Stanford Law Review Online

Volume 79 September 2026

ESSAY

Disparate (Algorithmic) Advantage

Yunsieg P. Kim & Jacob Eisler*

Introduction

When a human gatekeeper generates disparate outcomes, the causes are opaque, the data are noisy, and isolating the source of the disparity from "holistic judgment" is nearly impossible. When an algorithm produces a disparity, practices are specified, outputs are reproducible, and biases can be measured under controlled conditions. Yet scholars argue that algorithms make civil rights enforcement *harder*, not easier. The concerns are familiar: biased training data reproduces historical discrimination at scale,[1] opaque models defeat accountability,[2] and the elements of disparate impact do not map cleanly onto algorithmic systems.[3] This anxiety has produced a near-consensus

* Kim is an Associate Professor of Law, Maurice A. Deane School of Law at Hofstra University. J.D., Yale Law School; Ph.D., University of Michigan; M.S. in Cybersecurity, New York University Tandon School of Engineering. Eisler is the James Edmund and Margaret Elizabeth Hennessey Corry Professor at Florida State University College of Law. We thank Haven Branca, Tiffany Chen, Alexander Holt, and Joseph Serrone for their research assistance and the editors of the *Stanford Law Review Online* for their diligent work.

1. *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 671-701 (2016) (identifying mechanisms by which data mining produces discriminatory outcomes, including biased training data and proxy discrimination); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2218 (2019) ("In a racially stratified world, any method of prediction will project the inequalities of the past into the future.").
2. *See* Brandon L. Garrett & Cynthia Rudin, *The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice*, 109 CORNELL L. REV. 561, 563-64 (2024) (" '[B]lack box' AI designed to be non-interpretable[] mean[s] that its processes cannot be fully understood by laypeople or even by experts . . . . [In] the criminal justice system . . . black box AI poses risks to both public safety and to fundamental human and constitutional rights.").
3. *See* Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 907 (2017) (arguing that disparate impact doctrine is "a poor fit" for algorithmic employment decisions because the statutory framework's elements do not map cleanly onto data-driven systems); Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 629-30 (2021) (algorithmic opacity and trade secret protection obstruct disparate impact claimants).

that existing civil rights law cannot keep up with algorithmic decisionmaking, and that new law is needed.[4]

This consensus has it backwards. For directional disparities, disparate impact law is better suited to algorithms than to the humans it was designed to police.[5] Algorithms apply the same criteria to every input, producing the cleanest statistical evidence the law has ever seen. Their outputs are deterministic and reproducible, making it possible to test a gatekeeper's justification for a practice and to propose concrete, less exclusionary alternatives. And because an algorithm is a single, flat process rather than a web of cognition, culture, and bias, it is identifiable as the particular practice that statutes require plaintiffs to specify.[6]

Even the most commonly cited challenge—that algorithms are "black boxes"[7]—gets the problem backwards: human cognition is the *original* black box, and disparate impact was built to police opaque decisions through outcomes, not mechanisms.[8] If this structural advantage has not produced

4. *See, e.g.*, Robert Bartlett, Adair Morse, Richard Stanton & Nancy Wallace, *Consumer-Lending Discrimination in the FinTech Era*, 143 J. FIN. ECON. 30, 32 (2022) (identifying racial disparities in algorithmic lending); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1301-13 (2008) (proposing safeguards for automated government decisionmaking).

5. This Essay synthesizes two research agendas. One of us has argued that U.S. consumer protection suffers from a structural detection gap: The legal system's picture of harm is drawn from the subset of injuries that become visible enough to litigate, and this conspicuous subset is mistaken for the whole. Yunsieg P. Kim, *Incognito Consumer Harm*, 104 WASH. U. L. REV. (forthcoming Apr. 2026) (manuscript at 3-4) (on file with author). The other has developed a constitutional framework establishing disparate impact's permissibility under contemporary equal protection doctrine. Jacob Eisler, *Racial Gateways*, 79 ALA. L. REV. (forthcoming Jan. 2026) (manuscript at 5-7) (on file with authors). These frameworks suggest that the doctrine is structurally sound but operationally starved of information. We use "disparate impact" here to refer to directional disparities between protected groups. This Essay does not address algorithmic harms that manifest as convergence rather than skew. For that account, see Yunsieg P. Kim & Jacob Eisler, *Artificial Sameness* (August 5, 2026) (working manuscript) (on file with authors).

6. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (requiring complainants to demonstrate "a particular employment practice" causing a disparate impact); *see* Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338, 355-60 (2011) (plaintiffs failed to identify a sufficiently uniform "practice" to establish commonality across 1.5 million claims)*; see also* Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, 576 U.S. 519, 535 (2015) (extending disparate-impact liability to the Fair Housing Act by analogy to Title VII).

7. *See, e.g.*, Landyn Wm. Rookard, *The Common Threats of Artificial Intelligence and Privatization*, 12 TEX. A&M L. REV. 831, 870 (2025) ("[A]dvanced machine learning algorithms may be black boxes by their nature, making it inherently impossible to trace their logic from the inputs to their outputs."); Benedict Sheehy & Yee-Fui Ng, *The Challenges of AI Decision-Making in Government and Administrative Law: A Proposal for Regulatory Design*, 57 IND. L. REV. 665, 672 (2024) ("[T]he inability to explain AI decision-making is a significant problem when it comes to giving reasons.").

8. *See* Griggs v. Duke Power Co., 401 U.S. 424, 431-32 (1971) (holding that Title VII "proscribes not only overt discrimination but also practices that are fair in form, but

footnote continued on next page

successful litigation, the reason is not legal but informational: plaintiffs cannot challenge what they cannot see. But the civil rights community has misdiagnosed information asymmetry as legal inadequacy. Algorithms may be preferable for the pursuit of equity not because they are fair, but because they make disparities easier to expose.

