disparate impact to refine OLC's approach—and to adapt it to rapid changes in the way American employers tend to hire and manage employees.

The AI Conundrum

Algorithms and AI pose new problems for antidiscrimination law, including disparate-impact doctrine itself and efforts to reform it.

AI increasingly permeates the business world, including HR departments. In surveys, the vast majority of companies now report using some form of AI in their hiring process.[19] AI tools are used to do more than scrutinize basic qualifications or quantify information provided on applications. Instead, in the words of one academic paper, "new forms of assessment include online games, video-based interviews, and big data pulled from many sources, including test responses, test-taking behavior, applications, resumes, and social media."[20]

In many ways, this is promising for those who want hiring to be as fair and meritocratic as possible. Human decision-makers are prone to all manner of mistakes, biases, and inconsistencies, which automated systems can avoid by processing all applications in exactly the same way. Computers, moreover, can analyze far more information than any HR manager could keep in his head at once. Modern algorithms and AIs can analyze past data about hiring decisions and employee performance, digest copious information about new job applicants, and rank candidates or recommend the best ones for a given position.

If we don't want algorithms to base decisions directly on characteristics like race and sex, we can simply not provide them that information. However, critics argue that, even with such safeguards in place, computer algorithms can still be unfairly biased. This allegation is not as frivolous as it may seem at first blush. Algorithms can exhibit biases both against and in favor of traditionally underrepresented groups.

About a decade ago, Amazon discovered a flaw in a hiring algorithm that it was working on. By design, the program did not consider sex directly. Yet it systematically down-ranked candidates who had graduated from women's colleges or mentioned female-specific activities on their résumés.[21]

The full details of the experiment, which was quickly scrapped, are not publicly known.[22] But this type of issue can easily occur when math meets real-world, human data. For example, imagine a model trained on data from past hires, with an emphasis on how those individuals later fared in performance reviews. If female hires tended to receive low ratings from their managers, the model could learn to penalize female applicants. Even if it is not told applicants' sex, the model could lean on proxies—variables that predict performance reviews by measuring sex indirectly, rather than by (or in addition to) measuring attributes directly relevant to job performance.

It is worth thinking about why sex (or other demographic traits) might predict job performance in an algorithm's training data. If human managers give women low ratings out of sexism, an algorithm trained on those data will perpetuate that discrimination in future hiring decisions. Even if women really do tend to perform worse on average after being hired, the law still does not permit an employer to refuse to hire someone *because* she is a woman, a practice that economists call "statistical" discrimination.[23] At most, employers can figure out the on-the-job reasons for the disparity, such as upper-body strength, and incorporate those legitimate qualifications into the hiring process. However, as we have seen, even this might risk a messy disparate-impact lawsuit under

the problematic legal status quo. It is thus crucial to distinguish between (a) genuine job qualifications on which demographic groups happen to differ; and (b) variables that are, at least partly, measuring demographic characteristics, a difficult question to which we shall soon return.

A decade after Amazon's experiment, modern AI systems—including large language models (LLMs) like ChatGPT, which debuted in 2022—behave in far more complicated ways than older generations of technology. They have been "trained" on enormous libraries of text and other data and "aligned" with more deliberate training to minimize unethical behavior, including to restrain output that the programmers deem, for example, offensive or racially insensitive. The training data reveal how words and other "tokens" such as punctuation relate to each other; given a user query, AIs can use these relationships to predict a series of tokens that works as a response. Modern AIs are sometimes described as "black boxes," meaning that the connections they make can be buried under mountains of complex calculations and the models' behavior can be mysterious even to their programmers.

Experiments that are run on modern AI, mirroring experiments run with real employers to detect discrimination, often reveal bias. In a Manhattan Institute report, David Rozado detailed a series of randomized experiments in which he asked major AI models to decide between two fictional people in various contexts (hiring, eviction, etc.).[24] The overall results showed a modest bias in *favor* of women and a slight bias against white applicants as well, especially when sex and race were available to the model directly instead of merely implied by names. A different study, by Raluca Alexandra Fulgu and Valerio Capraro and published in *Computers in Human Behavior Reports*, confirms an antimale bias in several AI models—including a greater willingness to abuse a man than a woman in order to prevent a nuclear holocaust—and speculates that "[i]nclusivity efforts may have generated unintended forms of discrimination."[25]

