

Artificially Intelligent Class Actions

Peter N. Salib*

Class actions are supposed to allow plaintiffs to recover for their high-merit, low-dollar claims. But current law leaves many such plaintiffs out in the cold. To be certified, classes seeking damages must show that, at trial, “common” questions (those for which a single answer will help resolve all class members’ claims) will predominate over “individual” ones (those that must be answered separately as to each member). Currently, many putative classes with important claims—mass torts, consumer fraud, employment discrimination, and more—are regarded as uncertifiable for lack of predominance. As a result, even plaintiffs with valid claims in these areas have little or no access to justice. This state of affairs is exacerbated by a line of Supreme Court cases beginning with Wal-Mart Stores, Inc. v. Dukes. There, the Court disapproved of certain statistical methods for answering individual questions and achieving the predominance of common ones.

This Article proposes a first-of-its-kind solution: AI class actions. Advanced machine learning algorithms would be trained to mimic the decisions of a jury in a particular case. Then, those algorithms would expeditiously resolve the case’s individual questions. As a result, common questions would predominate at trial, facilitating certification for innumerable currently uncertifiable classes. This Article lays out the AI class action proposal in detail. It argues that the proposal is feasible today; the necessary elements are predated in both complex litigation and computer science. The Article also argues that AI class actions would survive scrutiny under Wal-Mart, though other statistical methods have not. To demonstrate this, the Article develops a new, comprehensive explanation of the higher-order values animating Wal-Mart and its progeny. It shows that these cases are best understood as approving statistical proof only if it can deliver accurate answers at the level of individual plaintiffs. Machine learning can deliver such accuracy in spades.

INTRODUCTION.....	520
I. WAL-MART’S CRITIQUE OF STATISTICAL PROOF	525
A. The Rise and Fall of Statistical Proof to Facilitate Class Certification	527

* Assistant Professor of Law, The University of Houston Law Center; Associated Faculty Member, The University of Houston Hobby School of Public Affairs. Thanks to Omri Ben-Shahar, Kiel Brennan-Marquez, Anthony Casey, Ryan Copus, Colin Doyle, Maggie Gardner, William H.J. Hubbard, Louis Kaplow, Francesca Procaccini, Nicholas Stephanopoulos, Jay Tidmarsh, Sarah Winsberg, Jonathan Zittrain, and Carly Zubrzycki for their helpful comments. Thanks to Michael Hornzell for excellent research assistance. Finally, thanks to the editors of the *Texas Law Review* for their hard work.

B.	<i>Wal-Mart's Under-Theorized Rule</i>	528
C.	A Theory of <i>Wal-Mart</i>	532
1.	<i>Three Sets of Candidate Values</i>	532
2.	<i>Accuracy Values Dominate the Wal-Mart Analysis.</i>	535
II.	AI CLASS ACTIONS	540
A.	Machine Learning: A Brief Overview	541
B.	How Algorithms Can Facilitate Class Certification	544
C.	Generating Sufficient Training Data.....	548
D.	Litigating Algorithmic Design.....	550
E.	Settlement Dynamics	553
III.	CAN AI CLASS ACTIONS OVERCOME THE <i>WAL-MART</i> PROBLEM?	555
A.	AI Class Actions Are More Than Accurate Enough to Satisfy <i>Wal-Mart</i>	555
B.	Weak AI Class Actions Provide Even More Bang-for-the-Buck	559
IV.	CAN WE TRUST THE MACHINES?	567
A.	Algorithmic Discrimination	567
B.	The “Black Box” Problem	572
	CONCLUSION	575

Introduction

By many accounts, *Wal-Mart Stores, Inc. v. Dukes*¹ killed “trials by statistics”—a family of adjudicatory techniques designed to facilitate class certification.² For a class to be certified, Rule 23 of the Federal Rules of Civil Procedure generally requires that “common” questions of law and fact must “predominate” over individual ones.³ Yet for whole categories of putative classes in a range of important doctrinal areas—mass torts, consumer fraud, employment discrimination, and more—*individual* questions invariably

1. 564 U.S. 338 (2011).

2. See, e.g., Alexandra D. Lahav, *The Case for “Trial by Formula,”* 90 TEXAS L. REV. 571, 574 (2012); Jay Tidmarsh, *Resurrecting Trial by Statistics*, 99 MINN. L. REV. 1459, 1459 (2015). Some scholars argue that *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442 (2016) subsequently resurrected them. See Robert G. Bone, *Tyson Foods and the Future of Statistical Adjudication*, 95 N.C. L. REV. 607, 633–54 (2017) (arguing that “the general principles that best fit and justify what the Court says and does in *Tyson Foods* have normative extension and justify sampling in a wider range of cases”). But this point is disputed. See *infra* note 76 and accompanying text.

3. FED. R. CIV. P. 23(b)(3). While not explicit in their text, Rules 23(b)(1) and (2) likewise impose functional, though perhaps less strict, predominance requirements. Richard A. Nagareda, *Class Certification in the Age of Aggregate Proof*, 84 N.Y.U. L. REV. 97, 131–32 (2009). Note also that, while I focus in this Article on predominance as a hurdle to certification, the issues here also implicate, for example, superiority and manageability. See FED. R. CIV. P. 23(b)(3).

predominate.⁴ In a mass tort, for example, determining that the defendant's asbestos proximately caused one class member's cancer says little about the other members' illnesses.⁵ Instead, resolving the crucial question of causation for the whole class would require innumerable mini-trials. As a result, individual questions would dominate the merits litigation, so certification must be denied.⁶ For reasons like this, untold numbers of meritorious claims in hard-to-certify legal areas simply evaporate. Class certification is unavailable, and without it, the claims cannot be economically litigated.⁷ Potential defendants are thus allowed to violate the law with impunity, and victims are left without a remedy.

The plaintiffs in *Wal-Mart*, a Title VII sex discrimination case, proposed a trial plan—based on statistical adjudication—designed to efficiently dispatch that case's individual questions.⁸ Issues like whether each individual class member would have been promoted, but for unlawful discrimination,

4. See *Wal-Mart*, 564 U.S. at 355, 357–59 (making certification effectively impossible when allegedly discriminatory employment actions are the result of broadly distributed “discretion,” rather than a company “policy”); *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 609–10 (1997); MANUAL FOR COMPLEX LITIGATION (FOURTH) § 22.7 (2004) [hereinafter MANUAL] (“Mass tort personal injury cases are rarely appropriate for class certification for trial.”); *McLaughlin v. Am. Tobacco Co.*, 522 F.3d 215, 224 (2d Cir. 2008).

5. See *Amchem*, 521 U.S. at 609–10. Individual questions have this character—resolving one as to a single class member does not resolve it as to the others. See, e.g., *Wal-Mart*, 564 U.S. at 352. By contrast, common questions are ones for which a “determination of [their] truth or falsity will resolve an issue that is central to the validity of each one of the [class members'] claims in one stroke.” *Id.* at 350.

6. See, e.g., *Wal-Mart*, 564 U.S. at 350.

7. See, e.g., *Amchem*, 521 U.S. at 617 (stating that Rule 23 “does not exclude from certification cases in which individual damages run high,” but making small-value claims economically litigable is a core goal); *In re Rhone-Poulenc Rorer, Inc.*, 51 F.3d 1293, 1299 (7th Cir. 1995) (“In most class actions—and those [] ones in which the rationale for the procedure is most compelling—individual suits are infeasible because the claim of each class member is tiny relative to the expense of litigation.”); Jonathan R. Macey & Geoffrey P. Miller, *The Plaintiffs' Attorney's Role in Class Action and Derivative Litigation: Economic Analysis and Recommendations for Reform*, 58 U. CHI. L. REV. 1, 8–9 (1991) (“The class action procedure partially overcomes [economic] difficulties by providing an effective and inexpensive procedure for joining large numbers of individual plaintiffs.”).

Some such claims may be funneled into alternative forms of collective and pseudo-collective litigation. Multidistrict litigation and bellwether trials are potential avenues for resolving at least some of them. See Alexandra D. Lahav, *Bellwether Trials*, 76 GEO. WASH. L. REV. 576, 581 (2008). But these procedures have serious drawbacks. Not least of them, “parties can . . . ignore [bellwether] results and insist on an individual trial[,]” again raising litigation costs beyond the practicable for small-dollar claims. *Id.* Multidistrict litigation also offers fewer protections for plaintiffs than true class actions, reducing, for example, judicial oversight of settlements. See Elizabeth Chamblee Burch, *Judging Multidistrict Litigation*, 90 N.Y.U. L. REV. 71, 79–80, 83–84 (2015). Moreover, bellwether trials suffer from the same individual accuracy problems discussed herein. Lahav, *supra*, at 581. Thus, the proposal here, for machine learning in class actions, could easily be adapted to improve multidistrict litigations and bellwether trials.

