the CENTER for Civil Rights and Technology

The Leadership Conference on Civil and Human Rights

WHEN MACHINES DISCRIMINATE

THE CRITICAL ROLE OF DISPARATE IMPACT IN AI ACCOUNTABILITY

A SNAPSHOT BY CHIRAAG BAINS

Acknowledgements

This briefer is a continuation of the Snapshot series, an initiative of The Leadership Conference's Center for Civil Rights and Technology (Center) meant to educate and empower civil and human rights organizations looking to engage on technology policy issues.

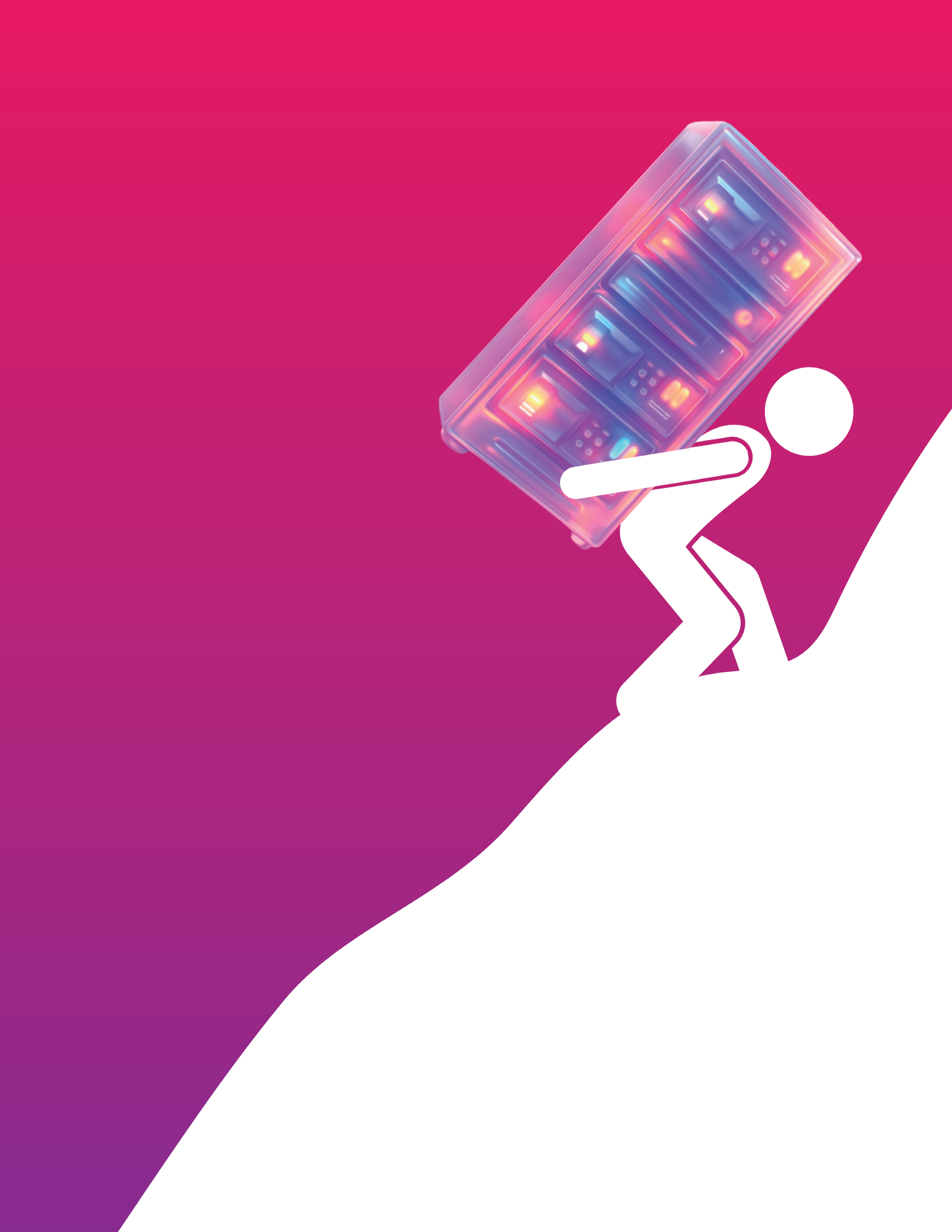
The Center is a joint project of The Leadership Conference on Civil and Human Rights and The Leadership Conference Education Fund. The Center, launched in September 2023, serves as a hub for advocacy, education, and research at the intersection of civil rights and technology policy. Our experts dive into the most pressing policy issues in three key areas: AI and privacy, voting and platform accountability, and broadband access.

Chiraag Bains is a senior fellow at Democracy Fund, nonresident senior fellow at the Brookings Institution, and civil rights lawyer. He served as Deputy Director of the White House Domestic Policy Council and Deputy Assistant to the President for racial justice and equity under President Biden.

He is grateful to Olga Akselrod, Charlotte Burrows, Stephen Hayes, Alondra Nelson, Spencer Overton, and Jenny Yang for their thoughtful comments and conversations. The Center would also like to thank staff for their assistance for this Snapshot, including Alejandra Montoya-Boyer, Mariah Wildgen, and Madeline Chawla. The Leadership Conference's Christian Andrew Madison was the lead designer, and Mark Ricks designed the web version.

The author and publisher are solely responsible for the accuracy of statements and interpretations contained in this publication.

Table of Contents



I. Introduction

Scientific breakthroughs in machine learning, natural language processing, computer vision, and other advanced techniques have taken artificial intelligence out of the realm of science fiction and directly into our lives. AI systems are being used today to make decisions about us. They screen job applications, evaluate creditworthiness for home loans, help decide who can rent an apartment, flag people for suspicion of benefits fraud, target and surveil immigrant communities, make recommendations impacting healthcare, and influence who goes to jail through bail and sentencing recommendations. Investors are pouring billions of dollars into the technology on the promise that it can do even more.[1] As AI's role grows in decisions that affect our rights and opportunities, it is imperative that the technology be fair and nondiscriminatory.

Ideally, the use of AI would produce more fairness by minimizing the influence of human bias and artificial barriers to opportunity. But like other technological advances before it, AI is not neutral.[2] An AI system is influenced by how it was created. It may be trained on biased data. The design team may be from a narrow demographic and reflect that team's lived experiences and biases. Designed to recognize and learn

1 *Who Will Pay for the AI Boom?*, The Economist (July 31, 2025), https://www.economist.com/business/2025/07/31/who-will-pay-for-the-trillion-dollar-ai-boom.

2 *See* Reva Schwartz et al., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, Special Publication, National Institute of Standards and Technology, at ii (2022), https://doi.org/10.6028/NIST.SP.1270 ("Bias is neither new nor unique to AI and it is not possible to achieve zero risk of bias in an AI system.").

from patterns, an AI system can deepen disadvantage by applying stereotyping and replicating the effects of past discrimination at unprecedented scale and speed. Or, an algorithm might overweight unnecessary factors that correlate with identity and thereby introduce new forms of discrimination. The result is that the use of AI can produce biased predictions, bad decisions, and harmful outcomes.

Systems that disadvantage people based on arbitrary and irrelevant factors, rob us of the chance to succeed on our own merit. The discrimination can ripple through communities, denying opportunities based on personal traits such as race, sex, sexual orientation, gender identity, religion, age, disability, or other illegal bases (known as protected characteristics).

Civil rights laws that prohibit ***disparate treatment***—usually involving intentional discrimination—are inadequate to combat such harms. After all, most AI systems are not deliberately designed to discriminate based on protected characteristics. Moreover, datasets and model designs are often proprietary corporate secrets, or so complex that they are effectively black boxes.[3] This can make it exceedingly difficult, if not impossible, to discover when algorithms do treat people differently based on their identity.

Fortunately, there is another legal doctrine that has protected civil rights for over half a century: ***disparate impact liability***. This doctrine tests for invisible barriers[4] to equal opportunity—hidden unfairness

3 Matthew Kosinski, IBM, *What Is Black Box AI?* (Oct. 29, 2024), https://www.ibm.com/think/topics/black-box-ai (defining black box AI as an AI in which "[u]sers can see the system's inputs and outputs, but they can't see what happens within the AI tool to produce those outputs").

4 ReNika Moore, Sept. 16, 2025, in conversation with Author.

based on race, sex, or another irrelevant factor that may be baked into a decisionmaking system. Under disparate impact law, an apparently neutral system that in practice hurts people with a shared protected characteristic is unlawful unless (a) it serves a substantial and important interest, and (b) there is no less discriminatory way to design the system. This doctrine allows victims of algorithmic discrimination to challenge unfair AI systems and seek justice without having to prove the creators' intent to discriminate. It also creates the right incentives for AI developers to test for discriminatory results and adjust their training data and model architecture to make their systems fair.[5] (Disparate impact, as explained below, is entirely different from affirmative action.)

Put more simply, disparate impact liability helps make sure AI-based decisionmaking systems identify qualified people—like strong job applicants and good credit risks—rather than allowing outputs to be skewed unfairly because of race and other traits.