The wider status of disparate impact remains contested,[9] but the prevalent view is that the doctrine is useful for smoking out discriminatory intent that would otherwise escape proof.[10] Whether it also serves a broader structural-correction function is a question this Essay need not reach. Its claim holds on the common ground: disparate impact is easier to detect when the decisionmaker is an algorithm, because algorithms are easier to test under controlled conditions.

This Essay proceeds in two parts. Part I identifies three properties of algorithmic decisionmaking—consistency, replicability, and flatness—that make algorithms more amenable to disparate impact analysis than the human judgment that disparate impact was originally built to police. Part II explains why this legal advantage has not been realized in practice: plaintiffs lack the data necessary to invoke a framework that is structurally sound but operationally blind. The priority is not new law for the algorithmic age, but new infrastructure to activate the law we already have.

---

discriminatory in operation" and evaluating the employer's practices by their consequences, not their internal rationale). The greater ease of identifying algorithmic disparate impact is especially strong if the "core" purpose is taken to be "smoking out" hidden or unconscious discriminatory intent. See *infra* note 10 and accompanying text.

9. *See* Inclusive Cmtys., 576 U.S. at 540 (disparate-impact liability "plays a role in uncovering discriminatory intent" by allowing plaintiffs to reach biases that evade disparate-treatment classification); *cf.* Ricci v. DeStefano, 557 U.S. 557, 594-95 (2009) (Scalia, J., concurring) (questioning whether disparate-impact provisions are consistent with the Equal Protection Clause); Washington v. Davis, 426 U.S. 229, 239-42 (1976) (discriminatory purpose required under the Equal Protection Clause). For a discussion of the broader debate, 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, 1136 (2016). In a recent Office of Legal Counsel memo, the Trump administration indicated that it would no longer pursue disparate impact claims, based on an interpretation of the constitutionality of disparate impact including the cases such as *Ricci*. Constitutionality of Disparate-Impact Liability Under Title VII, 50 Op. O.L.C. (June 9, 2026) (slip op. at 1-2). Notably, however, while such memos might guide the executive, they can be overruled by later executive instruction. Trevor W. Morrison, *Stare Decisis in the Office of Legal Counsel*, 110 COLUM. L. REV. 1448, 1451 (2010).

10. *See* Bagenstos, *supra* note 9, at 1135 (discussing Justice Scalia's concurrence in *Ricci*, 557 U.S. at 594, suggesting that the most legitimate purpose of disparate impact is as an evidentiary tool to identify discriminatory activity that would otherwise be undetectable).

I. Three Algorithmic Properties

An algorithm is a set of instructions that takes an input value or values and, through execution of some defined instructions or steps, thereby produces a determined output value. In addition, these instructions must be executed by a computational mechanism that requires no additional judgment.[11] The most familiar "simple" algorithms implement operations that are visibly entirely deterministic, yielding the same output given the same input.[12] The types of algorithms that drive contemporary artificial intelligence (AI) share the essential properties of all algorithms—they take an input, perform a fixed operation, and yield an output determined by the input and operation—but the operation itself is a series of calculations, with each calculation conditioning the next step.[13] The probabilistic nature of each individual calculation explains why the same input put into the same AI algorithm can yield different outputs,[14] but also explains why those outputs *tend* to have similar substance.[15] However, the complexity of contemporary AI does not change its character as fundamentally algorithmic. The neural networks conditioned by repeated exposure to existing data (known as deep learning training) that underlie contemporary AI technology remain an operation (or, more precisely, a series of operations) consisting of an input, a fixed set of computational instructions, and an output determined by these two elements.[16]

---

11. *See generally* Julia de Miguel Velazquez & J. Ángel Velázquez-Iturbide, *What Is an Algorithm? Traditional vs. Intelligent Algorithms*, *in* ENCYCLOPEDIA OF INFORMATION SCIENCE AND TECHNOLOGY (Mehdi Khosrow-Pour ed., 6th ed. 2025) (collecting sources which support this definition).

12. *Id.* at 4.

13. For the seminal paper that has driven the current AI boom, see Ashish Vaswani et al., *Attention Is All You Need*, *in* 1 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30, at 5999, 6000 (2017) (describing the "transformer" that iteratively performs a series of probabilistic calculations to generate an output, with each subsequent iteration of the operation dependent upon prior states).

14. For a description of how the technical foundation of contemporary AI is "'attention," which is a probabilistic weighting of a given outcome, see Dave Bergmann & Cole Stryker, *What Is an Attention Mechanism?,* IBM, https://perma.cc/Q43H-EXE8 (archived Aug. 15, 2026).

15. *See* Kai Riemer & Sandra Peter, *Conceptualizing Generative AI as Style Engines: Application Archetypes and Implications*, 79 INT'L J. INFO. MGMT. no. 102824, at 3, 12 (2024) (describing generative AI as probabilistically producing novel content based on patterns, yielding "sophisticated style[s]" but also retaining some capacity for "unpredictability.").

16. For a seminal technical discussion of the probabilistic nature of deep learning, see IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, DEEP LEARNING 128-33, 178-81 (2016). For a discussion of how such probabilistic neural network technology informs contemporary AI algorithms, see CHRISTOPHER M. BISHOP & HUGH BISHOP, DEEP LEARNING: FOUNDATIONS AND CONCEPTS 1-4 (2024). For a less technical discussion of the nature of neural networks and how they inform deep learning, see Larry Hardesty, *Explained: Neural Networks*, MIT NEWS (Apr. 14, 2017), https://perma.cc/872B-82BH.