8. 564 U.S. at 343–46.

would be resolved using statistical sampling.⁹ A random representative sample of the class would be chosen, their individual questions resolved, and their individual backpay awards determined.¹⁰ Then, *everyone* in the class would be awarded the sample-average recovery, discounted by the sample-derived probability of a non-meritorious claim.¹¹ With this high-efficiency plan in place, common questions—like whether Wal-Mart maintained a companywide “policy” enabling discrimination—would again predominate in the litigation.¹² Rule 23 would be satisfied, and the class could be certified.

The Supreme Court disapproved.¹³ It pejoratively dubbed the plan “Trial by Formula,” concluding that it violated Rule 23, the Rules Enabling Act, and Due Process.¹⁴ By many accounts, this was the last word on new, creative, and certification-facilitating methods of statistical proof.¹⁵ Such designs were forbidden, and any hope of revitalizing class litigation in hard-to-certify doctrinal areas was lost.¹⁶

This Article proposes a first-of-its-kind method of statistical proof designed to overcome *Wal-Mart*’s critique. In short, cutting-edge machine learning algorithms would efficiently resolve individual questions in class actions, thereby allowing common ones to predominate. Such artificially intelligent (AI) class actions would revolutionize aggregate litigation. By enabling class certification across a broad range of doctrinal areas, they would facilitate the vindication of countless meritorious—but currently unvindicable—claims. At the same time, they would help avoid recovery by plaintiffs who were not owed it. In doing so, AI class actions would serve the interests of plaintiffs, defendants, and society at large.

Why should AI class actions survive scrutiny under *Wal-Mart* when other methods of statistical proof have failed? To answer that question, this Article develops a new, comprehensive account of the higher-order principles animating *Wal-Mart* and its progeny. The cases’ stated logic is confusing—and perhaps confused. But careful analysis reveals that the Court did *not* flatly forbid ambitious, new statistical strategies for efficiently resolving individual questions. Instead, the case law is driven by a demand for *accuracy* in the resolution of claims—not only in the aggregate, but as to individual class members. After *Wal-Mart*, the Supreme Court continued to endorse a small handful of long-established statistical designs that it believed could

9. *Id.* at 348, 366–67.

10. *Id.* at 367.

11. *Id.*

12. *Id.* at 355.

13. *Id.* at 367.

14. *Id.* at 365–67.

15. See, e.g., Lahav, *supra* note 2, at 574 n.9.

16. See, e.g., Tidmarsh, *supra* note 2, at 1459.

produce such individual accuracy.¹⁷ At the same time, it rejected novel approaches that could not adequately minimize individual error.¹⁸ Modern machine learning simply *is* a collection of related algorithmic techniques for producing automated individual decisions with high accuracy. Thus, well-designed algorithms can deliver accuracy to at least the degree that the post-*Wal-Mart* Court has found sufficient—and likely do even better.

This Article proceeds in four Parts. Part I begins by recounting the rise and fall of statistical proof as a mechanism for enabling class certification. It discusses the pre-*Wal-Mart* experiments with sample trials. And it examines *Wal-Mart*'s rejection of that method of statistical proof, along with another method—a regression analysis—advocated by the plaintiffs in that case. In *Wal-Mart* and subsequent cases, the Court explained its acceptances and rejections of statistical proof by reference to preexisting substantive law. If the underlying substantive doctrine—Title VII, securities fraud, antitrust, etc.—permitted some form of statistical proof in an ordinary, individual lawsuit, then it could likewise be used in a class action. But if not, then it could not.

Part I shows why this substantive-law rationale cannot adequately explain the Court's statistical-proof jurisprudence. In short, whenever courts encounter a truly novel method for proving a claim—statistical or otherwise—they must, and do, decide whether it *ought to* be treated as sufficient. Preexisting sources of substantive law rarely provide a definitive answer. Instead, higher-order values must inform the court's decision to make *new* law—by either accepting or rejecting the novel method.

Finally, Part I develops a new account of those higher-order values animating *Wal-Mart* and the cases interpreting it. The Part argues that of the plausible candidate values, those associated with adjudicatory accuracy—fairness, efficient deterrence, corrective justice, and more—best explain the case law.

Part II introduces machine learning as a tool for resolving individual questions and facilitating class certification. First, it briefly summarizes how the newest generation of machine learning algorithms work. These advanced algorithms perform extraordinarily complex categorization and quantification tasks in essentially every sector of commercial and private life. They recognize individual people's faces in photos,¹⁹ power Siri and other

17. See, e.g., Bone, *supra* note 2, at 633.

18. See, e.g., *id.* at 662–64.

19. See generally Navneet Jindal & Vikas Kumar, *Enhanced Face Recognition Algorithm Using PCA with Artificial Neural Networks*, 3 INT'L J. ADVANCED RSCH. COMPUT. SCI. & SOFTWARE ENG'G. 864 (2013).

digital assistants,²⁰ detect credit card fraud,²¹ diagnose diseases,²² determine creditworthiness,²³ and more. They do all of this by learning from “training data”—sets of sample inputs drawn from known cases and pre-labeled with correct determinations that the algorithm should mimic. The algorithms uncover complex relations between the sample inputs and pre-labeled outputs. And they then apply what they have learned to deliver high-accuracy determinations in new cases—those that neither the algorithm nor its human trainers have seen before.

Next, Part II describes how such algorithms would be used in class litigation. As in *Wal-Mart*, a representative sample of the class would be selected and their individual questions tried to the jury.²⁴ But instead of applying the sample’s average answers to every individual class member, the jury’s determinations would serve as labels for training data. Those labels, paired with the evidence that produced them, would be used to train an algorithm to mimic the jury’s decision function. Once trained, the algorithm would analyze the evidence relevant to the remaining, unsampled class members’ claims. Simulating the jury’s decision procedure, it would automatically answer the case’s individual questions as to all of the remaining class members—with high accuracy at the individual level. With individual questions minimized, the litigation could focus primarily on common questions, thus satisfying Rule 23’s predominance requirement.

Would such AI class actions survive scrutiny under *Wal-Mart*? Part III argues that they would. Their algorithmic decisions would be at least as accurate—and likely much more so—than the handful of statistical approaches that the Supreme Court continues to endorse. Part III also explores a weakened version of the AI class action proposal. Under it, algorithmic answers to individual questions would be merely presumptive—challengeable by either party. This design increases accuracy by allowing for

20. See, e.g., SIRI TEAM, *Hey Siri: An On-device DNN-Powered Voice Trigger for Apple’s Personal Assistant*, APPLE (Oct. 2017), <https://machinelearning.apple.com/2017/10/01/hey-siri.html> [<https://perma.cc/UU54-FB2Z>].

21. See generally Sam Maes, Karl Tuyls, Bram Vanschoenwinkel & Bernard Manderick, *Credit Card Fraud Detection Using Bayesian and Neural Networks*, in PROCEEDINGS OF THE FIRST INTERNATIONAL NAISO CONGRESS ON NEURO FUZZY TECHNOLOGIES (2002).

22. See, e.g., Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swelter, Helen M. Blau & Sebastian Thrun, *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 115 (2017).

23. See, e.g., Jochen Kruppa, Alexandra Schwarz, Gerhard Armingier & Andreas Ziegler, *Consumer Credit Risk: Individual Probability Estimates Using Machine Learning*, 40 EXPERT SYS. WITH APPLICATIONS 5125 (2013).

24. Here and throughout, this Article talks about individual questions as being resolved by juries. I mean this as shorthand for the entire process of ordinary litigation. Such questions may also be resolved by judicial rulings on, say, motions for summary judgment. But juries may also be particularly useful in resolving the kinds of fact-intensive questions that can give rise to predominance problems.

error correction. And it simultaneously advances certain non-accuracy-related values championed in the literature on procedural justice—like legitimacy, dignity, and autonomy.