Still other work has shown a remarkable ability of AI to identify anonymous writers—not just what race or sex they are, but their individual identities—by matching their prose styles and other clues to writing that has appeared online elsewhere, such as social media[26] (which employers now routinely inspect when evaluating candidates).[27] And speaking of prose styles, experiments published in *Nature* in 2024 found that major AIs were biased against speakers who use an African-American dialect, despite the models' training not to exhibit bias against black people directly. Specifically, the AIs were less likely to match a fictional person to an occupation requiring higher education—and more likely to think that they were guilty of a crime—if the person's statements were presented in black vernacular English instead of standard American English.[28] Some firms are now using AI to analyze video interviews, which would easily provide information about a candidate's race. AIs' ability to identify and act upon protected characteristics that they have not been explicitly provided is thus quite extensive.

Those on the Left fret about algorithms reproducing historical discrimination in ways that are hard to detect. Conservatives, worried about woke tech companies and DEI, are also unlikely to be reassured by the idea of AI handling job applications. Each group has a point, even as we remember that the main alternative—human hiring decisions—is hardly guaranteed to be better.

Applying Disparate Impact to Algorithms

How might traditional concepts of disparate impact apply to algorithmic bias? Several areas are especially worth exploring here: the legal literature touching on the subject; a potentially groundbreaking lawsuit testing the issue in court; state and local statutes that impose new rules; and proposals offered by the Biden administration.

The Legal Literature

Legal scholars have been exploring how AI poses fresh difficulties for traditional disparate-impact doctrine.

A 2016 law-review article by Solon Barocas and Andrew D. Selbst noted that, because algorithmic bias "is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court."[29] They further noted that, under the disparate-impact doctrine, "a practice can be justified as a business necessity when its outcomes are predictive of future employment outcomes"—which is precisely the type of statistical relationship that data-mining algorithms look for, and sometimes find, in variables that serve as proxies for race or sex rather than representing real qualifications.

Matthew U. Scherer and coauthors have stressed that EEOC's guidelines for employment tests "reflect half-century-old conceptions both of the nature and format of employment tests and of what makes a test valid."[30] That framework may be nearly impossible to apply to modern algorithms that analyze reams of data through complicated formulas that humans may not be able to interpret. They also note that even tiny gaps may be statistically significant in a large enough dataset, potentially exposing employers to liability. Rather than simply accepting algorithms' linking of data to job performance as a defense, courts may demand elaborate justifications for every element of algorithmic decisions and require justifications for ever-smaller disparities.

Another article, from Nicole Capp, points out that employers might face liability for using tools that even they do not understand, and that plaintiffs might struggle to recognize and prove the connection between the use of a black-box system and the ultimate hiring decision, particularly if they are not notified that AI tools contributed to their rejection.[31] Pauline T. Kim, meanwhile, suggests that biased algorithms might be better understood not as implicating disparate impact but rather as a violation of the Civil Rights Act's prohibition on "classify[ing]" employees in ways that tend to expose them to unfavorable treatment based on protected characteristics.[32]

Writing with two coauthors in 2022, Keith E. Sonderling (now the acting labor secretary under President Trump) offered another view of the potential of current law.[33] Arguing against burdensome new regulations, these authors contended that existing U.S. antidiscrimination law applies to the decisions of algorithms as much as to the decisions of people and that courts are "well-equipped to deal with the complexities of reconciling new AI technology with long-established legal principles." The authors argued that disparate-treatment doctrines—which are generally used for intentional discrimination—might

be applied to algorithms alongside disparate-impact rules, especially when algorithms learn to penalize protected characteristics in a way analogous to explicit, straightforward discrimination by humans.

Sonderling proposed voluntary compliance programs, in which companies audit their own systems with the help of federal guidance and receive some protection in return. This would improve one of the most frustrating aspects of the current disparate-impact doctrine—namely, that firms cannot know whether a test is legal until after they have been sued and received a ruling.

The proposal closest to this report's own comes from Robert Bartlett and coauthors. They propose the "Input Accountability Test" for algorithms. Mapping the *Griggs* ruling and other case law onto the realities of modern algorithms, Bartlett et al. suggest that, in evaluating claims of algorithmic bias, the law should first ask whether the algorithm is designed to predict a legitimate job qualification, and then, whether the errors in its measurements are correlated with membership in protected classes.[34]

AI Hiring in Court

The interplay of AI and disparate impact is already being litigated, most prominently in *Mobley v. Workday*. The class-action plaintiffs are pursuing a wide range of claims under both federal and California law, contending that Workday's software produces illegal disparate impacts by race, sex, age, and disability.