For the purposes of this Essay, the simplest as well as the most complex algorithms share three properties that make them better suited to disparate impact analysis than human decisionmakers.[17] Algorithms are *consistent*: they apply the same criteria to every input. They are *replicable*: their outputs can be reproduced and tested under controlled conditions. Finally, they are *flat*: an algorithm is a unified, fixed process, not a multifaceted system of cognition, culture, and bias. Each of these properties resolves a problem that has long undermined disparate impact claims against humans. While disparate impact claims do not comprise a constitutional harm,[18] a variety of regulatory frameworks provide remedies for otherwise unjustified disparate impact. At the core of establishing liability for a disparate impact claim, it must be shown that the conduct of the relevant actor (employer, housing provider, and so forth) produced a disproportionate unfavorable impact on the relevant class without justification.[19] Thus a quantified demonstration of impact of a practice on a vulnerable group is the essence of a disparate impact claim.

A. Consistency

An algorithm applies the same criteria, weighted the same way, to every input. A human decisionmaker does not. A hiring manager's evaluation of two identical resumes may differ depending on the time of day, the order in which they are reviewed, and the manager's mood, fatigue, and implicit associations. These variables are uncontrolled and unobservable. Algorithms' behavior, in contrast, is "consistent: given the same inputs, they should reliably produce the same outputs without significant variation."[20] Even large language models (LLMs) like ChatGPT, which introduce controlled randomness, draw from a fixed distribution: individual outputs may vary, but aggregate patterns across a population remain stable and testable. One study, for example, used prominent LLMs to perform 500 iterations of evaluation of approximately 361,000

17. As de Miguel Velazquez and Velázquez-Iturbide note, one of the defining features of algorithms is that they *exclude* the discretionary use of human judgment. *See* de Miguel Velazquez & Velázquez-Iturbide, *supra* note 11, at 4-5.

18. Washington v. Davis, 426 U.S. 229, 239 (1976) ("[O]ur cases have not embraced the proposition that a law or other official act, without regard to whether it reflects a racially discriminatory purpose, is unconstitutional solely because it has a racially disproportionate impact").

19. Griggs v. Duke Power Co., 401 U.S. 424, 432-33 ("[T]he Act [creates liability] for consequences of employment practices, not simply motivation…The facts of this case demonstrate the inadequacy of broad and general testing devices as well as the infirmity of using diplomas or degrees as fixed measures of capability.").

20. Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 463-64 (2023).

resumes, yielding consistent results with regards to racial and gender disparities.[21]

This consistency resolves the statistical proof problem that has long plagued disparate impact claims.[22] When human decisionmaking is inconsistent, as it often is, the signal is noisy and confounded. It is difficult to demonstrate that a pattern of outcomes reflects a single practice rather than a collection of idiosyncratic decisions. When an algorithm is consistent, as it always is, the signal is clean: run the model on a population, disaggregate the outputs by race, and perform a standard significance test.[23] Algorithms produce the cleanest evidentiary environment that disparate impact has ever encountered.[24]

But clean statistical evidence is only the first step. The disparate impact framework also requires an identifiable practice and no defensible justification, or testable alternatives.[25] Algorithmic replicability delivers all three.

---

21. Jiafu An, Difang Huang, Chen Lin & Mingzhu Tai, *Measuring Gender and Racial Biases in Large Language Models: Intersectional Evidence from Automated Resume Evaluation*, 4 PNAS NEXUS, no. 3, 2025, at 1, 2-5.

22. *See Griggs*, 401 U.S. at 431 (explaining that statutory civil rights regulations may prohibit conduct that produces disparate impact, but the nature of any conduct that generates disparate-impact liability must be "invidious[]" and an unjustified "barrier"); Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, 576 U.S. 519, 542 (2015) ("A robust causality requirement ensures that '[r]acial imbalance . . . does not, without more, establish a prima facie case of disparate impact' and thus protects defendants from being held liable for racial disparities they did not create." (quoting Wards Cove Packing Co. v. Atonio, 490 U.S. 642, 653 (1989))).

23. *See, e.g.*, Bartlett et al., *supra* note 4, at 34-45, 53 (Latino and Black borrowers paid approximately five-eight basis points more on home-purchase loans).

24. *See* Eisler, *supra* note 5, at 46-47 (observing that the difficulty facing the disparate impact regime is extracting a sufficient causal account related to a protected category in the absence of intentional discrimination and developing an efficient-cause framework establishing disparate impact's permissibility under contemporary equal protection doctrine). Because AI identifies correlates in a pure fashion, it can do so more convincingly than a human actor who may have a confounding moral or policy agenda. *Id. See also* Louisiana v. Callais, 146 S. Ct. 1131, 1146-67 (2026) (synthesizing the narrowness by which the current Supreme Court interprets current statutory efforts that prohibited discriminatory racial effect where there is an absence of established discriminatory intent). The purity of AI-demonstrated disparate impact answers an evidentiary aspect of this challenge. *Id.*; Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll., 143 S. Ct. 2141, 2173 (2023) (finding race-conscious admissions programs violate the Equal Protection Clause including when they work to ameliorate social discrimination, thereby showing the headwinds facing disparate impact regimes with regards to protected classes).

25. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (requiring identification of a particular employment practice and imposing on defendants the burden of demonstrating job-relatedness and business necessity); *id.* § 2000e-2(k)(1)(A)(ii) (complainants may prevail by demonstrating an alternative employment practice that respondent refuses to adopt).