Part IV takes up normative questions. First, it argues that, unlike other legal applications of machine learning, AI class actions are relatively unlikely to generate or entrench invidious discrimination. The procedural and substantive safeguards of ordinary litigation would help to ensure that the training data—the evidence and the jury’s sample determinations—remained unbiased. An algorithm trained on such data would remain, to the same extent, bias-free. Second, the Part argues that, unlike with certain commercial uses of machine learning, we should not worry here about the “black box” nature of algorithmic “reasoning.” Juries’ reasons are similarly unknowable. We therefore employ procedural mechanisms to hold juries accountable and prevent them from relying on forbidden reasons. The algorithmic decisions in AI class actions would be subject to at least those same—and arguably stronger—protections.

I. *Wal-Mart’s* Critique of Statistical Proof

As noted above, *Wal-Mart* is commonly thought to have been the last stand for data-minded wonks set on finding new ways to satisfy Rule 23’s predominance requirement.²⁵ Forbidding “Trial[s] by Formula,”²⁶ the *Wal-Mart* Court abruptly ended years of attempts to open the courts to plaintiffs with high-merit, but low-dollar, claims.

This is not the only way to read *Wal-Mart*. The long-run implications of almost *any* Supreme Court decision are debatable, and *Wal-Mart* is an especially cryptic decision. The majority insisted that its holding turned exclusively on Rule 23(a)(2), which requires only that a putative class share “even a single common question.”²⁷ Finding that the case lacked *any* common questions, the court putatively did not reach the predominance issue at which statistical proof is usually aimed.²⁸ Under this reading, the Court’s objection to statistical proof was merely that it cannot transform an archetypal individual question—like who, in particular, was injured—into a common one.

This Article rejects that narrow, commonality-centered reading of *Wal-Mart* for two reasons. First, it is hard to square with basic Rule 23 jurisprudence. As Justice Ginsburg’s opinion pointed out, the plaintiffs in *Wal-Mart* quite obviously *did* present at least one important common

25. See notes 15–16 and accompanying text.

26. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011).

27. *Id.* at 359 (internal quotation marks and alterations omitted) (quoting Richard A. Nagareda, *The Preexistence Principle and the Structure of the Class Action*, 103 Colum. L. Rev. 149, 176 n.110 (2003)).

28. *Id.*

question: Did Wal-Mart have a company-wide policy or practice governing pay and promotion that could serve as the foundation of a disparate impact claim?²⁹ The majority wrote that this was *not* a common question because the plaintiffs “provide[d] no convincing proof” of such a policy.³⁰

This response, however, confuses the issue of whether plaintiffs *present* a common question with the issue of whether they can *win* on that question.³¹ That distinction is commonplace in class actions. To prove securities fraud, for example, a plaintiff—or class of plaintiffs—must prove that a company’s statement was material to stock buyers.³² And if plaintiffs fail to prove materiality, that is a reason to dismiss the case as meritless.³³ But it is not a reason to deny class certification for lack of commonality.³⁴ Indeed, in such instances, defendants may affirmatively *insist* on certification so that the whole class is bound by a dismissal with prejudice. Likewise, *International Brotherhood of Teamsters v. United States*,³⁵ on which the *Wal-Mart* majority relied,³⁶ was about the evidence necessary to prove the existence of a company-wide practice, and thus carry a Title VII claim.³⁷ *Teamsters* therefore shows why the question *is* common. Proving a company-wide policy—or failing to prove it—“will resolve an issue that is central to the validity of each one of the claims in one stroke.”³⁸

A second reason to reject the narrow reading of *Wal-Mart* is that it risks making things too easy. If *Wal-Mart* held only that statistical proof cannot transform individual questions into common ones, that would be uncontroversial. Such a holding thus does not disfavor statistical proof, as such proof is normally used in class actions. As discussed above, the promise of statistical proof is that it could be used to achieve *predominance* in cases that *do* present common questions.³⁹ And it does so by efficiently answering *individual* questions, not transmuting them into common ones.⁴⁰ Thus, reading *Wal-Mart* narrowly weakens its critique of statistical proof

29. *Id.* at 369–74 (Ginsburg, J., concurring in part and dissenting in part).

30. *Id.* at 359.

31. *See id.* at 369 (Ginsburg, J., concurring in part and dissenting in part). Also note the majority’s suggestion that the plaintiffs might have *won* on the question of a company-wide policy had they adduced more anecdotal evidence about promotions at Wal-Mart. *Id.* at 358. This relegates the Court’s ruling in *Wal-Mart* to mere error correction. It suggests that the exact same case would have come out the other way had it been litigated slightly differently.

32. *Amgen Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 468 (2013).

33. *E.g., id.*

34. *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 457 (2016).

35. 431 U.S. 324 (1977).

36. *Wal-Mart*, 564 U.S. at 358.

37. *Teamsters*, 431 U.S. at 334–35.

38. *Wal-Mart*, 564 U.S. at 350.

39. For an explanation of the *Wal-Mart* plaintiffs’ proposed trial plan, see *supra* notes 9–12 and accompanying text.

40. *Wal-Mart*, 564 U.S. at 368 (Ginsburg, J., concurring in part and dissenting in part).

substantially, making the most important uses of statistical proof trivially distinguishable.

Better then, for purposes of this Article, to treat the *Wal-Mart* majority as Justice Ginsburg did. On her view, the Court jumped ahead and resolved predominance under Rule 23(b)(3), rather than remanding the issue to the district court.⁴¹ This reading places the majority's denial of certification on firmer doctrinal footing: Even if the plaintiffs *did* present common questions about company policies, their proffered statistical proof was not an acceptable way—as discussed below⁴²—to render those questions predominant. Moreover, this reading takes *Wal-Mart* seriously as a critique of statistical proof's core use in class actions—as a facilitator of predominance. If this Article's proposal for AI class actions can withstand scrutiny under *that* version of *Wal-Mart*, it can withstand scrutiny under *any* version.

A. *The Rise and Fall of Statistical Proof to Facilitate Class Certification*

The years leading up to *Wal-Mart* saw numerous experiments with statistical proof designed to achieve predominance and thus facilitate class certification. In particular, a handful of courts approved trial plans calling for the kind of “sampling” design that *Wal-Mart* would eventually quash.⁴³ In each case, a jury resolved the individual questions of a representative sample of the class, and those results were extrapolated to everyone else.⁴⁴ Academics have long favored this methodology because it imposes the correct *aggregate* damages award, despite inaccuracies at the level of individual plaintiffs.⁴⁵

41. *Id.*

42. See *infra* notes 55–65 and accompanying text.

43. See, e.g., *Hilao v. Est. of Marcos*, 103 F.3d 767, 782–84 (9th Cir. 1996) (137 sample claims); *In re Fibreboard Corp.*, 893 F.2d 706, 708–09 (5th Cir. 1990) (thirty sample claims); *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris, Inc.*, 178 F. Supp. 2d 198, 250–52, 254 (E.D.N.Y. 2001) (156 sample claims); *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 653 (E.D. Tex. 1990) (160 sample claims), *aff'd in part, vacated in part*, 151 F.3d 297 (5th Cir. 1998); *Leverence v. PFS Corp.*, 532 N.W.2d 735, 737–39 (Wis. 1995) (thirteen sample claims); *Bell v. Farmers Ins. Exch.*, 9 Cal. Rptr. 3d 544, 551–52 (Cal. Ct. App. 2004) (295 sample claims); *Scottsdale Mem'l Health Sys., Inc. v. Maricopa Cnty.*, 228 P.3d 117, 123 (Ariz. Ct. App. 2010) (1,150 sample claims).

44. See, e.g., *Hilao*, 103 F.3d at 782–84; *In re Fibreboard Corp.*, 893 F.2d at 709; *Phillip Morris*, 178 F. Supp. 2d at 252, 254–55; *Cimino*, 751 F. Supp. at 653; *Leverence*, 532 N.W.2d at 738; *Bell*, 9 Cal. Rptr. 3d at 551; *Scottsdale Mem'l Health Sys.*, 228 P.3d at 123.

45. See, e.g., Robert G. Bone, *Statistical Adjudication: Rights, Justice, and Utility in a World of Process Scarcity*, 46 VAND. L. REV. 561, 568 (1993); Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815, 815 (1992) (“[A]ggregation adds an important layer of process which, when done well, can produce more precise and reliable outcomes.”).