Unfortunately, disparate impact liability is currently under attack. In April 2025, President Donald Trump announced his administration would "eliminate the use of disparate-impact liability in all contexts to the maximum degree possible."[6] He ordered agencies to repeal federal disparate impact regulations—which the Justice Department promptly did, upending over 50 years of law without taking public comment.[7] Trump is also pushing Congress to preempt state laws regulating AI,

5 *See* Chiraag Bains, *The Legal Doctrine that Will Be Key to Preventing AI Discrimination*, Brookings (Sept. 13, 2024), https://www.brookings.edu/articles/the-legal-doctrine-that-will-be-key-to-preventing-ai-discrimination.

6 President Donald J. Trump, Executive Order 14281, Restoring Equality of Opportunity and Meritocracy, 90 Fed. Reg. 17537 (Apr. 23, 2025), https://www.federalregister.gov/documents/2025/04/28/2025-07378/restoring-equality-of-opportunity-and-meritocracy.

7 *Id.;* Final Rule, Rescinding Portions of Department of Justice Title VI Regulations to Conform More Closely With the Statutory Text and to Implement Executive Order 14281, 90 Fed. Reg. (Dec. 10, 2025), https://www.govinfo.gov/content/pkg/FR-2025-12-10/pdf/2025-22448.pdf.

including state disparate impact statutes. Having failed thus far, he issued an executive order in December 2025 directing federal agencies to argue preemption under existing law based on far-fetched legal theories.[8]

This report explains why disparate impact is needed now more than ever and why undermining the doctrine is wrong. In brief, the article: (1) discusses the origin and nature of disparate impact liability; (2) explains how bias materializes in automated systems and how disparate impact can remedy and prevent discriminatory AI; and (3) demonstrates that President Trump's attempt to eliminate disparate impact rests on serious legal errors. Notwithstanding the federal government's abandonment of the doctrine, **disparate impact liability remains the law**. Robust state and private enforcement will help ensure that technological progress does not come at the expense of equality and that everyone can benefit from the promise of AI.

8 President Donald J. Trump, Executive Order 14365, Ensuring a National Policy Framework for Artificial Intelligence 90 Fed. Reg. 58499 (Dec. 11, 2025), https://www.federalregister.gov/public-inspection/2025-23092/artificial-intelligence-efforts-to-ensure-national-policy-framework-eo-14365. See Charlie Bullock, Legal Obstacles to Implementation of the AI Executive Order (Dec. 2025), https://law-ai.org/legal-obstacles-to-implementation-of-the-ai-executive-order.

II.

The Origins of Disparate Impact and How it Works

The landmark statutes passed at the height of the Civil Rights Movement made it unlawful to "discriminate" against people—or for people to be "subjected to discrimination"—based on certain protected characteristics.[9] Those statutes, however, did not expressly define whether "discrimination" meant *only* the explicit differential treatment of people based on a trait like race or sex or *also* the use of facially neutral procedures that in practice unfairly disadvantaged people based on such traits.[10]

9 *See*, e.g., 42 U.S.C. § 2000d (Title VI of the Civil Rights Act of 1964); 42 U.S.C. § 2000e-2(a) (Title VII of the Civil Rights Act of 1964); 29 U.S.C. § 623(a) (Age Discrimination and Employment Act, 1967); 42 U.S.C. § 3604 (Fair Housing Act, 1968); 20 U.S.C. § 1681 (Title IX of the Educational Amendments of 1972); 29 U.S.C. § 794 (Section 504 of the Rehabilitation Act of 1973).

10 The text of several of these statutes strongly suggested they should be read to prohibit unjustified discriminatory effects. For example, Title VII forbade employers to "limit, segregate, or classify" 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" based on race, color, religion, sex, or national origin. 42 U.S.C. § 2000e-2(a)(2). The Age Discrimination and Employment Act contained the same textual prohibition based on age. 29 U.S.C. § 623(a)(2). The Voting Rights Act originally outlawed the application of procedures to "deny or abridge" the right to vote on account of race or color. Pub. L. No. 94-73 (amended in 1982 to prohibit the application of procedures "in a manner which results in a denial or abridgement," 52 U.S.C. § 10301). The Fair Housing Act made it unlawful to "make" housing "unavailable" based on race, color, religion, and national origin, and later sex, disability, and familial status. 42 U.S.C. § 3604(a), (f).

Federal agencies tasked with applying the new statutes interpreted them to cover such discriminatory effects.[11] For example, Title VI of the Civil Rights Act of 1964 empowered agencies to write implementing rules to ensure nondiscrimination in the use of federal funds. That year, the predecessor agency to the Department of Education and the Department of Health and Human Services issued a rule prohibiting recipients of federal funds from "utiliz[ing] criteria or methods of administration which have the effect of subjecting individuals to discrimination."[12]

The Equal Employment Opportunity Commission (EEOC), meanwhile, issued guidance in 1966 and 1970 interpreting Title VII of the Act, which prohibits employment discrimination. Title VII contains an exemption for the use of "any professionally developed ability test" that "is not designed, intended or used to discriminate."[13] Southern companies that previously discriminated openly against Black workers began adopting such tests after Title VII's enactment. The EEOC saw the potential for these tests to produce discriminatory effects—for example, by using written tests requiring significant reading comprehension for jobs that involved little or no reading. The agency therefore took the position that ability tests had to measure qualities relevant for the specific job in question in order to pass muster under Title VII.[14]

11 *See* Olatunde C. Johnson, *The Agency Roots of Disparate Impact*, 49 HARV. C.R.-C.L. L. REV. 125, 127 133-34, 138-39 (2014) (arguing that agency action in the immediate wake of the Civil Rights Act's passage "allows us to understand disparate impact not as a separate offshoot of antidiscrimination law invented by courts, but as a reasonable agency implementation choice given the potentially broad and conflicting meanings of the antidiscrimination directive of civil rights law").

12 29 Fed. Reg. 16298, 16299 (Dec. 4, 1964), codified at 45 C.F.R. § 80.3(b)(2).

13 42 U.S.C. § 2000e-2(h).

14 Alfred W. Blumrosen, *Strangers in Paradise:* Griggs v. Duke Power Co. *and the Concept of Employment Discrimination*, 71 MICH. L. REV. 59, 60-61, 64 (1972); Johnson, *supra* note 11, at 134, 140-41.

Chief Justice Warren E. Burger and the 1970–71 Supreme Court

The Supreme Court validated this view of discrimination in the seminal 1971 case *Griggs v. Duke Power Company*.[15] The Court held that Title VII prohibits "not only overt discrimination but also practices that are fair in form, but discriminatory in operation."[16] In doing so, it recognized a legal framework for what came to be called disparate impact liability.

15 401 U.S. 424 (1971).

16 *Id*. at 431.

In *Griggs*, Black workers at a North Carolina power plant challenged the company's policy of requiring employees to have a high school diploma and pass two written aptitude tests for positions above the lowest job tier. While these requirements appeared neutral, they were not related to skills needed for jobs at the power plant, and they disproportionately excluded Black applicants who, due to historical educational discrimination in the state, graduated from high school and achieved passing test scores at much lower rates than White applicants.[17] (Notably, the company first adopted its high school diploma requirement in 1955, the year after the Supreme Court's landmark desegregation case *Brown v. Board of Education*,[18] and instituted the aptitude tests the day Title VII took effect.[19])

Chief Justice Warren Burger—a conservative jurist appointed by President Richard Nixon—wrote the unanimous decision. He explained that in Title VII, Congress required "the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification."[20] His opinion emphatically rejected the idea that the purity of intent insulated employment practices from Title VII's mandate: "Congress directed the thrust of the Act to the consequences of employment practices, not simply the motivation."[21] Burger wrote that barriers that have a discriminatory effect can be maintained if warranted by "business necessity," meaning in the employment context that they are "related to job performance."[22]

17 *Id*. at 426, 429-30 & n.6.

18 347 U.S. 483 (1954).

19 *Griggs*, 401 U.S. at 427.

20 *Id*.

21 *Id*. at 432 (emphasis in the original); *see also id*. ("good intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups and are unrelated to measuring job capability").

22 *Id.* at 431. The Court also reviewed Title VII's legislative history and validated the EEOC's view that Title VII exempts only employment tests that are job-related. *Id.* at 433-34.

In this case, the evidence showed there was no relationship between either high school graduation or the aptitude tests and job performance. Those requirements therefore violated Title VII.[23]

In subsequent cases the Supreme Court developed a three-part process for evaluating disparate impact claims. First, a plaintiff must make "a prima facie case of discrimination," relying on statistical evidence to show that the challenged employment tests or requirements "select applicants" of a particular race, religion, sex, or national origin in a "pattern significantly different from that of the pool of applicants."[24] This causal showing must compare the demographics of the people selected to the demographics of the people who were qualified and available for the job—*not* to the demographics of the broader population.[25] Moreover, the disparity must be "statistically significant," typically meaning that there is less than a 5% probability that the disparity occurred by chance.[26] Second, the burden then shifts to the employer to prove business necessity by demonstrating that the requirements have a "manifest relationship to the employment in question"—that is, that they are job-related.[27] An employer that has good business reasons for the challenged practices generally can continue to use them. Finally, the plaintiff will still prevail if the evidence shows "that other tests or selection devices, without a similarly undesirable racial [or other prohibited]

23 *Id.* at 431-33.