B. Replicability

Disparate impact requires plaintiffs to identify a "particular employment practice" that causes the challenged disparity.[26] That requirement has been the doctrine's most persistent obstacle. *Wal-Mart Stores, Inc. v. Dukes* held that a policy of delegating discretion to individual managers was not a sufficiently identifiable "practice."[27] The problem was that the alleged discrimination was dispersed: thousands of managers made subjective calls in thousands of different ways, producing a pattern too diffuse to attribute to any single practice. With human decisionmakers, the challenged "practice" is a cognitive process that cannot be isolated, specified, or re-run.

An algorithm eliminates this problem. An algorithm *is* a particular employment practice: a specified, centralized process that applies defined criteria to every input. Because its outputs are deterministic, the algorithm can be run on a population, re-run with altered inputs, and shown to be the same process producing materially identical results. In *Mobley v. Workday, Inc.*, a federal court described an AI screening system as a "unified policy" sufficient to support a nationwide collective action.[28] The algorithm was exactly the kind of identifiable practice that was missing in *Wal-Mart*.

Replicability also makes the subsequent stages of the burden-shifting framework tractable. Once a plaintiff establishes a prima facie case, the burden shifts to the defendant to demonstrate that the challenged practice is job-related and consistent with business necessity.[29] When the practice is a human manager's "holistic judgment," the defense is vague and difficult to challenge with specificity. When the practice is an algorithmic model, the defendant must explain why these particular features, weighted in this particular way, producing these particular decision boundaries, are necessary. The specificity of the model compels specificity in the defense. When the model is an AI system whose internal reasoning the defendant itself cannot fully explain, the

26. This language underlies the statutory basis of the contemporary disparate impact regime. Civil Rights Act of 1964, Pub. L. No. 88-352, § 703(k)(1)(A)(i), 78 Stat. 254, 255 (codified as amended at 42 U.S.C. § 2000e-2(k)(1)(A)(i)).

27. Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338, 355-57 (2011). The Court acknowledged that delegated discretion can constitute a cognizable employment practice, Watson v. Fort Worth Bank & Tr., 487 U.S. 977, 990 (1988), but held that plaintiffs failed to identify a common mode of exercising discretion sufficient for Rule 23(a)(2) commonality. *Dukes*, 564 U.S. at 356.

28. Mobley v. Workday, Inc., 740 F. Supp. 3d 796, 804 (N.D. Cal. July 12, 2024) (allowing disparate impact claims to proceed against Workday under an agent liability theory); Mobley v. Workday, Inc., No. 23-cv-00770, 2025 WL 1424347, at *5, *10 (N.D. Cal. May 16, 2025) (granting preliminary certification and describing Workday's AI recommendation system as a "unified policy" applicable to all putative collective members).

29. 42 U.S.C. § 2000e-2(k)(1)(A)(i); *see* Albemarle Paper Co. v. Moody, 422 U.S. 405, 425-36 (1975) (operationalizing the business necessity standard by validating employment tests against job performance).

burden becomes heavier still.[30] A company that deploys a model it cannot interpret will struggle to carry that burden.

The same logic applies to the statutory requirement that plaintiffs identify a less exclusionary alternative.[31] With a human decisionmaker, that requirement is practically meaningless as to discrete actors: one cannot propose a less exclusionary "alternative" to a cognitive process because one cannot redesign how a person thinks. The only alternative is elaborate institutional redesign that raises its own knock-on effects regarding human decisionmaking and which in any case raises questions regarding constitutionality.[32] With an algorithm, concrete alternatives are routine—remove a feature correlated with race, retrain on a more representative dataset, or substitute a simpler model.[33] Replicability means that each alternative can be tested and the difference measured.[34] The doctrine's burden-shifting framework, originally designed for subjective human judgment, becomes *more* tractable when the decisionmaker is a machine. What remains is the most fundamental objection, and the one most frequently directed at AI: that the machine is a black box.

C. Flatness

The black box objection has dominated the scholarly response to algorithmic disparate impact.[35] This objection may appear to be the strongest

30. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1094-1100 (2018) (distinguishing inscrutability from non-intuitiveness as sources of the "black box" problem); Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y, Jan.-June 2016, at 1, 1-5 (model opacity is distinct from corporate secrecy or technical illiteracy).

31. Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 762 (2006) (describing the statutory role of less discriminatory alternatives built into the current statutory regime).

32. Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943, 952-958 (1995) (explaining that attempts to regulate values often have unpredictable and countervailing effects); Ricci v. DeStefano, 557 U.S. 557, 595 (2009) (questioning the constitutionality of deliberately pursuing disparate-impact-seeking policies).

33. *See* de Miguel Velazquez & Velázquez-Iturbide, *supra* note 11, at 3, 9-10 (explaining that multiple algorithms may solve the same problem, that designers test different input data, and that model performance depends substantially on the quantity and quality of training data).

34. *See* Emily Black, John Logan Koepke, Pauline T. Kim, Solon Barocas & Mingwei Hsu, *Less Discriminatory Algorithms*, 113 GEO. L.J. 53, 62-71 (2024) (arguing that "model multiplicity"—the existence of many comparably accurate models for a given prediction task—means that less discriminatory alternatives will often be available and testable); *cf.* 42 U.S.C. § 2000e-2(k)(1)(A)(ii) (requiring complainants to identify a less discriminatory "alternative employment practice").