In *Hilao v. Estate of Marcos*,⁴⁶ for example, the parties conducted full trials for the torture, summary execution, and “disappearance” claims of 137 sample class members.⁴⁷ The jury found liability as to 135 of the sampled members and awarded damages to them.⁴⁸ The district court then awarded a lump sum of damages to the remaining class members, multiplying the total number of members by the proportion of the sample with valid claims, and then by the average sample award.⁴⁹ Other trial courts approved similar schemes, with the number of sample adjudications ranging between 13 and 2,000.⁵⁰ Some of these designs were upheld on appeal,⁵¹ but others were reversed on grounds presaging *Wal-Mart*.⁵² And in *Wal-Mart* itself, the Ninth Circuit had approved sampling as a viable method for resolving class members’ individual questions.⁵³ Ultimately, the Supreme Court disagreed, outlawing sampling as a method of class-wide adjudication for individual questions.⁵⁴

B. *Wal-Mart’s Under-Theorized Rule*

Sampling was not the only method of statistical proof that *Wal-Mart* forbade. There, the plaintiffs also presented a regression model designed to show that the discretionary promotion and pay decisions made by thousands of individual managers caused discriminatory harm.⁵⁵ The plaintiffs’ expert conducted a “region by region” analysis, concluding that there were “statistically significant disparities between men and women at Wal-Mart . . . [and] these disparities . . . can be explained only by gender discrimination.”⁵⁶ But the Court rejected that method of proof, too.⁵⁷

Wal-Mart’s explanation for rejecting these methods of proof was not rooted in Rule 23 per se, but rather in the Court’s understanding of substantive law.⁵⁸ As Justice Ginsburg noted, Title VII cases—including

46. 103 F.3d 767 (9th Cir. 1996).

47. *Id.* at 782–84.

48. *Id.* at 784.

49. *Id.* at 783, 784 & n.10.

50. See *supra* note 43 for a list of illustrative cases.

51. See, e.g., *Hilao*, 103 F.3d at 786–87.

52. See, e.g., *Cimino v. Raymark Indus., Inc.*, 151 F.3d 297, 319 (5th Cir. 1998).

53. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 348 (2011).

54. *Id.* at 367.

55. *Id.* at 356.

56. *Id.* (quoting *Dukes v. Wal-Mart Stores, Inc.*, 603 F.3d 571, 604 (9th Cir. 2010) (quoting *Dukes v. Wal-Mart Stores, Inc.*, 22 F.R.D. 137, 154 (N.D. Cal. 2004))).

57. See *id.* at 356–57.

58. See J. Maria Glover, *The Supreme Court’s “Non-Transsubstantive” Class Action*, 165 U. PA. L. REV. 1625 (2017) (“The Rule 23 cases are not so much *at odds* with one another as a matter of procedure—rather, they are perhaps not predominantly *about* procedure.”); Tobias Barrington Wolff, *Managerial Judging and Substantive Law*, 90 WASH. U. L. REV. 1027, 1036 (2013) (“[T]he *Dukes* Court offers an answer to those [Rule 23] questions that sounds entirely in Title VII policy.”).

class actions—can succeed with a showing of disparate impact,⁵⁹ which regularly *does* involve regression analysis.⁶⁰ But in the majority’s view, such evidence was not permitted *in a case like Wal-Mart*. That was because, as the Court held, a disparate impact claim must challenge some “specific employment practice.”⁶¹ The plaintiffs’ collection of millions of discretionary decisions did not qualify as specific enough.⁶² Similarly, as to sampling, the Court found that Wal-Mart had a substantive right to insist on “individualized determinations of each employee’s eligibility for backpay.”⁶³ Sampling designs—although they impose accurate *total* damages awards—contain no mechanism for determining which *individual* plaintiffs’ claims are valid.

Ostensibly, then, the core principle of *Wal-Mart* is that procedural exigency cannot trump substantive law. As the Court wrote, “[T]he Rules Enabling Act forbids interpreting Rule 23 to ‘abridge, enlarge or modify any substantive right’”⁶⁴ Endorsing statistical proof for the sake of class certification, when it would not be allowed in an ordinary individual suit, would constitute such a modification. The Court also opined that using Rule 23 to deny a defendant’s substantive right to demand individualized liability determinations violated the Due Process Clause.⁶⁵ In sum, if some substantive doctrine forbids a method of proof in an individual suit, it is also forbidden in a class action.

Despite ending the recent experiments with sampling, *Wal-Mart* was not the last Supreme Court case to evaluate statistical proof in class actions. In a series of follow-on cases, the Court both upheld and reversed class certification decisions predicated on various—more traditional—statistical

59. *Wal-Mart*, 564 U.S. at 373–74 (Ginsburg, J., concurring in part and dissenting in part).

60. *E.g.*, *McClain v. Lufkin Indus., Inc.*, 519 F.3d 264, 279–80 (5th Cir. 2008).

61. *Id.* at 357.

62. *Id.*

63. *Id.* at 366.

64. *Id.* at 367 (quoting 28 U.S.C. § 2072(b)).

65. *Id.* at 366. Lower federal courts have critiqued sampling using similar arguments rooted in the Seventh Amendment. *See Cimino v. Raymark Indus., Inc.*, 151 F.3d 297, 319 (5th Cir. 1998). The Supreme Court has not taken this approach—either in *Wal-Mart* or afterward. So, I do not treat it at length here. It is an interesting question whether the Seventh Amendment critique of statistical proof is coextensive with the critique grounded in Due Process and the Rules Enabling Act. If not, the facial Seventh Amendment critique of statistical proof generally—and AI class actions specifically—is that it means mathematical models, not juries, resolve factual questions. But this cannot be the whole story, unless *all* of the methods of statistical proof discussed here, including those the Court has repeatedly approved, are unconstitutional. *See infra* notes 68–82. Or perhaps it is good enough to present the statistical results to a jury for wholesale adoption or rejection. Such a step could easily be incorporated into the AI class action proposal. Finally, insofar as Seventh Amendment concerns linger, the weak, presumption-based version of the AI class action proposal should allay them entirely. *See infra* subpart III(B).

methods of proof.⁶⁶ In each case, the Court purported to rely on *Wal-Mart's* “substantive law” principle.⁶⁷

In *Tyson Foods, Inc. v. Bouaphakeo*,⁶⁸ the Court upheld the certification of a class of plaintiffs asserting claims under the Fair Labor Standards Act.⁶⁹ To certify the class, the district court had endorsed the plaintiffs’ statistical plan for proving the amount of time each class member spent donning and doffing protective equipment.⁷⁰ The plaintiffs commissioned a study to determine the average time required for such dressing and undressing.⁷¹ They argued that this average speed should be imputed to every individual class member.⁷² The Supreme Court approved. Citing *Anderson v. Mt. Clemens Pottery Co.*,⁷³ it held that substantive employment law allowed such statistical proof, at least under the right circumstances.⁷⁴ When employers are derelict in their legal duty to keep individual records of employees’ work, they cannot then object to plaintiffs’ reliance on aggregate figures.⁷⁵ Instead, in such cases, “representative evidence . . . [is] a permissible means of” proving hours worked.⁷⁶

Likewise, in both *Amgen Inc. v. Connecticut Retirement Plans & Trust Funds*⁷⁷ and *Halliburton Co. v. Erica P. John Fund, Inc.*,⁷⁸ the Supreme Court reaffirmed the “fraud-on-the-market” method of proving individual reliance in securities fraud cases.⁷⁹ Under it, if one can show *general* market reliance—proved using sophisticated regression models of price

66. See *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 455–56 (2016) (allowing the use of representative time-worked data); *Amgen Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 461 (2013) (allowing the use of the fraud-on-the-market theory to prove reliance in securities fraud cases); *Halliburton Co. v. Erica P. John Fund, Inc.*, 573 U.S. 258, 268 (2014) (same); *Comcast Corp. v. Behrend*, 569 U.S. 27, 37–38 (2013) (disallowing a regression model that failed to match plaintiff’s theories of monopolization).

67. See *Tyson Foods*, 577 U.S. at 458–59; *Amgen*, 568 U.S. at 465–66; *Halliburton*, 573 U.S. at 275; *Comcast*, 569 U.S. at 33, 35.

68. 577 U.S. 442 (2016).

69. *Id.* at 452.

70. *Id.* at 450–52.

71. *Id.* at 450.

72. *Id.* at 454.

73. 328 U.S. 680 (1946).

74. *Tyson Foods*, 577 U.S. at 456.

75. *Id.* at 456–57.

76. *Id.* Robert G. Bone has argued that *Tyson* should be read more broadly, as reversing course on *Wal-Mart's* disapproval of statistical proof. See Bone, *supra* note 2, at 610, 633, 636. I am skeptical of this view. As described above, I read *Tyson* only as reaffirming that average proof may be used when employers have failed to keep the records that would be necessary for employees to adduce individual proof. Bone says that this view is “reasonabl[e].” *Id.* at 634. Insofar as Bone is right, all the better for my AI class action proposal. Moreover, even if the Court threw open the floodgates to statistical proof of all kinds, AI class actions would still have numerous policy advantages over alternative methods. See *infra* Parts II–III.