24 *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975).

25 *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 650-51 (1989) ("a comparison . . . between the racial composition of the qualified persons in the labor market and the persons holding at-issue jobs . . . generally forms the proper basis for the initial inquiry in a disparate impact case"), superseded on other grounds by Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (1991); *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308 (1977) ("a proper comparison was between the racial composition of Hazelwood's teaching staff and the racial composition of the qualified public school teacher population in the relevant labor market").

26 *Jones v. City of Boston*, 752 F.3d 38, 43-44, 46-47 & n.9 (1st Cir. 2014).

27 *Albemarle Paper Co.*, 422 U.S. at 425 (cleaned up).

effect, would also serve the employer's legitimate interest."[28]

Congress amended Title VII in 1991 to codify disparate impact liability and write the Supreme Court's standards into the statute's text.[29] Several other statutes also impose liability for neutral practices with unjustified discriminatory effects. These include the Fair Housing Act, the Equal Credit Opportunity Act, the Age Discrimination in Employment Act, the Safe Streets Act, Title VI, Title IX, the Voting Rights Act, the Americans with Disabilities Act, and Section 504 of the Rehabilitation Act of 1973. The precise standards and rules about which party has the burden vary by statute and jurisdiction, but with some notable exceptions[30] they typically involve the same sort of framework as Title VII:

1
Adverse Impact

Plaintiff shows through a significant statistical disparity that the challenged practice disproportionately harms people with a shared protected trait (race, religion, sex, etc.)

28 *Id.*

29 *See* Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071, § 3(3) (1991) (listing among its purposes "to confirm statutory authority and provide statutory guidance for the adjudication of disparate impact suits under title VII").

30 *See, e.g.*, Chiraag Bains, *What Just Happened: The Trump Administration's Dismissal of Voting Rights Lawsuits*, JUST SECURITY (May 27, 2025) (explaining that results claims under Section 2 of the Voting Rights Act "differ from disparate impact claims in important ways," including that "VRA plaintiffs must adduce evidence in certain enumerated categories concerning past and present discrimination"), https://www.justsecurity.org/113745/wjh-trump-dismissal-voting-rights-lawsuits.

2
Legitimate Interest

Defendant must prove that the challenged practice is necessary to serve a valid interest

3
Less Discriminatory Alternatives

Plaintiff can still prevail by showing that the defendant's valid interest could be served by a different practice with a less discriminatory effect

Although the coverage of these statutes is incomplete, leaving gaps in important sectors of our economy and society,[31] they protect Americans in a host of contexts from private-sector and government decisionmaking systems that appear neutral on their face but discriminate in practice.

31 See Bains, *The Legal Doctrine that Will Be Key to Preventing AI Discrimination, supra* note 5.

III. Disparate Impact as Uniquely Relevant In the Age of AI

A. How AI Systems Perpetuate and Amplify Discrimination

Understanding how AI systems create discriminatory outcomes is crucial for grasping why the disparate impact doctrine is essential to protect people from harm.

With appropriate safeguards, AI may be able to increase reliance on objective factors and reduce the opportunity for human bias to skew decisionmaking. For example, an AI tool that accurately measures skills might be preferable to a human being susceptible to stereotypes or preferences for applicants in their social network. AI tools might even help close opportunity gaps by directing resources to historically underserved neighborhoods and populations. Consider an underwriting algorithm that uses non-standardized, nontraditional information like cash flow data to expand access to credit,[32] or an AI tool that better predicts cardiovascular risk by analyzing diagnostic test results, health records, and activity data from smartwatches.[33] One can see how automation creates the tantalizing possibility of improving fairness.

32 FinRegLab, *The Use of Machine Learning for Credit Underwriting*, 9-10, 12 (2021), https://finreglab.org/wp-content/uploads/2023/12/FinRegLab_2021-09-16_Research-Report_The-Use-of-Machine-Learning-for-Credit-Underwriting_Market-and-Data-Science-Context.pdf.

33 Ariana Mihan et al., *Artificial Intelligence Bias in the Prediction and Detection of Cardiovascular Disease*. NPJ CARDIOVASC HEALTH, 1-2 (2024), https://doi.org/10.1038/s44325-024-00031-9.

AI systems are not inherently neutral, however. They can internalize and cause discrimination in several ways, based on the data they're trained on and how they're designed and deployed:

1. Biased training data

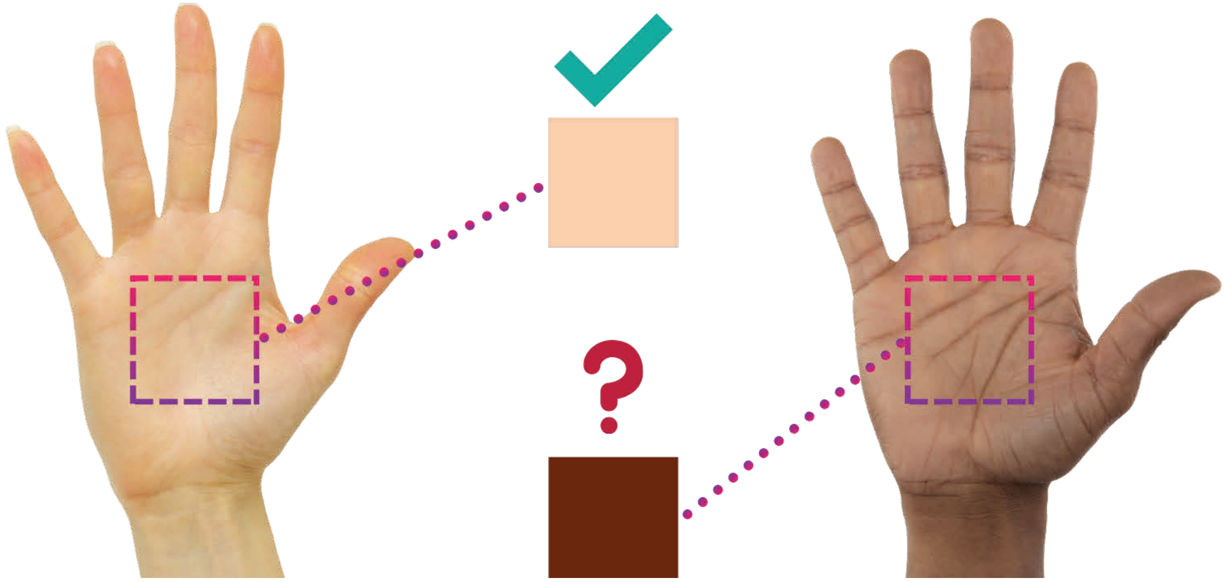
An AI system typically learns by iteratively analyzing and recognizing patterns in huge amounts of "training data"—text, images, audio, video, and other inputs that are "fed" into the system's mathematical algorithm.[34] It then applies those learnings to make predictions, recommendations, or decisions based on likely future outcomes (predictive AI) or to generate material (generative AI) based on new inputs.[35]

The training data may suffer from ***representation bias***, under- or over-representing certain traits in the population the AI will be used to evaluate and thereby skewing its outputs. For example, the underrepresentation of women and people of color in images used to train certain facial recognition tools may explain researchers' findings that the tools worked nearly perfectly in identifying lighter-skinned men but repeatedly failed to recognize darker-skinned women.[36]

34 Cole Stryker, IBM, *What Is Training Data?* (May 2, 2025), https://www.ibm.com/think/topics/training-data.

35 Rina Diane Caballar, IBM, *Generative AI vs. Predictive AI: What's the Difference?* (Aug. 9, 2024), https://www.ibm.com/think/topics/generative-ai-vs-predictive-ai-whats-the-difference.

36 Joy Buolamini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, Proceedings of Machine Learning Res. 81:1–15 (2018), https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; *see also* Patrick Gother et al., National Institute of Science and Technology (NIST), *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects* (2019), https://doi.org/10.6028/NIST.IR.8280 (analyzing 189 facial recognition algorithms and finding elevated false-positive rates for East Asian and Black faces).



Representation bias is the under- or over-representation of certain traits.

In supervised learning—a type of model training in which the AI learns from data manually tagged by human beings—***labeling bias*** can occur when these human annotators systematically label the training data incorrectly, inconsistently, or in ways that reflect social biases. One study of crowdsourced hate speech datasets found that annotators disproportionately labeled Twitter posts in Black vernacular English as offensive or abusive. Automated content moderation models trained on those datasets "acquire and propagate" that bias, flagging Black vernacular posts in test runs at disproportionately high rates.[37]

AI systems can also internalize stereotypes through ***embedding bias***. Word embeddings are numerical representations of words that map how close they tend to appear to other words in an AI system's training data. AI uses word embeddings to understand and process natural language data. Due to pervasive stereotypes, researchers have found in huge corpora of internet text that words like "computer programmer," "pilot," and "champion" appear closer to words like "man," while "homemaker," "maid," and sexual profanities

37 Maarten Sap et al., *The Risk of Racial Bias in Hate Speech Detection*, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1668-70 (2019), https://doi.org/10.18653/v1/p19-1163.

appear closer to words like "woman."[38] One study even found that names associated with being European American are more closely associated with positive words like "loyal" and "honest" and names associated with being African American are more closely associated with negative words like "sickness" and "assault."[39] Studies show that image classifiers can reflect similar biases—for example, identifying a man as a woman because he is standing in a kitchen.[40] Such embedding biases could cause harmful results in AI systems used to screen resumes, recommend people for promotion, assess recidivism risk, respond to chatbot queries, or even rank web search results.