35. *See* Garrett & Rudin, *supra* note 2, at 567-68; Rookard, *supra* note 7, at 868-70; Sheehy & Ng, *supra* note 7, at 672-73; Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1829 (2019) ("A recurrent concern about machine learning algorithms is that they operate as 'black boxes.'").

for AI systems. Deep learning models can contain billions of parameters, and even their developers cannot fully explain the reasoning process that produces a given output.[36] If we cannot see inside the model, we cannot hold it accountable. But human cognition is the *original* black box: opaque, unreplicable, and resistant to inspection. Disparate impact was designed from the beginning to police decisionmaking processes whose internal workings could not be observed.[37]

The reason that the human black box is deeper than the algorithmic one is not simply opacity. It is that human decisionmakers are multifaceted systems. A hiring manager's decision reflects the interaction of explicit criteria, implicit bias, institutional culture, interpersonal dynamics, emotional states, and cognitive heuristics: multiple overlapping motivational systems, each capable of producing exclusionary outcomes independently. These multiple facets make it extraordinarily difficult to isolate which aspect of the decisionmaker's process produced the disparity. By comparison, algorithms are flat: one model, one set of inputs, one set of weights, and one output, with no overlapping motivational systems whose independent contributions must be disentangled. An AI neural network with billions of parameters and multiple iterative rounds of probabilistic calculation is computationally complex, but motivationally flat and cognitively homogenous—a uniform process with defined inputs and a defined output, not a web of cognitive and cultural systems. An AI operates by repeatedly performing the *same* attention-maximizing operation iteratively, rather than, like a person, having to reconcile multiple fundamentally incommensurable values into a single judgment and subsequent consolidated, synthetically motivated action. While humans are cognitively reflexive, AI algorithms are cognitively iterative. This comparative flatness manifests at multiple points in the AI technical architecture. At the most computationally granular level, contemporary generative AI outputs rely upon repeat iterations of a single type of operation, attention maximization guided by the probabilistic SoftMax function.[38] At the level of operation, an AI outputting results is simply performing a settled function informed by fixed weights after being trained—unlike a person, it is not adaptively reflexive as part of its ongoing constitutive nature.[39] Finally, at

36. *See* Tomek Korbak et al., Chain of Thought Monitorability: A New and Fragile Opportunity for AI Safety 2 (Dec. 7, 2025) (unpublished manuscript), https://perma.cc/Y3VS-HRBT (underscoring how "[t]he opacity of advanced AI agents underlies many of their potential risks" and explaining that even "CoT [chain of thought] reasoning traces"—the most advanced method for interpreting AI reasoning—are "incomplete representations" of the model's actual reasoning process). This source was authored by researchers representing OpenAI, Google, and Anthropic, among others.

37. *See* Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971) ("Congress directed the thrust of the Act to the *consequences* of employment practices, not simply the motivation.").

38. Vaswani et al., *supra* note 13, at 4.

39. *Compare* GOODFELLOW ET AL., *supra* note 16, at 96-98 (describing the fixed nature of an AI output), *with* CHRISTINE M. KORSGAARD, SELF-CONSTITUTION: AGENCY, IDENTITY,

footnote continued on next page

the most philosophically sophisticated level, AIs do not possess the autonomous capacity to interrogate their moral purposes. This can be framed either deontologically (AI cannot, given its current technical algorithmic nature, engage in truly autonomous reasoning) or teleologically (an AI's purpose is merely to implement its technical coding, rather than to realize a "deeper" telos).[40]

Yet for the purposes of identifying disparate impact violations, the flatness of algorithms is a feature, not a bug. Disparate impact is an outcomes-based doctrine, and it is on strongest constitutional footing when those outcomes can be unequivocally established and have no alternative explanation. The prima facie case requires showing *that* a practice produces disparate results, not *how* it does so internally.[41] That showing is straightforward even when the model is a complete black box: run it on a dataset, disaggregate the outputs by race, and measure the disparity.[42] The model's internal opacity is irrelevant to this analysis because the analysis operates entirely on inputs and outputs. This is true whether the challenged selection system is a simple algorithm or an AI model like ChatGPT, so long as the legal question is whether its outputs produce directional disparity.[43] Moreover, at the business necessity stage, AI opacity does not help the defendant. A company deploying a model whose reasoning even its own engineers cannot explain faces a steeper burden to demonstrate that the practice is job related. The doctrine thus creates incentives for interpretability, which is the very outcome that the explainable AI community has been pursuing, now backed by legal force.[44]

AND INTEGRITY 116 (2009) (exploring, by contrast, the ongoing process of self-reflection characteristic of human morality).

40. This cuts to the foundations of human nature. For the idea that the basis of human morality is autonomy, see IMMANUEL KANT, GROUNDWORK OF THE METAPHYSICS OF MORALS 41-47 (Mary Gregor ed. & trans., Cambridge Univ. Press 1998) (1785). For the teleological foundation, see ARISTOTLE, DE ANIMA (ON THE SOUL) 165-66 (Hugh Lawson-Tancred trans., Penguin Books 1986). Some contemporary philosophers would argue that human cognition and moral ontology is reductive in a way that makes it more analogous to AI, in which case only the features of complexity and plasticity might differentiate humans from current AI systems. *See* DEREK PARFIT, REASONS AND PERSONS 210-11 (rev. ed. 1986); DANIEL C. DENNETT, BRAINSTORMS: PHILOSOPHICAL ESSAYS ON MIND AND PSYCHOLOGY 3 (1978). In such cases, the two previous limitations of AI (that it lacks metaphysical basis for deontological autonomy and that it lacks intrinsic moral telos) still stand, but from the perspective of these philosophers, the metaphysical uniqueness of human cognition vanishes.

41. *See Griggs*, 401 U.S. at 431-32. Even Andrew D. Selbst and Solon Barocas's category of "inscrutability" does not defeat the prima facie case. *See* Selbst & Barocas, *supra* note 30, at 1094-1100.