77. 568 U.S. 455 (2013).

78. 573 U.S. 258 (2014).

79. *Amgen*, 568 U.S. at 462–63; *Halliburton*, 573 U.S. at 283–84.

movements⁸⁰—one may presume that every *individual* class member likewise relied.⁸¹ The Court again endorsed this method of aggregate proof not because Rule 23 itself authorizes it but because substantive securities law does.⁸²

In *Comcast Corp. v. Behrend*,⁸³ the Court rejected the plaintiffs' regression model designed to show class members' pecuniary losses from Comcast's monopolization.⁸⁴ The problem was not that regression analysis is generally forbidden to prove antitrust injury. Such analyses are regularly used in antitrust cases, both collective and individual.⁸⁵ But the *Comcast* Court rejected the regression model because it failed to match the plaintiffs' substantive theory of monopolization.⁸⁶ The model presupposed four theories of injury-causing monopolization, but three of the four were ruled invalid.⁸⁷

In the end, though, the “substantive law” explanation of *Wal-Mart* and its progeny cannot tell the whole story. Certainly, if a preexisting and directly-on-point substantive authority blesses or condemns a method of proof for vindicating a variety of claim, procedural considerations cannot override. But when litigants offer *new* methods of proof—or old methods applied in new ways—there will be ample room to disagree—as the Justices have—about how precedent applies.⁸⁸ And sometimes, the question presented will be whether the precedents were right in the first place.⁸⁹ In all of these situations, some court must bless or condemn, *in the first instance*, the proffered proof.

This is, of course, how the traditional methods of statistical proof that have survived *Wal-Mart* scrutiny first arose. All of them—disparate impact,⁹⁰ “fraud-on-the-market,”⁹¹ the *Mt. Clemens* presumption,⁹² and antitrust

80. *Halliburton*, 573 U.S. at 279–80.

81. *Amgen*, 568 U.S. at 462 (“[I]t is reasonable to presume that most investors—knowing that they have little hope of outperforming the market . . . based solely on their analysis of publicly available information—will rely on the security’s market price as an unbiased assessment of the security’s value in light of all public information.”).

82. *Id.* at 460–61 (resting its endorsement of the “fraud-on-the-market” presumption on § 10(b) of the Securities Exchange Act); *Halliburton*, 573 U.S. at 267–68 (same).

83. 569 U.S. 27 (2013).

84. *Id.* at 34.

85. *See, e.g.*, *Messner v. Northshore Univ. Healthsystem*, 669 F.3d 802, 808 (7th Cir. 2012); MANUAL, *supra* note 4, § 23.1, at 470–71 (“Statistical evidence is routinely introduced and explained by experts in antitrust litigation.”).

86. *See Comcast*, 569 U.S. at 35–38.

87. *Id.* at 36–37.

88. *Comcast*, 569 U.S. at 38–40 (Ginsburg & Breyer, JJ., dissenting); *Wal-Mart v. Dukes*, 564 U.S. 338, 367–68 (2011) (Ginsburg, J., concurring in part and dissenting in part).

89. *Halliburton Co. v. Erica P. John Fund, Inc.*, 573 U.S. 258, 264 (2014).

90. *Griggs v. Duke Power Co.*, 401 U.S. 424, 429–30 (1971).

91. *Basic Inc. v. Levinson*, 485 U.S. 224, 241–42 (1988).

92. *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687–88 (1946).

market reconstruction⁹³—are judicial creations. The creation of such methods is the classic province of the judiciary. Legislators define the elements of a claim; courts determine what constitutes acceptable proof of them.⁹⁴

Thus, the question for a court applying *Wal-Mart* to a new proposed method of statistical proof is not whether preexisting substantive law already permits it. The question must instead be whether substantive law *ought to be* interpreted to permit it.⁹⁵

To resolve that kind of question, courts must appeal to higher-order principles. What values are served by understanding the defendant in *Wal-Mart* as lacking a “specific employment practice”—Title VII’s prerequisite for regression-based proof of discrimination?⁹⁶ What harm would arise from understanding a policy of managerial discretion as constituting such a specific practice?⁹⁷ More broadly, what values are served by endorsing the abbreviated procedures of statistical proof—and thus facilitating class certification—rather than insisting on an individual mini-trial for every individual question?

Wal-Mart and the Supreme Court cases interpreting it supply few answers. Or perhaps it is better to say that their answers are implicit, surfacing only occasionally from beneath the cases’ stated reasoning. But implicit hints are not enough. In order to evaluate new methods of statistical proof, one needs a strong, explicit understanding of the principles underpinning the *Wal-Mart* line of cases.

C. A Theory of Wal-Mart

This subpart supplies the missing theory of the values driving *Wal-Mart* and its progeny.

1. Three Sets of Candidate Values.—Observe that every proposed method of statistical proof is also a proposal for abbreviated adjudicatory process. In class actions, the whole point is to avoid individualized processes—like mini-trials—that would cause individual questions to predominate over common ones. And defendants, to defeat certification, insist that such arduous, member-by-member processes are indispensable.⁹⁸ *Wal-Mart* and its

93. *Comcast*, 569 U.S. at 36; see also *Messner v. Northshore Univ. Healthsystem*, 669 F.3d 802, 808 (7th Cir. 2012).

94. See, e.g., FED. R. CIV. P. 51 (jury instruction); FED. R. CIV. P. 50 (judgment as a matter of law). Except, of course, for common-law claims. Then, courts do both.

95. For more on such judicial acts of hybrid substantive–procedural lawmaking, see Glover, *supra* note 58, at 1636–50; Wolff, *supra* note 58, at 1038–39.

96. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 357 (2011).

97. *Id.* at 372–74 (Ginsburg, J., concurring in part and dissenting in part).

98. See *id.* at 366; see also, e.g., *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 609–10 (1997) (demanding individual process to resolve medical causation in the mass personal injury context).

progeny, then, are intrinsically about what kind of judicial process—and how much—is necessary in a given set of circumstances.

A substantial scholarly literature on procedural justice is devoted to investigating the values that are served—or undermined—when adjudicatory process is increased or decreased.⁹⁹ Lawrence B. Solum has helpfully typologized those multifarious values into three rough categories: accuracy values, process values, and cost values.¹⁰⁰ As is often the case in law, not all of these families of values can be maximized simultaneously.¹⁰¹ Thus, legal rules often embody judgments about which values matter most in a given set of circumstances. And as will be demonstrated, accuracy values drive *Wal-Mart* and subsequent rulings on statistical proof.

Accuracy values are vindicated, unsurprisingly, as adjudications become more accurate.¹⁰² Fulsome judicial process allows more evidence to be entered, the evidence to be scrutinized more closely, and—at least potentially—decision-makers to more often decide correctly.¹⁰³ Accuracy values include, among others: efficient deterrence, fairness, equality of outcome, desert, and corrective justice.¹⁰⁴ When legal decisions are accurate at the level of individual plaintiffs and defendants, both parties are efficiently deterred from engaging in socially costly behavior.¹⁰⁵ It is also more fair, just, and equitable when only those plaintiffs who deserve judgments receive them.¹⁰⁶ Like cases ought to be resolved alike.¹⁰⁷ Finally, principles of corrective justice dictate that individuals who were wrongfully injured should be made whole.¹⁰⁸ But injured parties are not fully restored if they are forced to split their damages with uninjured parties or if they go uncompensated by the hand of rough procedural justice.

99. Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L.J. 455, 474–90 (1986).

100. See Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 305 (2004) (“Accuracy, cost, and participation must all play a role in a theory of procedural justice.”). Other typologies are possible. For example, Robert G. Bone presents a scheme that sorts similar values into an “efficiency metric,” a “rights-based metric,” and a “process-based metric.” Robert G. Bone, *The Process of Making Process: Court Rulemaking, Democratic Legitimacy, and Procedural Efficacy*, 87 GEO. L.J. 887, 919 (1999).

101. Solum, *supra* note 100, at 242–43.

102. See *id.* at 184–85.

103. See *Mathews v. Eldridge*, 424 U.S. 319, 347–48 (1976). This assumes that decision-makers are unbiased and respond rationally to new relevant information.