Training data may also be rife with ***historical bias***, causing AI to replicate discrimination in past human decisions. When Amazon trained a recruiting algorithm on ten years of resumes from its predominantly male workforce, the system learned to privilege men's resumes and penalize women's.[41] A health care algorithm affecting millions of people consistently underestimated the medical needs of Black patients because it based its assessments on past health care spending and learned that American health care systems had historically spent "less money caring for Black patients than for White patients."[42] Similarly, lending algorithms trained on past loan

38 *See* Aylin Caliskan et al., *Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics*, Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 156-170 (2022), https://doi.org/10.1145/3514094.3534162; Tessa E.S. Charlesworth et al., *Gender Stereotypes in Natural Language: Word Embeddings Show Robust Consistency Across Child and Adult Language Corpora of More Than 65 Million Words*, PSYCH. SCIENCE, 32(2), 218-240 (2021), https://doi.org/10.1177/0956797620963619; Tolga Bolukbasi et al., *Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings* (2016), https://doi.org/10.48550/arXiv.1607.06520.

39 Aylin Caliskan et al., *Semantics Derived Automatically from Language Corpora Contain Human-like biases*, SCIENCE 356.6334, 183-186 (2017), http://opus.bath.ac.uk/55288; Aylin Caliskan, *Detecting and Mitigating Bias in Natural Language Processing*, Brookings (May 10, 2021), https://www.brookings.edu/articles/detecting-and-mitigating-bias-in-natural-language-processing.

40 Jieyu Zhao et al., *Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-level Constraints,* Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, ACL, pages 2979–2989, 2980 (2017).

41 Jeffrey Dastin, *Insight – Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women,* Reuters (Oct. 11, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG.

42 Ziad Obermeyer et al., *Dissecting Racial Bias in An Algorithm Used to Manage the Health of Populations*, SCIENCE 366(6464): 447-453 (2019), https://doi.org/10.1126/science.aax2342.

decisions could reproduce historical redlining patterns. Criminal justice algorithms trained on arrest data could penalize people and communities subjected to racial profiling.

Worse, AI systems that learn statistical patterns from biased data may also optimize for them. An AI tool might not just reproduce historical discrimination, but could supercharge it by applying it as a rule, at staggering scale and speed.[43]

2. Biased algorithmic design

Choices about model development and deployment create further opportunities for discrimination.

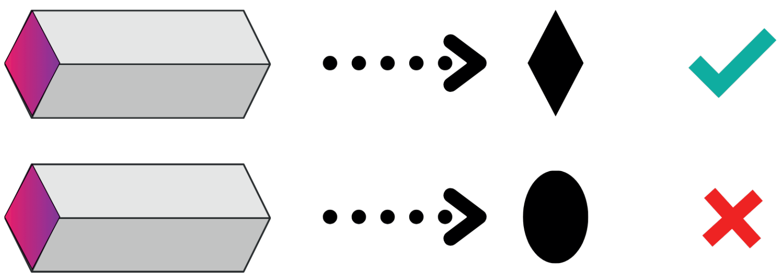
Developers can introduce bias through ***feature selection and weighting*** in the model architecture. For example, if a lending algorithm is designed to weight ZIP code as a predictor of creditworthiness, it could systematically disadvantage applicants of color. Due to persistent residential segregation, ZIP code can function as a proxy for race.[44] Similarly, surnames or language preference

43 Reva Schwartz et al., NIST Special Publication 1270, *Toward a Standard for Identifying and Managing Bias in Artificial Intelligence,* 10, 33 (2022), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf; Klas Leino et al., *Feature-Wise Bias Amplification*, ICLR (2019), https://arxiv.org/abs/1812.08999.

44 Alexandra George, *Thwarting Bias in AI Systems*, Carnegie Mellon University Engineering News (Dec. 2018), https://engineering.cmu.edu/news-events/news/2018/12/11-datta-proxies.html.

can also be proxies for race or national origin.[45] Incorporating these proxies into an algorithm may allow the AI to make decisions based on protected characteristics without doing so expressly. Another example of biased feature selection is a pretrial detention or sentencing algorithm designed to predict recidivism risk based on arrest data. Arrest data is a better measure of police activity than criminal conduct and the likelihood to reoffend.[46]

Another problem is ***deployment bias***, which happens when AI systems are used in contexts different from their training environment. A hiring tool trained on one company's data might not work fairly across different industries or regions. A tool trained on urban area data may have a high failure rate in rural areas.



Deployment bias happens when AI systems are used in contexts different from their training environment.

The creation of ***feedback loops***, in which a model's outputs influence its further training and refinement, can also produce bias. For example, studies have shown that when predictive policing models cause officers to be deployed to an area, data on the arrests they make there are fed back into the model. The predictive model can incorrectly interpret an increase in arrests as an increase in crime, and thus send

45 *See, e.g.,* Nathan Kallus et al., *Assessing Algorithmic Fairness with Unobserved Protected Class Using Data Combination*, MANAGEMENT SCIENCE 68(3):1959-1981 (2021), https://doi.org/10.1287/mnsc.2020.3850.

46 *See* Christine Lindquist, *Racial Equity Considerations When Using Recidivism as a Core Outcome in Reentry Program Evaluations*, RTI International & Center for Court Innovation, at 1 (2021), https://nationalreentryresourcecenter.org/sites/default/files/inline-files/racialEquityRecidivismBrief.pdf; Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2218, 2221 n.4, 2251-52 (2019), https://www.yalelawjournal.org/pdf/Mayson_p5g2tz2m.pdf.

more officers to the same area, "regardless of the true crime rate."[47] This type of self-fulfilling prediction can perpetuate over-policing in low-income neighborhoods and communities of color, create the false impression that their residents are dangerous, extend mass incarceration, and leave crime unaddressed elsewhere.

Many AI systems, particularly those using deep learning and neural networks, are ***black boxes*** in which the internal processes are proprietary and therefore secret. Some are so complex that they are opaque even to their creators. This opacity makes it impossible to detect whether a model is using protected characteristics. The problem is compounded by ***automation bias***, our tendency to defer to automated systems, and the closely related concept of ***technochauvism***, the belief that tech is always superior to other solutions.[48]

B. How Disparate Impact Helps Us Combat Algorithmic Discrimination

Disparate impact doctrine helps surface and root out these sources of bias in ways that disparate treatment doctrine alone cannot.

Machines act based on the programming that human developers give them.[49] Developers, meanwhile, tout their reliance on data and math as a sign that their algorithms are objective, neutral, and trustworthy. When AI nonetheless discriminates in practice, disparate treatment law typically won't help. Disparate impact liability, however, gives people an avenue for redress.

The potential for liability also creates the incentive to *prevent*

47 Danielle Ensign et al., *Runaway Feedback Loops in Predictive Policing*, Proceedings of the 1st FAccT Conference, PMLR 81:160-171 (2018), https://proceedings.mlr.press/v81/ensign18a.html.

48 ROB REICH ET AL., SYSTEM ERROR: WHERE BIG TECH WENT WRONG AND HOW WE CAN REBOOT, 102 (2021); MEREDITH BROUSSARD, ARTIFICIAL UNINTELLIGENCE: HOW COMPUTERS MISUNDERSTAND THE WORLD, 7 (2018); MEREDITH BROUSSARD, MORE THAN A GLITCH, 2 (2023).

49 *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 906-21 (2018).

discrimination before it happens. It encourages system design and testing to minimize bias before deployment, rather than after things have gone wrong. For example, disparate impact law gives a mortgage lender a reason to make sure its AI models do not include factors that overstate risk of default or understate likelihood of repayment for borrowers of a certain race. It requires employers who use AI to ensure their algorithms assess applicants for qualities relevant to the job in question. It pushes AI developers and deployers to explore less discriminatory alternatives—say, avoiding biased training data or changing the algorithm to be fairer while still serving the company's valid purposes.

There are ample ways to make discriminatory models fairer while retaining—or improving—the accuracy of their predictions.[50]

Some training bias may be avoided with forethought, like ensuring that facial recognition software is trained on representative images. Other biases such as historical, labeling, or embedding bias may be harder to remove, but researchers have identified multiple "de-biasing" techniques to reduce their effects.[51]

50 *See* Leadership Conference on Civil and Human Rights, *The Innovation Framework: A Civil Rights Approach to AI* (2025), https://innovationframework.org. *See also* Pauline T. Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action*, 110 CAL. L. REV. 1539, 1544, 1574-86 (2022) (discussing various de-biasing techniques and noting their lawfulness under anti-discrimination law, explaining that "many efforts to eliminate problematic features that cause bias in algorithms are more accurately characterized as non-discriminatory efforts to remove unfairness, rather than 'reverse discrimination'").