Historically, employers have faced liability for disparate impacts even when their hiring tools were designed by outside vendors (though vendors have typically been eager to help defend their products in expert testimony). In this case, though, the judge allowed the case to proceed on the theory that companies delegated enough hiring responsibilities to Workday to make it an "agent" of the employers. This drastically raises the stakes for AI developers going forward.[35]

The plaintiffs allege that they received many rejections from companies using Workday technology, sometimes at odd hours or almost immediately after applying, suggesting that the letters were automated. Harking back to Amazon's abandoned experiment, the lawsuit notes that when AIs are trained on past hiring decisions, they may re-create biases that were evident in those decisions by leveraging proxies for protected characteristics—pointing to lead plaintiff Derek Mobley's degree from a historically black college, as well as other variables that may serve as proxies, such as "zip code, educational institution, employment gaps, or language usage." The complaint alleges that Workday's "tools utilize machine learning models whose behavior depends on trained parameters, scoring thresholds, and data interactions, such that their discriminatory effects manifest in outcomes rather than observable rules."

Notably, the plaintiffs tried to bring claims of intentional discrimination as well, but the judge dismissed these, finding that they did not allege "specific facts from which the Court could plausibly conclude that Workday intended its screening tools to be discriminatory."[36] At this writing, the most recent developments in the case include legal wrangling over class certification, as well as a report from a plaintiff-side expert finding demographic disparities in rejections,[37] though the public versions of several key documents are heavily redacted.

The case could prove a key bellwether of how modern hiring algorithms fare against disparate-impact claims and whether software makers will face liability. Court-watchers should be especially interested in the dueling data analyses that the two sides present, especially regarding any variables allegedly used as proxies for protected characteristics, and how this evidence is ultimately evaluated in court.

State and Local Laws

Several state and local governments—mostly blue-leaning areas including New York City, Illinois, California, Maryland, and Colorado—have passed laws regulating "algorithmic discrimination" or requiring AI users to proactively screen their tools for problems. A groundbreaking law in this area, New York City's Local Law 144, provides a good example of left-of-center approaches to the issue.

Local Law 144 applies when an automated hiring tool's "simplified output" (for example, a ranking or score) is the most heavily weighted (or only) factor in an employer's hiring process, or when it is used "to overrule conclusions derived from other factors including human decision-making."[38] These criteria are vague and subjective; some have observed low compliance in the law's early days, coupled with varying perceptions of when the law even applies.[39]

If a tool qualifies, it must undergo an annual third-party "bias audit"—which is a bit of a misnomer, as these reports are not required to measure whether a tool is biased in the sense of treating equally qualified applicants differently. The audits simply document gaps in selection rates or scores across sex and race categories ("impact ratios"). Job applicants must also be informed that automated tools are being used.

The law does not require any particular response to the audits, but the statistics must be posted on companies' websites, publicly advertising which selection methods are likely to fail step 1 of the disparate-impact framework in court. Beyond encouraging disparate-impact suits, this may also prompt employers to avoid the problem by quietly gravitating toward tools that produce racially balanced recommendations (even at a cost in effectiveness), which is functionally similar to a racial quota. Alternatively, employers may eschew hiring algorithms entirely in favor of more subjective processes, which do not require this reporting.

Relative to other efforts in this area, New York's approach is on the aggressive side. Other states have adopted policies with interesting features, such as Illinois's specific prohibition on the use of zip codes to proxy for protected characteristics[40] and California's requirement of risk assessments when automated tools are used to make important decisions.[41] Red Texas, meanwhile, requires proof of discriminatory intent in a claim of algorithmic discrimination; an algorithm that learns to discriminate on its own is not covered.[42]

Also notable is a Colorado statute enacted in 2024 that would have required bias audits, extensive monitoring, and reports to the state attorney general of any "incidents" of "unlawful" disparate impacts.[43] Following strong pushback from the tech industry—including a lawsuit from xAI supported by the Trump administration—these policies have since been scaled back to more general disclosure and documentation rules, such as a requirement that employers tell applicants that automated tools are used in hiring.