104. See Bone, *supra* note 45, at 572 (deterrence); *id.* at 604 (corrective justice); *id.* at 576–617 (accuracy values generally); Lahav, *supra* note 2, at 579 (equality of outcome); Solum, *supra* note 100, at 238 (corrective justice); *id.* at 266 (efficient deterrence and fairness).

105. Solum, *supra* note 100, at 266; see also Bone, *supra* note 45, at 572.

106. See Lahav, *supra* note 2, at 594 (“Outcome equality is rooted in the ‘basic principle of justice that like cases should be decided alike.’” (quoting *Martin v. Franklin Cap. Corp.*, 546 U.S. 132, 139 (2005))).

107. *Id.*

108. See Bone, *supra* note 45, at 604.

Process values, by contrast, treat participation in the judicial process as an *inherent* good, regardless of its relation to accuracy.¹⁰⁹ Some scholars contend that all apparent process values are actually reducible to accuracy values.¹¹⁰ But as Solum has argued, truly independent process values do exist.¹¹¹ Chief among these is legitimacy.¹¹² A legal system must allow some quantum of procedural participation in order to be considered legitimate by the polity.¹¹³ This is, in part, because providing access to process demonstrates that the legal system is committed to additional democratic values—like human dignity, equal treatment, and autonomy.¹¹⁴ Of course, all the process in the world will not legitimize a system if that process does nothing to facilitate legally correct outcomes. Thus, while process values may not be *reducible* to accuracy, they do in some sense *depend* on having a relationship to accuracy.¹¹⁵

Cost values are just what they sound like. Adjudicatory resources are not infinite. No government in history has had the resources to stage a full-dress trial before deciding whether to issue each and every driver's license or before approving each and every claim for benefits.¹¹⁶ Thus, governments must allocate resources in a way that “ensure[s] that the systemic costs of adjudication are not excessive in relation to the interests at stake in the proceeding or type of proceeding.”¹¹⁷

109. Solum, *supra* note 100, at 287.

110. See generally Louis Kaplow & Steven Shavell, *Fairness Versus Welfare: Notes on the Pareto Principle, Preferences, and Distributive Justice*, 32 J. LEGAL STUD. 331 (2003); Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

111. See generally Solum, *supra* note 100.

112. *Id.* at 275.

113. *Id.* This is a kind of hybrid normative–empirical argument. It seems likely, as Solum argues, that the citizenry would reject autocratic rule even by a benevolent philosopher king. Government without any opportunity for participation by the governed is illegitimate, regardless of the policies it produces. If some readers disagree, that is fine. As discussed in Part III, *infra*, legitimacy values are not foundational to my arguments here.

114. Solum, *supra* note 100, at 286; see also Bone, *supra* note 45, at 619; Owen M. Fiss, *The Allure of Individualism*, 78 IOWA L. REV. 965, 978 (1993) (“[T]he value of individual participation has an important role to play in the legal process”); Lahav, *supra* note 2, at 573 (“Liberty in civil litigation is summed up as the ‘deep-rooted historic tradition that everyone should have his own day in court.’” (quoting *Ortiz v. Fibreboard Corp.*, 527 U.S. 815, 846 (1999))); Toni M. Massaro, *The Dignity Value of Face-to-Face Confrontations*, 40 U. FLA. L. REV. 863, 902 (1988); Redish & Marshall, *supra* note 99, at 484; see generally JERRY L. MASHAW, *DUE PROCESS IN THE ADMINISTRATIVE STATE* (1985).

115. Solum, *supra* note 100, at 285–86.

116. See *Mathews v. Eldridge*, 424 U.S. 319, 333–34 (1976).

117. Solum, *supra* note 100, at 306.

2. *Accuracy Values Dominate the Wal-Mart Analysis.*—So, which set of values animates the Supreme Court’s statistical-proof rulings in *Wal-Mart* and subsequent cases? The answer must be accuracy values. The Court’s pattern of rulings is readily explained by reference to accuracy values. But analysis of this constellation of decisions using only pure process or cost heuristics renders the pattern incomprehensible. To be sure, those values matter generally to the structure of class actions. But they cannot differentiate between class actions where statistical proof is permitted and those where it is forbidden.

Consider the statistical methods of proof that the post-*Wal-Mart* Court has approved. First, the fraud-on-the-market presumption. At its root, the Court’s endorsement of that presumption rested on its assessment that “most investors—knowing that they have little hope of outperforming the market . . .—will rely on the security’s market price as an unbiased assessment of the security’s value.”¹¹⁸ That is, the Court endorsed the fraud-on-the-market presumption because it believed the presumption to be quite accurate. It was accurate enough, the Court believed, that defendants might, at most, “attempt to pick off the occasional class member here or there through individualized rebuttal.”¹¹⁹

The same can be said for statistical methods of proving antitrust injury. Models designed to show antitrust injury are conceptual siblings of fraud-on-the-market. In both cases, a plaintiff class shows, through statistical evidence, that the defendant’s malfeasance adversely affected *most* market participants.¹²⁰ Having proved that, the plaintiffs are then entitled to assume that *every* market participant was likewise injured.¹²¹ Just as with securities fraud cases, it is plausible that some antitrust plaintiffs escaped injury via idiosyncratic market engagement.¹²² But the same logic—that consumers generally rely on market prices—applies to suggest that those instances are rare.

Likewise, the Court’s *disapprovals* of methods of statistical proof are easily explained in terms of accuracy. The *Wal-Mart* Court objected to the

118. *Halliburton Co. v. Erica P. John Fund, Inc.*, 573 U.S. 258, 273 (2014) (quoting *Amgen Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 462 (2013) (emphasis added)).

119. *Id.* at 276.

120. *See Atl. Richfield Co. v. USA Petroleum Co.*, 495 U.S. 328, 339 n.9 (1990); *Brunswick Corp. v. Pueblo Bowl-O-Mat, Inc.*, 429 U.S. 477, 489 (1977); *Comcast Corp. v. Behrend*, 569 U.S. 27, 31–32 (2013).

121. *See Comcast*, 569 U.S. at 36 (explaining that plaintiff’s expert needed “to establish a ‘but for’ baseline—a figure that would show what the competitive prices would have been if there had been no antitrust violations” and that “[d]amages [for each class member] would then be determined by comparing to that baseline what the actual prices were during the charged period”).

122. Consider a plaintiff with a futures contract to buy the relevant product. If the contract predated the anticompetitive behavior but contemplated a later purchase at the (eventual) super-competitive price, the plaintiff was not injured, despite first appearances.

plaintiffs' regression analysis because it showed sex-based disparities only at the regional level.¹²³ Thus, the Court thought, those disparities could have “be[en] attributable to only a small set of Wal-Mart stores, and [could] not by [themselves] establish . . . uniform, store-by-store disparity.”¹²⁴ Factors like individual managers' advancement philosophies or store-by-store differences in applicant pools made it unlikely, the Court thought, that every plaintiff suffered the regional average harm.¹²⁵ Instead, the Court reasoned that many individual plaintiffs probably suffered no discriminatory harm at all.¹²⁶ In statistics, such erroneous imputation of population-level effects to individuals in the population is called “aggregation bias.”¹²⁷

This accuracy rationale also helps to flesh out the *Wal-Mart* Court's interpretation of substantive Title VII doctrine. Recall that Title VII sometimes allows statistical proof to show a disparate impact, but only if the plaintiff ties that impact to some concrete employment “policy.”¹²⁸ A unified testing procedure for determining hiring and promotion is such a “policy.”¹²⁹ Even some “common *mode* of exercising discretion that pervades the entire company” will do.¹³⁰ But the mere *fact* of discretion—with innumerable, manager-by-manager differences in its exercise—is not concrete enough. Allowing statistical proof, but only if tied to a single concrete causal theory, is an accuracy-improving move. If all employees were subjected to a single *treatment*, then it becomes more plausible that gender disparities were caused by that treatment and not by legally inoperative factors like labor-pool composition.¹³¹ The application of a uniform treatment also makes it less likely that some women suffered extremely harsh injuries, while others suffered none.¹³²

Accuracy values also easily explain *Wal-Mart*'s rejection of sampling. Sampling as a method of proof suffers from obvious aggregation bias. It cannot, by its own terms, distinguish between plaintiffs that had a claim and those who do not—nor can it assess variations in damages.¹³³ Instead, sampling treats every unsampled class member as identical to the sample average, granting everyone the average damages, discounted by the

123. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 388, 356–57 (2011).

124. *Id.* at 357.

125. *Id.*

126. *Id.* at 355–56.