51 *See, e.g.*, Yunyi Li et al., *Mitigating Label Bias via Decoupled Confident Learning*, AI & HCI Workshop at 40th ICML (2023), https://doi.org/10.48550/arXiv.2307.08945; Jieyu Zhao et al., *Learning Gender-Neutral Word Embeddings*, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 4847–4853 (2018); Michael Feldman et al., *Certifying and Removing Disparate Impact*, ACM SIGKDD Conf. on Knowledge Discovery & Data Mining (2015), https://doi.org/10.48550/arXiv.1412.3756.

Variables can be added, removed, or weighted differently in an algorithm. Developers can refine a model's "hyperparameters," settings that instruct the algorithm on how to learn from training data.[52] They can also use "adversarial de-biasing," in which they build a second model that tries to find inputs that will cause the primary AI model to exhibit biased behavior, essentially acting as an automated red team identifying weaknesses.[53] The adversarial model and primary model are trained together in a competitive process that maximizes the robustness of the primary model's predictions while minimizing unfairness.[54]

Indeed, because of a concept called "model multiplicity"—the reality that "there are almost always multiple possible models with equivalent accuracy for a given prediction problem"—a model that produces discriminatory effects can frequently be replaced by a less discriminatory version.[55] But model developers may not test for discriminatory effects or consider alternative models. Disparate impact liability gives them a reason to do so—before they cause harm and get sued.[56]

Pushing bias-prevention efforts upstream, into the model development phase, can also be profitable. Models that are accurate and fair for all people help employers identify the most qualified employees, help lenders take prudent credit risks, and help businesses attract more customers. Many AI developers themselves

52 Nicholas Schmidt & Bryce Stephens, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination*, 73 QUARTERLY REPORT 130, 142 https://arxiv.org/pdf/1911.05755.

53 In AI development, a "red team" is a "structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI." NIST, Computer Security Research Center, *Glossary: Artificial Intelligence Red-Teaming*, https://csrc.nist.gov/glossary/term/artificial_intelligence_red_teaming (last visited Jan. 6, 2026).

54 *See id.*; Jenny Yang et al., *An Adversarial Training Framework for Mitigating Algorithmic Biases in Clinical Machine Learning*, NPJ DIGIT. MED. 6, 55 (2023), https://doi.org/10.1038/s41746-023-00805-y.

55 Emily Black et al., *Less Discriminatory Algorithms*, 113 GEO. L.J. 53, 56 (2024).

56 *See generally id.*; *see also* Upturn et al., Letter to Department of Justice regarding Comprehensive Use of Civil Rights Authorities to Prevent and Combat Algorithmic Discrimination, 4 (Feb. 1, 2024), https://www.upturn.org/static/files/2024-02-01%20Letter%20to%20DOJ%20re%20AI%20Executive%20Order%20Civil%20Rights.pdf.

recognize the risk of automated systems producing or replicating bias and have established responsible AI practices aimed at mitigating it. Companies that use biased AI will be outcompeted.[57]

Finally, although most AI systems may not specifically rely on race or sex to make predictions, some do. But injured parties often lack access to the information to find that out. Disparate impact helps here, too. By giving plaintiffs who allege discriminatory effects access to the discovery process in litigation, the doctrine gives them a chance to smoke out evidence that a model in fact classifies people based on protected characteristics. They could then file intentional discrimination claims they otherwise would never have known they had.[58]

In all of these ways, disparate impact is beneficial and indeed indispensable to combating algorithmic discrimination. By eliminating arbitrary barriers, it supports genuine meritocracy.

C. Examples of Recent Legal Actions

Recent legal actions demonstrate both the necessity and efficacy of disparate impact liability in reining in algorithmic discrimination. The examples below largely involve predictive AI systems. Generative AI systems are also beginning to shape processes that affect rights and opportunities.

EMPLOYMENT 💼

Job Application Screening: Derek Mobley, an African American man over 40 with anxiety and depression, applied for over 100 jobs through Workday Inc.'s AI-powered applicant screening and ranking platform. Despite his qualifications—including a finance degree from Morehouse

57 Stephen Hayes, *Why "Disparate Impact" Is Good for Business*, THE ROOFTOP (June 17, 2025), https://www.newamerica.org/future-land-housing/blog/disparate-impact-good-for-business.

58 *See* Tara K. Ramchandani, *Why "Disparate Impact" Matters for Tackling Intentional Housing Discrimination*, THE ROOFTOP (June 17, 2025), https://www.newamerica.org/future-land-housing/blog/disparate-impact-intentional-housing-discrimination ("Disparate impact allows litigants to expose covert intentional discrimination that would otherwise go undetected.").

College, a certification in server management, and work experience—Mobley was rejected in every case. Once, the rejection came within an hour of applying; another time, he was rejected for the job he was currently doing for the same company as a contractor. He sued Workday, alleging that its algorithm discriminated against him intentionally through disparate treatment and unintentionally through disparate impact based on race, age, and disability. In July 2024, a federal judge ruled that AI vendors like Workday can be held liable as agents of employers under federal anti-discrimination laws. Notably, the court dismissed Mobley's disparate treatment claim, finding insufficient indications of discriminatory intent, but allowed his disparate impact claims to proceed.[59] In May 2025, for his age discrimination claim, the judge preliminarily certified a collective action of others who were harmed by Workday's allegedly discriminatory AI.[60] In July 2025, the judge ruled that Workday must provide a list of employers that enabled its AI features to "score, sort, rank, or screen applicants."[61]

Photograph of Mr. Mobley
by Angela Owens/Wall Street Journal

The case could have nationwide consequences. Workday and its competitors provide AI-powered applicant tracking systems to thousands of companies,

59 Order Granting in Part and Denying in Part Motion to Dismiss (Doc. 80), *Mobley v. Workday, Inc.*, 3:23cv770 (N.D. Cal. July 12, 2024), available at https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.80.0.pdf.

60 Order Granting Preliminary Collective Certification (Doc. 128), *Mobley v. Workday, Inc.,* 3:23cv770 (N.D. Cal. May 16, 2025), available at https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.128.0.pdf. A collective action is a species of class action under 29 U.S.C. § 216(b).

61 Order Re HiredScore Dispute (Doc. 158), *Mobley v. Workday, Inc.*, 3:23cv770, at 1 (N.D. Cal. July 29, 2025), available at https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.158.0.pdf. *See also* Caroline Colvin, *Judge orders Workday to supply an exhaustive list of employers that enabled AI hiring tech*, HR DIVE (July 31, 2025), including over 98% of the Fortune 500, who would also have legal exposure for using biased tools.[62] These systems evaluate millions of job seekers each year.

Automated Personality Tests: The American Civil Liberties Union (ACLU) filed a complaint with the Federal Trade Commission (FTC) in 2024 over the consulting company Aon's algorithmic personality assessments, used by major employers to screen millions of applicants. The ACLU alleged that two of Aon's assessments have an adverse impact on autistic people and people with mental health disabilities because they test for "characteristics that are close proxies of their disabilities" and those characteristics are not job-related. Another tool, a gamified cognitive test, allegedly produces disparities based on race and disability. The ACLU argues that Aon has engaged in deceptive marketing, a legal violation within the FTC's jurisdiction, based on the company's claims that its products are "bias free" and "improve diversity." The complaint also argues that the company's "failure to take reasonable measures to assess or address the discriminatory harms" of its automated and/or AI-based assessments is an unfair act, also within the agency's authority to address.[63]

Housing 💼🏚️

Tenant Screening Algorithms: *Louis v. SafeRent Solutions* is a textbook example of how facially neutral algorithms can perpetuate systemic discrimination. Mary Louis, a Black woman with a Section 8 housing voucher, had her rental application denied by SafeRent Solutions' algorithmic screening system despite 16 years of perfect rent payment history. The discrimination arose from a design flaw in SafeRent's algorithm: it failed to properly account for housing vouchers in its scoring system. When voucher holders applied for housing, the algorithm treated them as having less income than they actually had available for rent,

62 Kelsey Purcell, *2024 Applicant Tracking System (ATS) Usage Report: Key Shifts and Strategies for Job Seekers*, JOBSCAN (July 14, 2025), https://www.jobscan.co/blog/fortune-500-use-applicant-tracking-systems.

63 ACLU Complaint to the FTC Regarding Aon Consulting, Inc. (May 30, 2024), https://www.aclu.org/documents/aclu-complaint-to-the-ftc-regarding-aon-consulting-inc;

since it didn't recognize that housing authorities would pay approximately 73% of the rent directly to landlords. This facially neutral error had a severely disparate racial impact because Black and Hispanic individuals make up a disproportionate percentage of voucher recipients.[64] The case revealed how algorithmic discrimination compounds existing inequalities. SafeRent's heavy reliance on credit scores also penalized Black and Hispanic applicants who have lower average scores due to historical discrimination. Property managers relied unquestioningly on the scores without understanding their flaws. The algorithm provided no meaningful avenue for appeal. SafeRent and its clients (including landlords who used the tool) had less discriminatory alternatives—such as adjusting scoring models to properly incorporate voucher income—but failed to adopt them.