42. *See, e.g.*, An et al., *supra* note 21, at 2 (performing such an analysis of resume scores given by commercial LLMs).

43. *See id.*

44. *See* Selbst & Barocas, *supra* note 30, at 1113-24 (arguing that the demand for explainability in machine learning reflects legitimate normative concerns about autonomy and accountability, but that the appropriate response varies by context).

This structural simplicity also addresses the judicial anxieties driving the Supreme Court's narrowing of effects-based liability. *Inclusive Communities* warned that disparate impact must not become a means to second-guess valid governmental and private priorities.[45] *Brnovich v. Democratic National Committee* narrowly interpreted Section 2 of the Voting Rights Act's results test in a manner reflecting broader discomfort with effects-based claims overriding legitimate state choices.[46] Both cases reflect anxiety about liability regimes that second-guess human discretion. That concern lacks force when the challenged practice is not a judgment call, but a model. There is no human discretion to chill: There is only a system to audit. Such modifications are more readily implemented without pushback at the level of initial design and do not face the unpredictability of nuanced human social responses.

As more decisions migrate from humans to algorithms, the factual predicates that trouble the Court will gradually disappear. Subjective discretion gives way to specified criteria. The impossibility of proposing concrete alternatives gives way to model multiplicity. At the detection stage, concerns about de facto quotas are reduced because the inquiry turns on measurable, auditable outputs. This allows disparate impact implementation to edge towards unequivocally performing the function of "smoking out" otherwise undetectable but also inexplicable discrimination hidden in facially neutral procedures. This is the purpose of disparate impact that is less controversial and also on constitutionally firmer footing, because it is more readily reconcilable with the prohibition of deliberate discrimination.[47]

Algorithmic detection of disparate impact could serve as the critical tool for pursuing this aim. The law that the Court is narrowing for human contexts may be most easily applied to algorithmic ones. Algorithms could be designed to identify disparate impacts in a manner that minimizes the role of discretionary and thus prospectively partial interests or ideologically motivated human beliefs. An algorithm-led approach for disparate impact, in other words, could create an objective and neutral process for finding unjustified disparities. The retrenchment of disparate impact doctrine is, paradoxically, clearing the ground for exactly the kind of claims that even the most ardent disparate impact skeptics should accept.[48]

---

45. *See* Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, 576 U.S. 519, 540 (2015) ("The FHA is not an instrument to force housing authorities to reorder their priorities."); *id.* at 544 (insisting that disparate impact liability must not "displace valid governmental and private priorities").

46. *See* Brnovich v. Democratic Nat'l Comm., 141 S. Ct. 2321, 2338-39 (2021) (adopting a multifactor test for Section 2 of the Voting Rights Act that significantly raised the bar for effects-based challenges to voting regulations). The Court distinguished Section 2's results test from Title VII-style disparate impact. *Id.* at 2340-41.

47. *See* Bagenstos, *supra* note 9, at 1135 (discussing Justice Scalia's narrow conception of when disparate impact might be a valid doctrine).

48. The constitutional permissibility of effects-based liability remains contested. *See* Eisler, *supra* note 5, at 46-47 (assessing the state of the doctrine); *cf.* Washington v. Davis, 426

footnote continued on next page

II. Starved of Information

The argument thus far still has an obvious weakness. If disparate impact is structurally better suited to algorithms than to humans, why has no wave of successful algorithmic disparate impact litigation materialized? The answer is not legal but informational: The law is starved of the information it needs.

A. Two Barriers

Consider what information a plaintiff must have in hand to bring a disparate impact claim against an algorithmic system, and what information is systematically unavailable.

The first barrier is identification. In many employment, housing, and lending contexts, applicants do not know that an algorithm processes their applications, let alone which algorithm produces the disparity. Even a plaintiff who suspects algorithmic screening cannot see the decision pipeline: which particular algorithm does what in a chain of intake screening, resume-parsing, and final scoring.[49]

The second barrier is evidence. Even plaintiffs who know that an algorithm was used lack access to aggregate outcome data disaggregated by, say, race or gender. They see only their own outcomes, accepted or rejected, and cannot see the statistical distribution across demographic groups. Performing

U.S. 229, 239-48 (1976) (requiring discriminatory purpose under the Equal Protection Clause while preserving statutory effects-based claims under Title VII); *Inclusive Cmtys.*, 576 U.S. at 540, 544 (sustaining statutory disparate impact while cautioning against overreach).

49. *See* Kim, *supra* note 5, at 34-37 (describing the "survivorship bias" in enforcement: how the legal system's picture of harm is drawn from the subset of injuries that become visible enough to litigate). In *Mobley*, the plaintiff applied to over 100 jobs using Workday's AI platform and was rejected by all of them, often during non-business hours and once in less than an hour. Mobley v. Workday, Inc., 740 F. Supp. 3d 796, 803, 811 (N.D. Cal. 2024). In *iTutorGroup*, the EEOC challenged an algorithm that automatically rejected applicants based on age. Consent Decree at 1, 15, EEOC v. iTutorGroup, Inc., No. 22-cv-02565 (E.D.N.Y. Sept. 8, 2023), 2023 WL 6261089, ECF No. 26 (approving a $365,000 settlement). *See also* Settlement Agreement and Final Judgment at 2, United States v. Meta Platforms, Inc., No. 22-cv-05187 (S.D.N.Y. June 27, 2022), ECF No. 7 (imposing a $115,054 civil penalty); Press Release, U.S. Dep't of Just., Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising (June 21, 2022), https://perma.cc/ZQF6-K5SU (describing the lawsuit as "the department's first case challenging algorithmic bias under the Fair Housing Act"); *see also* Miriam Vogel, Michael Chertoff, Jim Wiley & Rebecca Kahn, *Is Your Use of AI Violating the Law? An Overview of the Current Legal Landscape*, 26 N.Y.U. J. LEGIS. & PUB. POL'Y 1029, 1062 (2024) (identifying the Meta action as DOJ's first case challenging algorithmic bias under the Fair Housing Act).