127. *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137, 156–57 (N.D. Cal. 2004); *see also* A. E. LULOFF & P. H. GREENWOOD, DEFINITIONS OF COMMUNITY: AN ILLUSTRATION OF AGGREGATION BIAS 1–2 (1980).

128. *Wal-Mart*, 564 U.S. at 352–55.

129. *Id.* at 353.

130. *Id.* at 356 (emphasis added).

131. *Id.* at 357.

132. *Id.* at 356–57.

133. *See supra* subpart I(B).

probability of an invalid claim.¹³⁴ Every disparity between the sample average and the true fact-of-the-matter thus constitutes an inaccuracy at the individual level.

True, sampling is accurate along *one* dimension—producing the correct total damages. But that is not enough to vindicate all of the values associated with accuracy. As described above, values like fairness and corrective justice require figuring out who *in particular* was harmed and by how much.¹³⁵ And in some contexts, efficiency might require judgments to be accurate as to *both* the defendant and the plaintiff. A sampling scheme that compensates plaintiffs without a valid claim may create a moral hazard by rewarding plaintiffs who fail to take cost–benefit-justified precautions.¹³⁶ These factors can explain why the Court found that Title VII guarantees a right to demand “individualized determinations of each employee’s eligibility for backpay.”¹³⁷

A similar story emerges when one compares *Comcast*—which rejected a statistical market model to prove antitrust injury and damages¹³⁸—with cases that approve such models. In *Comcast*, accuracy was at the core of the dispute between the majority and the dissent. The majority thought that the model was wildly inaccurate because the model presumed the truth of four theories of monopolization, even though three of these theories were previously ruled meritless.¹³⁹ But in the dissenters’ view, if *any* of the theories was viable, then *all* supercompetitive pricing could be attributed to the valid theory of monopolization.¹⁴⁰ In other words, the model’s viability depended on whether it accurately represented the class members’ injuries.

Tyson, too, is best explained as promoting accuracy-related values. The Court in *Tyson* held that if an employer fails to maintain legally mandated timekeeping records, plaintiffs may assume that each class member worked the average amount.¹⁴¹ Normally, treating every employee as average risks individual inaccuracy from aggregation bias—as the sampling design in *Wal-*

134. *See Wal-Mart*, 564 U.S. at 367.

135. *See supra* notes 104–08 and accompanying text.

136. Unlike in, say, the tort context, liability rules need not incentivize Title VII plaintiffs to take precautions against discrimination, *per se*. The onus not to discriminate is rightfully on the employer. Put another way, discrimination is one area where we do not think of parties as imposing Coasian reciprocal harms on one another. *See* R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 2 (1960) (describing reciprocal harms as situations where “avoid[ing] the harm to B would inflict harm on A”). However, observe that Title VII rules denying recovery for individuals who were not actually qualified for a job or pay increase rewards those who *do* invest in such qualification. Thus, even here, individual accuracy improves the regulation of both defendants’ and plaintiffs’ behavior.

137. *Wal-Mart*, 564 U.S. at 366.

138. *Comcast Corp. v. Behrend*, 569 U.S. 27, 34 (2013).

139. *Id.* at 36–38.

140. *Id.* at 47 (Ginsburg & Breyer, JJ., dissenting).

141. *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 456–57 (2016).

Mart showed.¹⁴² But *Tyson* acts as a *prophylactic* against an inaccuracy-promoting moral hazard. Absent *Tyson*, employers would have a strong incentive to eschew the required recordkeeping, which would make class certification impossible, thereby allowing employers to avoid liability for small-dollar underpayment claims. The Court's ruling penalizes such behavior by making class certification easier.¹⁴³ This provides an *ex ante* incentive for employers to keep legally mandated records, and in the long run, those records increase the accuracy of determinations about the validity and size of individual claims. And even in *Tyson* itself, the Court viewed the aggregation error introduced by the prophylactic scheme as modest.¹⁴⁴ The variance in "donning and doffing times" between the fastest and slowest workers was only about "10 minutes a day."¹⁴⁵

Contrast these straightforward, accuracy-motivated explanations of *Wal-Mart* and its progeny with pure process-motivated explanations. Did the proof by regression that *Wal-Mart* rejected threaten process values like legitimacy, autonomy, dignity, or equal treatment? If so, it could only have been by depriving the parties of the opportunity to participate in the resolution of the class members' individual questions. But then what about the many Title VII class actions in which statistical proof of a disparate impact *is* allowed?¹⁴⁶ The only difference between those and *Wal-Mart* was the concreteness of the employer's hiring and promotion policy.¹⁴⁷ It is highly implausible that, say, legitimacy or dignity is threatened by statistical proof in a case without a concrete hiring policy, but not in a case with one.

The same logic applies to every other post-*Wal-Mart* case. Process values cannot plausibly allow for abbreviated statistical proof in securities fraud cases but forbid it in non-securities fraud cases.¹⁴⁸ Nor can process values demand individual mini-trials on antitrust injury in cases with a mismatch between market model and theory of monopolization, but not in cases with a close match.¹⁴⁹ Process values like legitimacy, autonomy, dignity, and equal treatment simply have *nothing to do* with the differences between these cases. They cannot provide an account of why statistical proof is allowed where it is allowed and forbidden where it is forbidden.

Indeed, in many of these cases, the case law seems to get pure process values backward. Surely if anyone was treated without dignity, it was the employees in *Tyson*. Their employer not only underpaid them but also tried

142. *Wal-Mart*, 564 U.S. at 356.

143. *Tyson*, 136 U.S. at 456–57.

144. *Id.* at 457–59.

145. *Id.* at 456.

146. *See supra* notes 59–60 and accompanying text.

147. *See supra* notes 58–61 and accompanying text.

148. *See supra* note 82 and accompanying text.

149. *See supra* notes 85–86 and accompanying text.

to cover its tracks and torpedo their claims by failing to keep proper records. If anyone was entitled on dignitary grounds to be heard during a day in court, it was them. Yet the rule of *Tyson* is the opposite. The statistical proof allowed there ensured that *none* of the disaffected employees would participate in the judicial process about donning and doffing time. In the real world, of course, most of the employees were perfectly happy with this arrangement. In exchange for reduced process, they got justice—damages for wages wrongfully withheld.

Such an exchange—process for access to the courts—is inherent in every class action. Cost values often demand that participation in process be reduced to promote judicial economy.¹⁵⁰ This explains why, for example, comparatively low-stakes, low-complexity administrative claims are often decided with very little process.¹⁵¹ The reality of limited resources does not necessarily lead to a crisis of process values. That is, a legal system does not become illegitimate—nor forsake autonomy, dignity, or equal treatment—simply because it makes reasoned concessions of process to cost.¹⁵²

Class actions themselves are the perfect example of such a reasoned trade-off. In authorizing them, Congress recognized that the ordinary rules of litigation made it economically impractical—for both individuals and the courts¹⁵³—to resolve certain smaller claims.¹⁵⁴ And the reasoned solution—class-wide litigation—*by its very design* contemplates a drastic reduction of participation in process. In an idealized class action, the plaintiffs would present *only* common questions. Then, the claims of everyone in the class would be disposed of simply by adjudicating the merits of the named plaintiffs' claims. The defendant and the named plaintiff would participate fully, but *no one* else would participate at all.¹⁵⁵

Cost values are therefore crucial to understanding class actions generally. But as with process values, cost values are little help in

150. See RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 549 (4th ed. 1992); Solum, *supra* note 100, at 247, 252–56, 305; see generally Kaplow & Shavell, *supra* note 110 (analyzing tradeoffs between accuracy and the cost of adjudication).

151. See *Mathews v. Eldridge*, 424 U.S. 319, 333–34 (1976).

152. See Solum, *supra* note 100, at 307–08.

153. In individual litigation, both private costs, like attorneys' fees, and public costs, like judicial attention, can easily swamp the value of small claims.

154. See *supra* note 7.

155. Scholars have occasionally questioned whether the lack of participation by non-named plaintiffs in class and other mass litigation undermines its legitimacy. See, e.g., Samuel Issacharoff, *Preclusion, Due Process, and the Right to Opt Out of Class Action*, 77 NOTRE DAME L. REV. 1057, 1058–63 (2002); Linda S. Mullenix, *Class Actions, Personal Jurisdiction, and Plaintiffs' Due Process: Implications for Mass Tort Litigation*, 28 U.C. DAVIS L. REV. 871, 916 (1995). But class actions are now a mainstay of American litigation, and they do not appear to have precipitated any crisis in legitimacy. And at any rate, most class actions permit opt-out—and full participation in an individual litigation—for those who reject the sacrifice of process values to cost values. See FED. R. CIV. P. 23(c)(2)(B)(v).

differentiating *which* class actions may be litigated using statistical proof. It is difficult to formulate an argument, for example, that judicial resources should be economized in securities fraud cases, but not for consumer fraud.