As in the *Workday* case, a federal judge rejected SafeRent's defense that it merely provided scores and didn't make final rental decisions.[65] The Department of Justice (DOJ) supported Louis's claims.[66] The case settled for almost $2.3 million, and SafeRent agreed that future scoring systems would be validated by third parties approved by the plaintiffs.[67]

Automated Criminal History Checks: In a case involving SafeRent's predecessor CoreLogic Rental Property Solutions (which CoreLogic later spun off), plaintiffs challenged an algorithmic tenant-screening tool called CrimSAFE. Plaintiffs alleged that CrimSAFE systematically denied housing to individuals—especially African American, Latino, and disabled applicants—based on automated criminal record checks. CrimSAFE's model combined unrelated offenses like traffic offenses and vandalism into single disqualifying categories. It conducted no individualized assessment, provided no underlying documentation, and issued "decline" decisions directly, effectively making decisions for landlords.

64 Complaint (Doc. No. 1), *Louis v. SafeRent Solutions*, No. 1:22cv10800, at 21 (D. Mass. May 25, 2022), https://clearinghouse.net/doc/160025.

65 *Louis v. SafeRent Solutions, LLC*, 685 F. Supp. 3d 19 (D. Mass. July 26, 2023).

66 Statement of Interest of the United States, *Louis v. SafeRent Solutions, LLC*, No. 1:22cv10800 (D. Mass. Jan. 9, 2023), https://www.justice.gov/crt/case-document/file/1562776/dl?inline.

67 Press Release, Cohen Milstein, "Rental Applicants Using Housing Vouchers Settle Ground-Breaking Discrimination Class Action Against SafeRent Solutions" (Apr. 26, 2024), https://www.cohenmilstein.com/rental-applicants-using-housing-vouchers-settle-ground-breaking-discrimination-class-action-against-saferent-solutions.

The plaintiffs alleged unlawful disparate impact under the Fair Housing Act, based on the model's compounding of racial disparities in arrest data and CoreLogic's failure to try to modify its algorithm.[68] The question of whether CoreLogic is subject to the Fair Housing Act is currently pending on appeal.[69]

Chatbots Against Vouchers: In 2023 the nonprofit organization Open Communities and a renter sued Harbor Group, a property rental company with units across the country, and its AI vendor PERQ for using a chatbot that automatically rejected applicants with Housing Choice Vouchers. While the chatbot was specifically configured to reject applicants with government rental assistance, the Fair Housing Act does not prohibit discrimination based on source of income even if intentional. However, the plaintiffs alleged disparate impact based on race—which the Act does cover—because Housing Choice Voucher holders are disproportionately Black. In a settlement, Harbor Group agreed not to turn voucher holders away, and PERQ agreed its AI leasing agents would not violate the Fair Housing Act.[70]

LENDING 🏦

Student Loan Underwriting: In July 2025, Earnest Operations LLC entered a $2.5 million settlement with the Massachusetts Attorney General over allegations that the company's AI was more likely to deny loans to Black and Hispanic borrowers, or to offer them worse terms, compared to White borrowers. The state alleged disparate impact in violation of the Equal Credit Opportunity Act and state law. It also faulted the company for "failing to test its models for disparate impact and

68 *Connecticut Fair Housing Center v. CoreLogic Rental Property Solutions, LLC*, No. 3:18cv705 (D. Conn. Apr. 4, 2018), https://www.cohenmilstein.com/wp-content/uploads/2023/07/CoreLogic-Complaint-04242018_0.pdf.

69 *Connecticut Fair Housing Center v. Corelogic Rental Property Solutions, LLC*, No. 23-1118.

70 Jeff Hirsch, *Fair Housing Group Wins Voucher Discrimination Settlement*, EVANSTON NOW (Feb. 5, 2024), https://evanstonnow.com/fair-housing-group-wins-voucher-discrimination-settlement.

training its models based on arbitrary, discretionary human decisions."[71] The company agreed to conduct such testing going forward as part of the settlement.[72]

Another significant matter involved Upstart Network, a financial technology company that uses AI to decide whether to make student loans and at what interest rate. The Student Borrower Protection Center (SBPC) accused the platform of racial discrimination, alleging that the model would cause a hypothetical Howard University graduate to pay almost $3,500 more for a five-year loan than a similar graduate from New York University.[73] After conversations with SBPC and the NAACP Legal Defense Fund, Upstart made some changes to its underwriting model, including dropping consideration of the average SAT and ACT scores at schools, relying instead on average post-graduation income, and adjusting inputs to ensure students at Minority Serving Institutions (MSIs) and non-MSIs are treated equally. The company also appointed the civil rights firm Relman Colfax PLLC as an independent monitor to analyze its lending model.

The monitor found that although Upstart's model did not use proxies for race, it did approve Black applicants for loans at lower rates. The monitor also identified less discriminatory alternatives—changes to the AI's structure that would reduce racial disparities while still serving the company's purpose of properly assessing creditworthiness.[74]

Upstart implemented the monitor's recommendations about how it conducts disparate impact testing and what level of disparity warrants a

71 Massachusetts Office of the Attorney General, Press Release, "AG Campbell Announces $2.5 Million Settlement With Student Loan Lender For Unlawful Practices Through AI Use, Other Consumer Protection Violations," (July 10, 2025), https://www.mass.gov/news/ag-campbell-announces-25-million-settlement-with-student-loan-lender-for-unlawful-practices-through-ai-use-other-consumer-protection-violations.

72 Assurance of Discontinuance, *In the matter of Earnest Operations LLC*, No. 2584-cv1895 (Mass. Super. Ct. July 8, 2025), https://www.mass.gov/doc/earnest-aod/download.

73 Student Borrower Protection Center, *Educational Redlining*, 4 (2020), https://protectborrowers.org/wp-content/uploads/2020/02/Education-Redlining-Report.pdf.

74 Relman Colfax PLLC, *Fourth and Final Report of the Independent Monitor, Fair Lending Monitorship of Upstart Network's Lending Model*, 3, 8-12 (Mar. 27, 2024), https://www.relmanlaw.com/news-upstart-final-report.

search for less discriminatory alternatives.[75] But it declined to adopt the monitor's suggested changes to its model. Upstart objected that those changes would cause a drop in the model's performance—its accuracy in predicting a borrower's risk of defaulting on the loan or paying it off early— while the monitor assessed that the drop was "so small as to not be meaningful" when applied in the real world. The monitor argued that disparate impact law requires a company to alter its model to reduce disparities, even if there's technically a small reduction in accuracy, where—as in this case—the altered model is likely to be equally effective at achieving the company's business needs. The monitor believed a court would interpret federal law this way, as well.[76]

Automated Valuation Models: When people apply for a mortgage to buy a home, refinance, or borrow money against the value of their home to pay for college or startup costs for a new business, the prospective lender has the property appraised. Researchers have found evidence that homes in majority-Black and majority-Latino neighborhoods are valued lower than comparable homes in majority-White neighborhoods; indeed, in home sales, they are more likely to be appraised below the contract price, which represents what the buyer is willing to pay.[77] This evidence is consistent with reported instances of Black families receiving a significantly higher valuation after hiding their family photos and having a White friend appear on

75 *Id.* at 12-13. Upstart defined MSIs as "schools where 80 percent or more of the student body are members of the same racial demographic group." *Id.* at 12.

76 *Id.* at 15.

77 Interagency Task Force on Property Appraisal and Valuation Equity (PAVE), *Action Plan to Advance Property Appraisal and Valuation Equity*, 2-3 (March 2022), https://archives.hud.gov/pave.hud.gov/PAVEActionPlan.pdf; Junia Howell & Elizabeth Korver-Glenn, *The Persistent Evaluation of White Neighborhoods as More Valuable Than Communities of Color* (Nov. 2, 2022); https://static1.squarespace.com/static/62e84d924d2d8e5dff96ae2f/t/6364707034ee737d19dc76da/1667526772835/Howell+and+Korver-Glenn+Appraised_11_03_22.pdf; Andre Perry et al., *The Devaluation of Black Assets: The Case of Residential Property,* Brookings (Nov. 27, 2018), https://www.brookings.edu/articles/devaluation-of-assets-in-black-neighborhoods (finding that "owner-occupied homes in Black neighborhoods are undervalued by $48,000 per home on average").

their behalf for a second appraisal.[78] A low valuation can prevent a family from purchasing a home by raising the downpayment required, cause a lender to deny refinancing, or depress a family's ability to borrow against their home equity to pay for college or start a small business. In these ways, discriminatory appraisals suppress wealth-building and widen racial wealth gaps.