The Trump administration has pushed for a single national standard that preempts state and local AI regulations, in order to alleviate pressure on businesses to eliminate disparities, the hodgepodge of different policies now in effect in different states, and the potential for these laws to reach beyond states' own borders.[44]

The Biden Administration's Blueprint

President Biden's administration, by contrast, released several guidance documents and policy proposals that would have applied traditional disparate-impact approaches to algorithmic hiring decisions.[45] *Its Blueprint for an AI Bill of Rights* is a good summary of that administration's overall principles.[46]

The document proposed that algorithms used to make important decisions should be rigorously evaluated to uncover "potential discrimination and effects on equity," including in tests that use a "broad set of measures" to determine whether these systems "produce disparities." Developers were advised to avoid giving demographic information to algorithms directly and to minimize the use of variables strongly correlated with demographics so that they do not serve as proxies. Where assessments reveal disparate impact, users should "document the disparity and a justification for any continued use of the system"; they should take steps to mitigate disparities that break the law or "violate equity goals." On a more general level, affected parties should be notified when automated systems are used, and systems should produce clear, accurate, plain-language explanations for the recommendations they make.[47]

Not all these suggestions are entirely off base. But they clearly reflect a focus on disparity as much as on bias.

Options for a Fair Policy That Promotes Colorblindness and Meritocracy

The disparate-impact doctrine, in starting with raw disparity rather than evidence of real bias, is profoundly unfair to those it accuses of discrimination. Yet we critics of the doctrine sometimes speak too narrowly—or, at least, too narrowly for the AI age—when we stress "intent" as the primary question when it comes to discrimination. For more than half a century, employers have been prohibited from making employment decisions "because of" protected characteristics, which algorithms can now do all by themselves.

Just as it is unclear where the law will head under the status quo, it is hard to predict how algorithmic bias would be treated if disparate-impact doctrine were eliminated or dramatically curtailed while the rest of antidiscrimination law remained intact. Courts might adapt intentional-discrimination doctrines to the situation or treat algorithmic decision-making as a form of "classifying" employees and develop new doctrines around that concept. Some courts might apply something like OLC's framework to algorithmic decisions. Doing so would require sidestepping the framework's stated purpose of identifying likely intent, perhaps by assigning intent for the machine's behavior to the

user or developer or plunging into the philosophical weeds to determine that machines do have some form of intent. Litigants might be eager to introduce other clever legal theories to make it as easy as possible to sue businesses notwithstanding the loss of the traditional disparate-impact framework.

As courts and the current administration continue their work, disparate-impact critics should advance a vision of what the final rules should look like. They should also work, if possible, to pass them through Congress rather than leaving the matter to the plaintiffs' bar and the judges they face.

Those rules should explicitly clarify that algorithms may not do the things that American humans have long been prohibited to do under the Civil Rights Act, while spelling out a narrow, fair test for determining whether this has occurred. Someone challenging decisions made by an algorithm should be able to show that: (a) it relies on factors that have no reasonable relationship to the job; or (b) it warps specific job-related factors into proxies for protected characteristics.

Crucially, (b) does not merely require a correlation between the job qualification and the protected characteristic. Plaintiffs must identify one or more specific qualifications that are functioning as proxies for protected characteristics and explain how the algorithm departs from a meritocratic assessment of those qualifications. A business should be able to defend itself by giving a reasonable explanation for the algorithm making the challenged distinctions.

A textbook example might be an algorithm that assigns oddly low scores to degrees from historically black colleges, relative to other schools of objectively similar academic quality and selectivity. This is a sign that the variable is measuring whether the applicant is black—instead of, or in addition to, measuring college quality. This stands in stark contrast to a "disparate impact" under the status quo, in which it would be problematic for black applicants' colleges to receive lower scores even if this happened because those colleges truly were less selective and rigorous. In response, the business could produce evidence that, say, the down-ranked colleges, despite their overall competitiveness, have substandard programs in areas of special importance to the job.

How will such problems come to light? In the purest black-box scenario—say, an employer simply tells a chatbot to rank résumés and accepts the results—these cases would be rather difficult to pursue. It would be hard even to identify a specific qualification to challenge—though, even in this case, experiments like Rozado's could be used to show which kinds of changes to the program's inputs affected its recommendations.