the analysis necessary to establish a prima facie case requires both data and technical capacity that many individuals and advocacy organizations lack.[50]

These information barriers are not unique to algorithms. Plaintiffs challenging human decisionmakers also lack access to the "deep" causal driver of a disparity and to aggregate outcome data. But with algorithms, each barrier is *informational*, not legal. The law could handle the case, but cases never get filed because necessary information is unavailable. Courts have created doctrinal friction such as standing limitations, trade secret objections to discovery, and uncertain vendor liability,[51] but these problems are secondary. The dominant constraint is that much algorithmic discrimination is never detected at all.[52] For directional disparate impact, this problem is entirely solvable. The same consistency, replicability, and flatness that make the doctrine work also make detection infrastructure feasible: An algorithm can be systematically tested in ways that a human mind cannot.

B. Detection-First Reforms

The civil rights community has misdiagnosed information asymmetry as legal inadequacy, proposing new law when existing law is structurally sound but operationally blind.[53] The priority should be detection infrastructure that makes algorithmic disparate impact observable so that the law can operate.

The most novel intervention is tester standing for algorithmic systems. Civil rights enforcement has a long history of using testers—individuals who apply for housing or work, for example, not to obtain an apartment or a job but to test for discrimination. Courts have recognized tester standing under the Fair Housing Act since *Havens Realty Corp. v. Coleman*.[54] Extending tester

50. Many scholarly proposals for addressing algorithmic discrimination presuppose that discrimination has already been detected. *See, e.g.*, Ajunwa, *supra* note 3, at 624 (proposing mandatory auditing of automated hiring systems but presupposing that discrimination has already been detected); Mayson, *supra* note 1, at 2275 (analyzing corrective measures that require knowledge of bias). This diagnosis-cure mismatch is a central topic of *Incognito Consumer Harm*. *See generally* Kim, *supra* note 5.

51. *See, e.g.*, TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2205-11 (2021) (heightening Article III standing requirements for statutory violations, with potential implications for algorithmic discrimination plaintiffs who suffered no traditional concrete injury). Algorithms can also produce disparities through proxy variables that do not map neatly onto traditional protected classes, further complicating the task of establishing standing and identifying the relevant "practice."

52. *See* Kim, *supra* note 5, at 11 (describing how "algorithmic manipulation or discrimination can affect thousands unaware" and discovery may thus depend on researchers, journalists, whistleblowers, or other unusual detection mechanisms); *id.* at 20-23 (arguing that aggregate litigation inherits the detection gap when victims cannot identify hidden injuries).

53. *See id.* at 38-48 (arguing for a "detection-first paradigm" in consumer protection and proposing regulatory reform designed to surface hidden harm).

54. 455 U.S. 363, 373-74 (1982) (concluding that a tester given false information about housing availability had standing to sue under the Fair Housing Act regardless of intent

footnote continued on next page

methodology to algorithmic systems would mean sending synthetic applicant profiles to detect disparate treatment or impact. Computer science literature has already proven audit methods using fictitious profiles.[55] It requires no change to substantive law, only an extension of testing long applied to human gatekeepers.

Another reform is mandatory outcome reporting. Employers, lenders, and landlords using algorithms in covered decisionmaking contexts should be required to report outcome data disaggregated by protected class. The model is Equal Employment Opportunity reporting, or EEO-1, which requires covered employers to report workforce composition data by race, ethnicity, and sex.[56] Extending this framework to algorithmic decisions would make statistical patterns necessary for disparate impact claims observable without requiring individual victims to detect them.

The third is algorithmic audit rights for regulators. Current transparency frameworks give individuals access to their own data but do not give regulators access to aggregate outcome distributions across protected classes.[57] An audit right requiring disclosure of aggregate results, not source code or trade secrets, would provide the evidentiary foundation that currently does not exist.[58] Each of these interventions is designed not to change what the law says, but to let the law see what it was built to expose.

---

to rent). On the legality of fictitious profiles used to audit algorithmic systems, see Sandvig v. Barr, 451 F. Supp. 3d 73, 76-77, 87-92 (D.D.C. 2020) (finding that fictitious audit profiles do not violate the Computer Fraud and Abuse Act (CFAA) because mere terms-of-service violations do not trigger liability); Van Buren v. United States, 141 S. Ct. 1648, 1658-59, 1662 (2021) (narrowing "exceeds authorized access" liability under the CFAA to a "gates-up-or-down inquiry" with liability only for obtaining "off limits" information—further supporting algorithmic audit legality).

55. *See, e.g.*, Joshua Asplund, Motahare Eslami, Hari Sundaram, Christian Sandvig & Karrie Karahalios, *Auditing Race and Gender Discrimination in Online Housing Markets*, 14 PROC. INT'L AAAI CONF. ON WEB & SOC. MEDIA 24, 24-27, 30-33 (2020) (demonstrating a controlled audit methodology using fictitious profiles to expose race and gender discrimination in online housing platforms).

56. *See* 29 C.F.R. §§ 1602.7-1602.11 (employer reporting obligations under Title VII); U.S. EQUAL EMP. OPPORTUNITY COMM'N, EMPLOYER INFORMATION REPORT (EEO-1 COMPONENT 1) § H (2023), https://perma.cc/UB9H-JBP4 (requiring employers to report workforce demographic data by job category, sex, and race or ethnicity); *cf.* 29 C.F.R. §§ 1602.12-1602.14 (recordkeeping obligations).