Automated valuation models (AVMs) are algorithms that use statistics and appraisals from comparable properties to estimate the value of a given home based on key data (e.g., square footage, number of bathrooms, yard size, location), without the involvement of an appraiser who visits the property. Their use can result in fairer, more accurate valuations by removing the possibility of conscious or unconscious human bias. However, AVMs can also bake in bias because they are trained on valuations made by human appraisers.[79] To combat this problem, six federal agencies issued a rule setting quality control standards for AVMs. Lenders that use these tools must take steps to ensure accuracy in valuation estimates and compliance with nondiscrimination laws such as the Equal Credit Opportunity Act and the Fair Housing Act, both of which prohibit disparate impact discrimination.[80]

78 *See, e.g.,* Debra Kamin, *Home Appraised With a Black Owner: $472,000. With a White Owner: $750,000,* N.Y. Times (Aug. 18, 2022), https://www.nytimes.com/2022/08/18/realestate/housing-discrimination-maryland.html; Debra Kamin, *Black Homeowners Face Discrimination in Appraisals*, N.Y. Times (Aug. 25, 2020), https://www.nytimes.com/2020/08/25/realestate/blacks-minorities-appraisals-discrimination.html.

79 Michael Neal et al., Urban Institute, *How Automated Valuation Models Can Disproportionately Affect Majority-Black Neighborhoods* (2020), https://www.urban.org/sites/default/files/publication/103429/how-automated-valuation-models-can-disproportionately-affect-majority-black-neighborhoods_1.pdf.

80 Final Rule, Quality Control Standards for Automated Valuation Models, 89 Fed. Reg. 64538 (published Aug. 7, 2024, effective Oct. 1, 2025), https://www.federalregister.gov/documents/2024/08/07/2024-16197/quality-control-standards-for-automated-valuation-models.

IV. Trump's Executive Order: Fundamentally Misunderstanding the Law

On April 23, 2025, President Trump signed Executive Order 14281, *Restoring Equality of Opportunity and Meritocracy*.[81] Its stated goal is "to eliminate the use of disparate-impact liability in all contexts to the maximum degree possible."[82]

The order directed enforcement agencies such as DOJ, EEOC, FTC, and the Consumer Financial Protection Bureau (CFPB) to "deprioritize" enforcement of disparate impact laws, pushing them not just to drop pending cases but also to ask courts to lift consent decrees and injunctions won based on the theory.[83] It also revoked presidential approval for disparate impact regulations under Title VI—the main authority to prevent

81 President Donald J. Trump, Executive Order 14281, Restoring Equality of Opportunity and Meritocracy, 90 Fed. Reg. 17537 (Apr. 23, 2025), https://www.federalregister.gov/documents/2025/04/28/2025-07378/restoring-equality-of-opportunity-and-meritocracy.

82 Id.

83 Agencies had already removed key guidance documents from their websites in the early days of the Trump Administration. *See, e.g.*, EEOC, *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964* (May 18, 2023), available at https://data.aclum.org/storage/2025/01/EOCC_www_eeoc_gov_laws_guidance_select-issues-assessing-adverse-impact-software-algorithms-and-artificial.pdf; U.S. Department of Housing and Urban Development, *Guidance on Application of the Fair Housing Act to the Screening of Applicants for Rental Housing* (April 29, 2024), available at https://www.fairhousingnc.org/wp-content/uploads/2024/08/FHEO_Guidance_on_Screening_of_Applicants_for_Rental_Housing.pdf.

discriminatory uses of federal funds[84]—and directed agencies to formally rescind them.[85] On December 10, 2025, without first publishing a proposed rule or seeking public comment, DOJ issued an immediately effective rule eliminating Title VI's disparate impact provisions, which had governed recipients of federal funding for over half a century.[86] The CFPB has also issued a proposed rule to eliminate disparate impact from regulations implementing the Equal Credit Opportunity Act.[87]

Finally, Trump's April order directed the Attorney General to determine "whether any Federal authorities preempt" state-level disparate impact liability and whether such state laws "have constitutional infirmities that warrant Federal action."[88] In December, he directed the Attorney General to create an "AI Litigation Task Force" to pursue lawsuits challenging state-level regulation of AI, and ordered other agencies to take steps to shore up the tenuous case for preemption.[89] These actions make clear that the administration does not merely intend to shirk its duty to enforce federal anti-discrimination law. It intends to interfere with state laws, as well.[90]

Trump's attack on disparate impact rests on at least three grievous legal errors.

84 *See Alexander v. Sandoval*, 532 U.S. 275 (2001) (holding that Title VI's statutory prohibition on discrimination, Section 601, prohibits only intentional discrimination, and that there is no private right of action to enforce disparate impact regulations promulgated under Section 602, meaning only the federal government can enforce them).

85 Executive Order 14281.

86 Final Rule, Rescinding Portions of Department of Justice Title VI Regulations, *supra* note 7.

87 90 Fed. Reg. 50901 (Nov. 13, 2025), https://www.federalregister.gov/documents/2025/11/13/2025-19864/equal-credit-opportunity-act-regulation-b

88 *Id.*

89 Executive Order 14365, *supra* note 6.

90 Trump issued another order about AI that also warrants comment. Executive Order 14319, *Preventing Woke AI in the Federal Government*, announced that the federal government—the world's largest buyer—would only purchase generative AI systems developed in accordance with "ideological neutrality." 90 Fed. Reg. 35389 (July 23, 2025), https://www.federalregister.gov/documents/2025/07/28/2025-14217/preventing-woke-ai-in-the-federal-government. By way of definition, the order specifies that large language models must not "encode" diversity, equity, and inclusion in their outputs. *Id.* Technologists have rightly pointed out that this mandate "positions one ideological perspective as the default standard for neutrality," and "efforts to align models" with it "risk introducing new distortions." Amy Winecoff & Chinmay Deshpande, Center for Democracy & Technology, *Anti-Woke AI Is a Technical Mirage* (Aug. 8, 2025), https://cdt.org/insights/anti-woke-ai-is-a-technical-mirage. Indeed, by preventing developers from addressing the known biases

LEGAL ERROR #1

First, the order wrongly asserts that under disparate impact liability, "a near insurmountable presumption of unlawful discrimination exists where there are any differences in outcomes in certain circumstances among different races, sexes, or similar groups, . . . even if everyone has an equal opportunity to succeed." The first half of this sentence is a misstatement of the law. The second half misunderstands equal opportunity.

Disparate impact does not turn on "any differences." In fact, under Supreme Court precedent and statutory law, plaintiffs must show that the challenged practice causes a "significant" or "substantial" statistical disparity to avoid having their case immediately dismissed.[91] That is just to make out a prima facie case, meaning showing that the claim appears to have merit on its face. At the second step of the analysis, companies acting in good faith typically do not have a problem establishing that the challenged practice is consistent with business necessity. Then the burden shifts back to the plaintiffs to identify a less discriminatory alternative. Proving the existence of a viable alternative can be complicated and expensive.

If plaintiffs manage to prove all of this, they have proved that the challenged practice unnecessarily harmed them based on race. They have proved that they *didn't* have "an equal opportunity to succeed." Put differently, they have proved discrimination.[92]

91 *Albemarle*, 422 U.S. at 405 (a plaintiff must show that employment tests "select applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants"); 29 C.F.R. § 1607.16(Q) (defining "adverse impact" as a "substantially different rate of selection in hiring, promotion or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group").

92 *See* EEO Leaders' Statement on Disparate Impact, *President Trump's Executive Order on Disparate Impact Analysis Is Legally Incorrect and Will Undermine Meritocracy and Equal Employment Opportunity* (May 2025), https://bit.ly/3F7A6bh ("[T]he entire concept of disparate impact is that unjustified and significant differences in outcome resulting from a 'neutral' policy means that people of different races or sexes are not being given an equal opportunity to succeed.").

LEGAL ERROR #2

Second, the order wrongly asserts that disparate impact requires defendants to "engage in racial balancing." Changing a practice that systematically harms people of a certain race—when the practice doesn't actually serve a company's legitimate interests, or when those interests can be advanced by a less discriminatory practice—is not racial balancing. It is removing an indefensible source of bias. The result is a fairer process for everyone. It is true that the analysis involves doing some math, but as noted, disparate impact doctrine does not demand parity. It allows considerable variation in outcomes before applying any scrutiny at all. In addition, Title VII itself explicitly bans quotas.[93]

It bears noting that outside of zero-sum contexts like hiring, decisionmakers are not making choices between two candidates. They may be assessing the value of a family's home or predicting a patient's cancer risk. Disparate impact law remains relevant to ensure they do not use tools or processes that skew results based on race or other irrelevant traits, and the concept of quotas has no plausible applicability.

LEGAL ERROR #3

Third, the order wrongly asserts that disparate impact is unconstitutional. Its claim that the doctrine "runs contrary to equal protection under the law" has its roots in a concurrence by Justice Antonin Scalia in *Ricci v. DeStefano*,[94] but ignores subsequent Supreme Court decisions.

In *Ricci*, a group of New Haven, Connecticut firefighters sued the city for intentional race discrimination under Title VII. The city had discarded the results of a promotion exam after seeing that almost all those who

93 42 U.S.C. § 2000e-2(j).

94 557 U.S. 557 (2009).

scored high enough were White and none of them were Black. The city explained that it scrapped the results because it feared being sued under Title VII for disparate impact. But the Court found that New Haven did not have a foundation to conclude the test was discriminatory.