To enable antidiscrimination enforcement—and to advance the separately desirable goal of giving workers some transparency about the processes to which they are subject—Congress and regulators should consider some transparency ground rules. These rules should make it possible to investigate claims of bias without hobbling new technology or requiring the public release of proprietary information.

As discussed above, many current proposals and regulations seek to document and minimize the disparities that algorithms produce, because raw disparity is a key focal point of the disparate-impact doctrine. However, some ideas in these regulations, particularly those having to do with transparency, can be adapted to more narrowly focus on uncovering bias.

Companies using AI in hiring might reasonably be required to inform applicants about the inputs and outputs involved—i.e., their own personal qualifications and ultimate scores—so that they can review the information for accuracy. While this does present a

compliance burden, it is similar to the way current law treats credit reports and background checks under the Fair Credit Reporting Act,[48] and AI-based systems should be especially efficient at automatically generating these reports and sending them to applicants.

Firms should be required to use hiring systems that are explainable and auditable. Such a system might, for example, score candidates on several specific dimensions and combine the scores into an overall recommendation through an explicit formula. If the system were challenged in court, the key questions would be whether the individual dimensions were reasonably related to the job and whether substantively equal qualifications were being scored differently, depending on the protected characteristics they signal. Importantly, this does not mean that every single calculation "under the hood" of these systems must be simple and easily interpretable. Rather, the output used in the hiring decision must measure specific job-related qualifications in ways auditors can inspect. The idea is to allow the use of the world's most advanced technologies, including LLMs, while ensuring that such technologies measure concrete job capabilities, not demographics.

This, too, creates a burden, but a reasonable one. Explainable algorithms are already considered a best practice when AI is used for important decisions.[49] A similar requirement is also all but mandated in the context of credit decisions, as the Equal Credit Opportunity Act requires that applicants for credit be provided the specific reasons they were denied.[50] Even apart from discrimination concerns, basic transparency in hiring decisions is a desirable outcome.

In addition to being easy to interpret, explicit scores for important types of qualifications have provided key evidence in cases where humans make decisions. In challenging Harvard's admissions practices as biased against Asian applicants, for example, lawyers and experts for Students for Fair Admissions were able to show that Asian applicants received oddly low "personality" scores notwithstanding high scores on more objective criteria, and that this disparity stemmed from scores given by admissions officers, as opposed to volunteer alumni who interviewed some candidates in person.[51] Similar patterns—where the scoring of a qualification appears warped in a way that corresponds to demographics—could be used to document that AI systems are down-ranking applicants of some groups in a targeted fashion.

There should also be a role for Sonderling et al.'s voluntary compliance proposal. Employers should be given the option of auditing their algorithms, using federally approved criteria, for patterns indicative of illegal discrimination before putting them to use. In return, they should receive some protection from liability—such as limiting relief to a change in policy going forward when plaintiffs find further problems in algorithms that had already been audited to regulators' satisfaction. Of course, these audits must be oriented toward finding not raw disparities, but proxy variables that cause the algorithm to discriminate based on projected characteristics.

To be clear, disparate-impact reform can clarify rules and forbid lawsuits that are rooted in little more than simple disparity, but it cannot end messy discrimination lawsuits entirely. For that, we would need simply to legalize discrimination. In my proposed system, there will still be difficult courtroom battles between expert witnesses about what does and does not indicate bias. As they develop new doctrines and enforcement guidelines on a case-by-case basis, courts and executive agencies will have to make judgment calls regarding when qualifications are substantively equal and how different their scores must be to qualify as a problem.

But such battles should focus on whether truly unfair treatment occurred, not on whether a selection method, while indisputably measuring job-related skills, is just technically imperfect enough that it is not "consistent with business necessity"—or on whether some other hypothetical method might be marginally better.

A Word About Data Collection

In addition to its assault on the disparate-impact doctrine, the Trump administration is in the process of paring back requirements that employers classify their employees by race, preserve the data, and file summary reports with EEOC.[52] It is unclear as to how many employers would keep such records even without a requirement (in which case, the records could still be accessed in relevant investigations and lawsuits). Of course, as with the OLC memo, a future presidential administration could undo these changes as easily as the Trump administration enacted them.