57. *Cf.* Kim, *supra* note 5, at 25-29 (arguing that disclosure-based regimes and self-reporting requirements leave enforcement dependent on outside detection); *id.* at 40 (arguing that consumer technology lacks routine inspection capacity, mandatory black box logging, and continuous compliance telemetry for hidden practices).

58. *See, e.g.*, CAL. CODE REGS. tit. 11, §§ 7150, 7155(a)(1), 7157(a)-(e), 7222(b)(2)-(3) (2026) (requiring pre-use risk assessments for certain high-risk processing, but imposing no obligation to routinely file disaggregated outcome data—businesses must only make summary submissions and executive attestations to the California Privacy Protection Agency); Commission Regulation 2024/1689, 2024 O.J. (L 1689) 69, 70, 123 (requiring certain pre-deployment fundamental rights impact assessments without periodic aggregate outcome reporting). These developments narrow but do not close the gap we

footnote continued on next page

Conclusion

The civil rights community has treated the rise of algorithmic decisionmaking as a threat. We, in contrast, suggest that it is an *opportunity*. Every human gatekeeper replaced by an algorithm is a decisionmaker whose disparate outputs become testable, auditable, and legally actionable. This is not a claim that current algorithmic systems are unbiased, but that algorithmic bias, unlike human bias, is the kind that civil rights law is best equipped to expose. If that detection infrastructure is built, the migration of high-stakes decisions from humans to algorithms would represent not a crisis for civil rights enforcement but the largest expansion of its practical capacity in decades.

But detection is only the beginning. Once an algorithm's disparate impact is identified, correcting it may require redesign that is conscious of protected classes, the very kind of classification that the Equal Protection Clause subjects to increasingly unforgiving scrutiny under *Students for Fair Admissions* and *Louisiana v. Callais*.[59] The irony is structural: a detection framework that successfully exposes algorithmic disparate impact may generate pressure for remedies that equal protection constrains.

However, the observation made by this Essay *may*, at least in the hands of a sympathetic enforcer, provide a way of sailing between the Scylla of surreptitious discrimination detectable only by disparate impact and the Charybdis of the prohibition on intentional use of protected classes. As this Essay has shown, algorithmic identification of disparate impact may avoid claims of ideological or self-interested motivation in identifying disparate impact; the very nature of the algorithm limits such a claim. Indeed, given an algorithm designed to pursue neutral goals, it may be possible to innovate an approach to disparate impact that is arguably neutral in a normative sense and more likely to survive constitutional scrutiny. Increasing dependence on

identify. Among states, Illinois makes it a civil rights violation in employment to use AI that "has the effect of subjecting employees to discrimination on the basis of protected classes." 775 ILL. COMP. STAT. ANN. 5/2-102(L)(1) (West 2026).

59. *See* Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll., 143 S. Ct. 2141, 2150 (2023) [hereinafter *SFFA*] ("Eliminating racial discrimination means eliminating all of it. Accordingly, the Court has held that the Equal Protection Clause applies 'without regard to any differences of race, of color, or of nationality'—it is 'universal in [its] application.'") (quoting Yick Wo v. Hopkins, 118 U.S. 356, 369 (1886)); Louisiana v. Callais, 146 S. Ct. 1131, 1146-47, 1152 (citing and discussing *SFFA* as reinforcing the extremely narrow scope of permissible governmental use of race). As *SFFA* highlighted in using Title VI to hold Harvard University to the same standard as the University of North Carolina, 143 S. Ct. at 2156-57 & n.2, statutory frameworks also prohibit private actors in many contexts from engaging in intentional racial classification. *Cf.* Bagenstos, *supra* note 9, at 1130 ("To the extent that [prohibitions on disparate impact] operate to encourage regulated entities to classify individuals based on race, though, disparate impact prohibitions raise 'serious constitutional questions' and must be appropriately cabined.").

algorithms may raise challenges of its own. As algorithmic decisionmaking becomes ubiquitous,[60] it may standardize and scale the criteria encoded in automated systems. And because deploying those systems requires translating context-sensitive and potentially competing values into machine-operationalizable variables and rules, they may compress broader norms of justice and equality into a narrower set of measurable criteria, creating distinct normative tradeoffs.[61] Such concerns follow from the use of algorithms to address inequality, in disparate impact and beyond, and should be of great interest to future scholars.

60. *See, e.g.*, Crootof et al., *supra* note 20, at 432 ("Artificially intelligent algorithms are being integrated into decisionmaking processes at mind-boggling speed and scale."); Kate Crawford & Jason Schultz, *AI Systems as State Actors*, 119 COLUM. L. REV. 1941, 1942 (2019) ("Every month, more algorithmic and predictive technologies are being applied in domains such as healthcare, education, criminal justice, and beyond."); Note, *Machine Rulemaking: Arbitrary and Capricious Review in the Age of AI*, 138 HARV. L. REV. 1821, 1821 (2025) ("[U]se of AI/ML is no longer limited to tech companies or the private sector"). One 2020 report prepared for the Administrative Conference of the United States documented "157 [AI/ML] use cases across 64 agencies." *Id.*; *see also* DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY & MARIANO-FLORENTINO CUÉLLAR, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 16 (2020).

61. *See* Yunsieg P. Kim & Jacob Eisler, *Artificial Sameness* 50-51 (May 2026) (unpublished manuscript) (on file with authors) (arguing that generative AI produces convergence rather than merely directional disparity, and that existing discrimination frameworks are poorly suited to detecting compression).