Specifically, city officials did not thoroughly evaluate whether the test was job-related and consistent with business necessity, or whether there existed any less discriminatory alternative that served its needs. The city therefore lacked "a strong basis in evidence" to believe it could be held liable under disparate impact for certifying the test results. Under those circumstances, the Court found that its decision not to use the results amounted to disparate treatment of the firefighters who had passed the test (i.e., declining to promote them because of their race).[95]

Scalia went further, suggesting that disparate impact liability itself might be unlawful under the Constitution. He contended that by "requiring employers to evaluate the racial outcomes of their policies, and to make decisions based on (because of) those racial outcomes," disparate impact forces employers to engage in discriminatory "racial decisionmaking."[96]

On this reasoning, any attempt to prevent racially disparate consequences—no matter how traceable to historical discrimination, how grounded in arbitrary considerations, how predictably unjust, or how easy to avoid while still serving an employer's legitimate interests—would itself be racial discrimination.

This is not the law.[97] In a pair of post-*Ricci* cases about University of Texas admissions, not a single Justice questioned the state's "Top Ten Percent" plan—under which the university automatically admitted the top 10% of each

95 *Id.*

96 *Id.* at 594 (Scalia, J., concurring).

97 Zachary Best & Stephen Hayes, *Executive Order on Disparate Impact: An Explainer*, 3 (May 9, 2025) ("No court has ever held that disparate impact runs afoul of the Constitution."), https://www.relmanlaw.com/media/cases/1965_Executive%20Order%20on%20Disparate%20Impact%20Explainer.pdf.

Texas high school—even though its purpose was to increase diversity.[98] As Professor Reva Siegel has observed, although the plan was *race-conscious in its purpose* of creating equal opportunity for students of color, it was constitutional because it was *race-neutral in form* and *did not classify students based on race*. The same is true of disparate impact. The doctrine is race-conscious in that it aims to avert unjustified adverse impacts based on race. But it does not require a decisionmaker to use race as a selection criterion, and it does not advantage or disadvantage any group based on race.[99] In this sense, it is wholly distinct from affirmative action.

Moreover, in 2015 the Supreme Court upheld the existence of disparate impact under the Fair Housing Act in *Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc.*[100] Constitutional arguments abounded in the case, with several amicus briefs arguing that disparate impact law is unconstitutional. The Court was not

98 *Fisher v. Univ. of Texas at Austin (Fisher I)*, 570 U.S. 297 (2013); *Fisher v. Univ. of Texas at Austin (Fisher II)*, 579 U.S. 365 (2016). *See also id.* at 532 (Thomas, J., dissenting) (describing the state law establishing the Top Ten Percent plan as "facially race-neutral law" that "served to equalize competition between students who live in relatively affluent areas with superior schools and students in poorer areas" and "tended to benefit African-American and Hispanic students, who are often trapped in inferior public schools").

99 Reva Siegel, *Race-Conscious but Race-Neutral: The Constitutionality of Disparate Impact in the Roberts Court*, 66 ALA. L. REV. 653, 672-78 (2013). Justice Kennedy, the author of both *Fisher* opinions, had previously explained that policymakers could pursue race-conscious goals through race-neutral means. *See Parents Involved in Cmty. Schs. v. Seattle Sch. Dist. No.* 1, 551 U.S. 701, 789 (2006) (Kennedy, J., concurring in part and concurring in the judgment) ("School boards may pursue the goal of bringing together students of diverse backgrounds and races through other means, including strategic site selection of new schools; drawing attendance zones with general recognition of the demographics of neighborhoods; allocating resources for special programs; recruiting students and faculty in a targeted fashion; and tracking enrollments, performance, and other statistics by race. These mechanisms are race conscious but do not lead to different treatment based on a classification that tells each student he or she is to be defined by race, so it is unlikely any of them would demand strict scrutiny to be found permissible."). Indeed, Justice Scalia himself had acknowledged as much 20 years before *Ricci*. *See City of Richmond v. J.A. Croson*, 488 U.S. 469, 526 (1989) (Scalia, J., concurring in the judgment) ("A State can, of course, act to undo the effects of past discrimination in many permissible ways that do not involve classification by race.") (internal quotation marks omitted).

100 *Tex. Dep't of Hous. & Cmty. Affairs v. Inclusive Cmtys. Project, Inc.*, 576 U.S. 519 (2015).

persuaded.[101] Writing for the majority, Justice Anthony Kennedy explained that "disparate-impact liability has always been properly limited in key respects that avoid the serious constitutional questions that might arise," such as by holding that statistical disparity alone is insufficient to establish liability.[102]

Doubtless, President Trump's order sets back the cause of non-discrimination. It does not change the law, however. Disparate impact remains prohibited under federal statutes—enforceable by state attorneys general and private parties—covering employment, housing, lending, and other spheres of life. Several state laws likewise impose disparate impact liability,[103] and some states have specifically targeted AI systems that have discriminatory effects.[104] The Trump administration's arguments that these statutes may be preempted by or violate federal law are weak and unlikely to prevail.[105]

101 *See* Samuel R. Bagenstos, *Disparate Impact and the Role of Classification and Motivation in Equal Protection Law after Inclusive Communities*, 101 CORNELL L. REV. 1115, 1127-28 (2016) ("Because the Fair Housing Act does not expressly provide for disparate-impact liability, if a majority of the Court had serious constitutional concerns about disparate impact claims *per se*, the Court would likely have avoided the constitutional problem by reading the statute not to provide for such claims. By holding that the Fair Housing Act *does* provide for disparate-impact liability, the Court must therefore have rejected the argument that disparate impact law is unconstitutional.").

102 *Inclusive Cmtys.*, 576 U.S. at 536-37.

103 *See, e.g.*, California Fair Employment and Housing Act, Cal. Gov't Code § 12955.8(b); Colorado Anti-Discrimination Act, C.R.S. §§ 24-34-402, 24-34-502; Illinois Human Rights Act, 775 I.L.C.S. 5/2-102; Mass. Gen. Laws ch. 151B § 4; N.J.S.A. §§ 13:13-2.5, 13:13-3.4(f)(2), 13:13-4.11; Washington Law Against Discrimination, R.C.W. § 49.60.

104 *See, e.g.*, Colorado SB 24-205, Consumer Protections for Artificial Intelligence, codified at C.R.S. § 6-1-1701 et seq. (2024), https://leg.colorado.gov/bills/sb24-205; Illinois H.B. 3773, amending the Illinois Human Rights Act (2024), https://legiscan.com/IL/bill/HB3773/2023; New Jersey Attorney General, Division on Civil Rights, *Guidance on Algorithmic Discrimination and the New Jersey Law Against Discrimination* (2025), https://www.nj.gov/oag/newsreleases25/2025-0108_DCR-Guidance-on-Algorithmic-Discrimination.pdf.

105 *See* Charlie Bullock, *supra* note 8 (stating that the case for preemption under the Commerce Clause is "legally dubious and unlikely to succeed in court); Gibson Dunn, *President Trump's Latest Executive Order on AI Seeks to Preempt State Laws* (Dec. 15, 2025), https://www.gibsondunn.com/president-trump-latest-executive-order-on-ai-seeks-to-preempt-state-laws (explaining why DOJ's preemption arguments "are unlikely to be successful" and why the contemplated FCC and FTC actions would not be a basis for preemption).

V. Conclusion

As AI increasingly determines our access to jobs, housing, credit, and more, the disparate impact doctrine stands as an essential safeguard against algorithmic discrimination. It offers a flexible and constitutionally sound framework to root out bias. Far from hindering innovation or imposing quotas, disparate impact liability helps identify hidden and unjustified barriers that disadvantage people based on demographic factors. The doctrine demands only reasonable changes consistent with legitimate business needs, and it incentivizes developers to design fairer systems from the outset. Attempts to dismantle this protection misunderstand both the law and the technical realities of algorithmic bias.

Given current presidential opposition, others must step up to ensure disparate impact serves as an effective check on AI-based discrimination. State legislators can codify disparate impact into state civil rights statutes, ensuring its durability against federal rollback. State attorneys general should bring more enforcement actions under state and federal law. Civil society organizations can also pursue strategic litigation and conduct public education campaigns to counter mischaracterizations of the law. Industry should adopt proactive compliance measures such as impact assessments, searches for less discriminatory alternatives, and disclosures to increase transparency. State and local governments can leverage their procurement power to demand these measures for AI systems they purchase. Ultimately, federal law should provide more comprehensive disparate impact protection, as well.

These strategies are critical to ensuring the AI revolution advances equal opportunity for all rather than entrenching and scaling discrimination.

the CENTER for Civil Rights and Technology

The Leadership Conference on Civil and Human Rights

1620 L Street NW, Suite 1100
Washington, DC 20036

(202) 466-3434

civilrights.org/value/center-civil-rights-technology/

@civilandhumanrights

@civilrightsorg

Pod for the Cause

Copyright © 2026
The Leadership Conference
Center for Civil Rights and Technology
All Rights Reserved