Yet Dan Morenoff, executive director of the American Civil Rights Project and an adjunct fellow at the Manhattan Institute, has pointed out that, although the Civil Rights Act empowers EEOC to require recordkeeping and reporting of information that could be relevant to civil-rights violations, the act also straightforwardly bans employers from classifying employees "in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin."[53]

Under a broad—but straightforward—interpretation of this language by the courts, racial classification of employees could shift from merely "not required" to effectively banned. Morenoff argues that an employee should be able to challenge a classification "simply because [it] is the kind of proxy for covered demography that a hypothetical bigot would want in hand in order to deny opportunities or otherwise adversely affect one's status because of covered demography." He points out that the act of classification is banned separately from the act of employment discrimination; the classification itself is the violation, and the classification ban is arguably redundant if paired with a requirement to show that the employer actually collected the information to implement discrimination.

Future developments in this area will profoundly shape disparate-impact doctrine and discrimination enforcement. If racial data are no longer even collected, it will still be possible to argue that hiring criteria are not job-related and to produce direct evidence of discriminatory intent, but it will become far more difficult to show whether algorithms are using variables as racial proxies.

Conclusion

After more than 50 years, the disparate-impact doctrine finally faces the prospect of serious reforms, and the OLC memo presents a way to address its defects without eliminating it entirely. At the same time, the emergence of algorithmic and AI-driven hiring raises fresh questions, both for the doctrine itself and for reform ideas that are centered on an intent to discriminate. The ideal path forward is to pare back disparate impact along the lines that OLC suggests—while developing a new framework to assess the real problem of discrimination by algorithms.

About the Author



Robert VerBruggen is a senior fellow at the Manhattan Institute where he provides policy research, writes for *City Journal*, and contributes to special projects and initiatives in the President's office. Having held roles as Deputy Managing Editor of *National Review*, Managing Editor of *The American Conservative*, Editor at *RealClearPolicy*, and Assistant Book Editor at *The Washington Times*, VerBruggen writes on a wide array of issues including economic policy, public finance, health care, education, family policy, cancel culture, and public safety. VerBruggen was a Phillips Foundation Journalism Fellow in 2009 and a 2005 winner of the Chicago Headline Club Peter Lisagor Award. He holds a B.A. in journalism and political science from Northwestern University

Endnotes

1 T. Elliot Gaiser and Joshua J. Craddock, "Constitutionality of Disparate-Impact Liability Under Title VII," memorandum opinion for the chair of the U.S. Equal Employment Opportunity Commission (EEOC), Office of Legal Counsel (OLC), U.S. Dept. of Justice (DOJ), June 9, 2026.

2 An example from "risk assessments" in criminal justice: if two groups have different recidivism rates, a risk rating can either: (a) mean the same thing across the groups—for example, with a "7" rating, meaning a 70% chance of recidivism regardless of whether the subject is black or white; or (b) have the same "false positive" rate across groups, which might be defined, in this case, as the percentage of non-recidivators who were initially classified in a higher-risk category. The disparity in (b) occurs because, as a group's recidivism rate increases, an accurate algorithm puts more members in the high-risk category—translating to more false positives in the numerator (as, unless the algorithm is perfect, some share of offenders considered high-risk will always avoid recidivism), divided by fewer non-recidivators in the denominator (owing simply to the higher recidivism rate, regardless of how anyone was classified). See Robert VerBruggen, "Does Pre-Crime Have a Race Problem?" *The American Conservative*, Aug. 2, 2016.

3 For a discussion of the four-fifths guidance and courts' sometimes dismissive treatment of it, see Matthew U. Scherer, Allan G. King, and Marko J. Mrkonich, "Applying Old Rules to New Tools: Employment Discrimination Law in the Age of Algorithms," *South Carolina Law Review* 71, no. 2 (Winter 2019): 449–552.

4 The third step was added with *Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975).

5 42 U.S.C. §2000e-2(h).

6 42 U.S.C. §2000e-2(k).

7 Robert VerBruggen, "Fight Bias and Legalize Meritocracy: A Unifying Vision for Antidiscrimination Law," Manhattan Institute, May 8, 2025.

8 Gail L. Heriot, "Title VII Disparate Impact Liability Makes Almost Everything Presumptively Illegal," *New York University Journal of Law & Liberty* 14, no. 1 (2020): 1–170.

9 For EEOC's detailed technical guidelines, to which the courts have often turned when assessing tests, see 29 CFR Part 1607, *Uniform Guidelines on Employee Selection Procedures*.

10 *Ricci v. DeStefano*, 557 U.S. 557 (2009).

11 Alison Somin, "Disparate Impact as a Non-Delegation Violation and Major Question," *Harvard Journal of Law and Public Policy: Per Curiam* (Summer 2024).

12 EEOC and DOJ, Civil Rights Division, "Memorandum of Understanding Between the U.S. Equal Employment Opportunity Commission and the U.S. Department of Justice–Civil Rights Division Regarding Title VII Employment Discrimination Charges Against State and Local Governments," Dec. 21, 2018.

13 EEOC, "What You Can Expect After a Charge Is Filed," accessed Sept. 9, 2026.

14 Gaiser and Craddock, "Constitutionality of Disparate-Impact Liability Under Title VII."

15 Ibid., n. 7.

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

17 For the statute's statement of the rule, note the use of "or" between parts (i) and (ii) of 42 U.S.C. §2000e-2(k)(1)(A). *Inclusive Communities* misstated the rule by claiming that courts must accept a proposed alternative "before rejecting a business justification," instead of before accepting one. *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, 576 U.S. 519 (2015).

18 Emile Loza de Siles, "AI, on the Law of Intentionality: Toward Proof of Intentional Discrimination by Artificial Intelligence," Social Science Research Network, updated Sept. 29, 2020.

19 ManpowerGroup, "90% of Companies Use AI in Hiring. Fewer than 5% Are Seeing It Work," news release, June 23, 2026; Utkarsh Amitabh and Ali Ansari, "Hiring with AI Doesn't Have to Be So Inhumane," World Economic Forum, Mar. 28, 2025.

20 Nancy T. Tippins, Frederick L. Oswald, and S. Morton McPhail, "Scientific, Legal, and Ethical Concerns About AI-Based Personnel Selection Tools: A Call to Action," *Personnel Assessment and Decisions* 7, no. 2 (2021): 1–22.

21 Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," Reuters, Oct. 10, 2018.

22 Ibid. The Reuters report notes that the training data were male-dominated, though it is not clear why this, by itself, would make the algorithm actively discriminate against women in the way it apparently did.

23 Jon Kleinberg et al., "Discrimination in the Age of Algorithms," *Journal of Legal Analysis* 10 (April 2019): 113–74.

24 David Rozado, "Fairness in AI Decisions About People: Evidence from LLM Experiments," Manhattan Institute, Jan. 8, 2026.

25 Raluca Alexandra Fulgu and Valerio Capraro, "Surprising Gender Biases in GPT," *Computers in Human Behavior Reports* 16 (December 2024): 100533.

26 Simon Lermen et al., "Large-Scale Online Deanonymization with LLMs," arXiv:2602.16800v2, Feb. 25, 2026.

27 David Cotriss, "How Social Media Screenings Affect Hiring Decisions," *Business News Daily*, updated Jan. 26, 2026.

28 Valentin Hofmann et al., "AI Generates Covertly Racist Decisions About People Based on Their Dialect," *Nature* 633, no. 8028 (September 2024): 147–54.

[29] Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104 (2016): 671–732.

[30] Scherer, King, and Mrkonich, "Applying Old Rules to New Tools."

[31] Nicole Capp, "Behind the Black Box: Employer Accountability for Algorithmic Hiring Bias," *Business, Entrepreneurship & Tax Law Review* 10, no. 1 (2026): 134–53.

[32] Pauline T. Kim, "Data-Driven Discrimination at Work," *William & Mary Law Review* 58, no. 3 (2016): 857–956.

[33] Keith E. Sonderling, Bradford J. Kelley, and Lance Casimir, "The Promise and the Peril: Artificial Intelligence and Employment Discrimination," *University of Miami Law Review* 77, no. 1 (2022): 1–86.

[34] Robert Bartlett et al., "Algorithmic Discrimination and Input Accountability Under the Civil Rights Acts," *Berkeley Technology Law Journal* 36, no. 2 (2021): 1–55. Also notable, in a similar vein but focused on evaluating overall decision-making processes more than specific criteria, is Jongbin Jung et al., "Measuring Disparate Impact in Human and Machine Decisions," *Proceedings of the National Academy of Sciences* 123, no. 30 (2026): e2509765122.

[35] Annette Tyman, "Mobley v. Workday: Court Holds AI Service Providers Could Be Directly Liable for Employment Discrimination Under 'Agent' Theory," Seyfarth Shaw Legal Update, July 19, 2024.

[36] Order Granting in Part and Denying in Part Motion to Dismiss, *Mobley v. Workday*, 3:23-cv-00770-RFL.

[37] Expert Report of Dr. Charles A. Scherbaum, *Mobley v. Workday*, 3:23-cv-00770-RFL.

[38] NYC Dept. of Consumer and Worker Protection, "Notice of Adoption of Final Rule: Automated Employment Decision Tools," April 2023.

[39] Lara Groves et al., "Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime," arXiv preprint, arXiv:2402.08101, Feb. 12, 2024.

[40] 775 Ill. Comp. Stat. 5/2-102(L)(1).

[41] Joseph J. Lazzarotti, "Is a CCPA Risk Assessment Required When Using AI-Powered Hiring and Screening Tools?" JacksonLewis Workplace Privacy, Data Management & Security Report, May 19, 2026.

[42] Duane Morris LLP, "Texas' AI Law Is Now in Effect—What Employers Need to Know About TRAIGA 2.0," May 27, 2026.

[43] Colorado General Assembly, "Senate Bill 24-205: Concerning Consumer Protections in Interactions with Artificial Intelligence Systems," 2024 regular session, signed May 17, 2024.

[44] White House, "Ensuring a National Policy Framework for Artificial Intelligence," Executive Order 14365, Dec. 11, 2025.

45 U.S. Dept. of Labor, Office of Federal Contract Compliance Programs, "Artificial Intelligence and Equal Employment Opportunity for Federal Contractors," updated Apr. 29, 2024.

46 White House Office of Science and Technology Policy, "Blueprint for an AI Bill of Rights," Oct. 4, 2022.

47 Ibid.

48 Vendors that produce reports about job applicants may already be subject to similar rules under the Fair Credit Reporting Act, an issue raised in a recent lawsuit against Eightfold AI. In reply, Eightfold says that its software merely "analyzes information the [applicants] provided for the employer's internal use" without drawing from any third-party sources of information about the candidates, which does not trigger the scrutiny usually applied to credit reports and background checks. *Kistler et al. v. Eightfold AI Inc.*, "Class Action Complaint," No. C26-00214 (Cal. Super. Ct., Contra Costa County, filed Jan. 20, 2026); and "Defendant Eightfold AI Inc.'s Motion to Dismiss Plaintiffs' Complaint for Failure to State a Claim," No. 4:26-cv-01768-YGR, document 29 (N.D. Cal., filed Apr. 20, 2026).

49 See, e.g., National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework* (AI RMF 1.0), NIST AI 100-1 (Gaithersburg, MD: National Institute of Standards and Technology, January 2023); Center for Industry Self-Regulation, *Principles for Trustworthy AI in Recruiting and Hiring* (McLean, VA: BBB National Programs Charitable Foundation, June 2023); U.S. Dept. of Labor, *Artificial Intelligence and Worker Well-Being: Principles and Best Practices for Developers and Employers*, May 16, 2024.

50 Consumer Financial Protection Bureau, "Adverse Action Notification Requirements in Connection with Credit Decisions Based on Complex Algorithms," circular 2022-03, May 26, 2022.

51 Anemona Hartocollis, "Harvard Rated Asian-American Applicants Lower on Personality Traits, Lawsuit Says," *New York Times*, June 15, 2018.

52 U.S. Office of Personnel Management, "Removal of References to the Uniform Guidelines on Employee Selection Procedures in Federal Personnel Regulations," interim final rule, 91 Fed. Reg. 48,234 (July 31, 2026); EEOC, "Rescission of Uniform Guidelines on Employee Selection Procedures—Recordkeeping," *Unified Agenda of Federal Regulatory and Deregulatory Actions*, RIN 3046-AB45, 2026; "EEOC Proposes Rescission of Annual Race and Sex Reporting Requirements," press release, July 21, 2026.

53 Dan Morenoff, "Title VII's Classification Bar: §2(a)(2), Disparate Impact, and EEO-1s," *Journal of Law & Civil Governance* 3, no. 1 (forthcoming, 2026).