Thus, only accuracy values adequately explain the Supreme Court's pattern of decisions in and after *Wal-Mart*. The lesson is that accuracy must be the lodestar for any new proposed methods of statistical proof designed to facilitate class certification. Methods that can provide sufficient accuracy at the level of *individual* class members should survive scrutiny under *Wal-Mart*. And those that cannot will not.

II. AI Class Actions

After *Wal-Mart*, the problem apparently remains: Aside from the handful of contexts where the Court has endorsed high-accuracy statistical proof, class certification in a range of important doctrinal areas appears impossible.¹⁵⁶ Putative classes of people who were exposed to harmful products—like asbestos—cannot be certified when members' claims present disparate questions of exposure and injury.¹⁵⁷ *Wal-Mart* itself shows the persistent difficulty of certification when a class alleges company-wide discrimination by many individual actors.¹⁵⁸ And unlike in securities fraud, class members in garden-variety fraud cases enjoy no presumption by which reliance can be assumed as to every individual class member.¹⁵⁹ Deprived of creative approaches like sampling, there is no obvious way to avoid the obstacle to certification that these individual questions pose. And without class certification, many meritorious claims will go unvindicated.

This Part proposes a new method for efficiently resolving individual questions and achieving class certification: AI class actions. In short, cutting-edge machine learning algorithms can be trained to provide high-accuracy, plaintiff-by-plaintiff answers to individual questions—like medical causation, individual discrimination, or reliance. They can accurately determine whether a *particular* class member—as opposed to the *average* one¹⁶⁰—relied, for example, on a fraudulent misrepresentation. With quick, machine-assisted decision-making available, district courts can confidently

156. *But see supra* note 76.

157. *See Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 609–10 (1997); MANUAL, *supra* note 4, § 22.7 (“Mass tort personal injury cases are rarely appropriate for class certification for trial.”).

158. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 355–56 (2011).

159. *McLaughlin v. Am. Tobacco Co.*, 522 F.3d 215, 224 (2d Cir. 2008) (“[T]he market for consumer goods, however, is anything but efficient . . .”); *Sikes v. Teleline, Inc.*, 281 F.3d 1350, 1364 (11th Cir. 2002) (“[E]ach individual plaintiff is the only person with information about the content of the advertisement upon which he relied.”).

160. Or, as with regression analysis, the average member bearing a set of regressor characteristics (where “average” means falling on the least-squares line).

rule that, come trial, common questions will predominate over individual ones. Thus, the otherwise uncertifiable classes are rendered viable.

A. *Machine Learning: A Brief Overview*

“Machine learning” is a broad term, encompassing a number of related algorithms that automatically improve themselves in response to data.¹⁶¹ The term is capacious enough to cover, for example, classic linear regression, which is often considered one of the simplest machine learning techniques.¹⁶² This Article, however, uses “machine learning”¹⁶³ to refer to the more complex, advanced learning algorithms that have recently begun to permeate nearly every corner of society.

Advanced machine learning algorithms power Siri and Alexa, converting sound waves into text and natural language into commands a computer can execute.¹⁶⁴ They detect unauthorized credit card use and identity theft.¹⁶⁵ They evaluate loan applications for creditworthiness.¹⁶⁶ They can recognize objects or even individual faces in photos, enabling services like Google’s image search and Facebook’s automatic tagging.¹⁶⁷ And they are even used in certain courthouses to help determine who should be released to await trial and who should remain jailed.¹⁶⁸ This is just a small sample of where machine learning is—and will soon be—used.

This new¹⁶⁹ generation of learning machines includes a number of specific and technically distinct algorithms—with names like “neural

161. TREVOR HASTIE, ROBERT TIBSHIRANI & JEROME FRIEDMAN, *THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION* 9 (2d ed. 2009).

162. *Id.* at 3–4.

163. And terms like “AIs” and “algorithms.”

164. *See, e.g.*, SIRI TEAM, *supra* note 20 (explaining how the “Hey Siri” feature recognizes speech by turning it into waveforms).

165. *See, e.g.*, Maes, *supra* note 21; A. Martin, Na. Ba. Anutthamaa, M. Sathyavathy, Marie Manjari Saint Francois & Prasanna Venkatesan, *A Framework for Predicting Phishing Websites Using Neural Networks*, 8 INT’L J. COMPUT. SCI. ISSUES 330, 330, 333–34 (2011), <https://arxiv.org/ftp/arxiv/papers/1109/1109.1074.pdf> [<https://perma.cc/9SVT-YEGD>].

166. Kruppa, *supra* note 23, at 5125, 5127, 5131.

167. Jianlong Fu & Yong Rui, *Advances in Deep Learning Approaches for Image Tagging*, APSIPA TRANSACTIONS ON SIGNAL & INFO. PROCESSING, Oct. 4, 2017, at 1, 3, 9; *see also* Aayushi Mangal, Himanshu Malik & Garima Aggarwal, *An Efficient Convolutional Neural Network Approach for Facial Recognition*, in 2020 10TH INTERNATIONAL CONFERENCE ON CLOUD COMPUTING, DATA SCIENCE & ENGINEERING 817, 817 (2020); Anna Bosch, Andrew Zisserman & Xavier Munoz, *Image Classification Using Random Forests and Ferns*, in 2007 IEEE 11TH INTERNATIONAL CONFERENCE ON COMPUTER VISION (2007).

168. Julia Angwin, Jeff Larsen, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<https://perma.cc/YPL4-9VRR>].

169. Many of these algorithms have been understood in theory for decades. But only with recent advances in data collection and computing power have they become viable options for performing practical tasks. *See* HASTIE, *supra* note 161, at 102 (describing early perceptron theory).

network,” “random forest,” “gradient booster,” and more.¹⁷⁰ But despite their differences, these cutting-edge techniques share important conceptual similarities. At their cores, they are methods for automatically uncovering fantastically complex correlations in data and then making high-accuracy determinations about what the data represent.

These algorithms optimize themselves, learning from sample data.¹⁷¹ Training generally takes the following form: First, a set of training data is collected. This data consists of example decisions or relations that the user wishes the machine to emulate. It includes both the input features to be analyzed and the correct decision in each instance.¹⁷² For example, to train an algorithm to recognize photos of dogs, the training data would comprise a set of photos pre-labelled as either featuring a dog or not. A loan-issuance algorithm might be trained using a set of known loan applications pre-labelled as to their actual creditworthiness.¹⁷³

Using semi-random processes, the algorithms generate a set of complicated relationships between inputs and outputs, thus producing tentative decisions.¹⁷⁴ They then compare those tentative decisions against the provided correct ones. Using an “error function,” they measure how good, on the whole, the decisions were.¹⁷⁵ Then, using an “optimization function,” the machines update their correlations between inputs and outputs, and they try again.¹⁷⁶ Training stops when error is minimized—that is, when the machine’s outputs match the provided correct ones as closely as possible.¹⁷⁷ The trained algorithm is then tested against *another* training dataset—one that was “held out” and not used in the initial training process.¹⁷⁸ If the

170. *See id.* at 337, 352, 389, 587.

171. Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 115 (2018). Note that the description here is streamlined for clarity.

172. *Id.* at 134–37.

173. *See* Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459, 463–64 (2019).

174. *See* Xu Sun, Xuancheng Ren, Shuming Ma & Huofeng Wang, *meProp: Sparsified Back Propagation for Accelerated Deep Learning with Reduced Overfitting*, in PROCEEDINGS OF THE 34TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING (2017) (describing forward and back propagation). I talk here and throughout about algorithmic “decisions,” implying categorical outputs. Advanced machine learning algorithms can also make quantitative decisions, like how *much* money a plaintiff is most likely owed. *See* HASTIE, *supra* note 161, at 2, 4.

175. *See* A.N. Gorban, E.M. Mirkes & A. Zinovyev, *Piece-Wise Quadratic Approximations of Arbitrary Error Functions for Fast and Robust Machine Learning*, NEURAL NETWORKS, Dec. 2016, at 28, 28–29.

176. *See* Sun, *supra* note 174 (describing backpropagation).

177. HASTIE, *supra* note 161, at 398. This is a slight simplification. Training does not necessarily proceed until error is absolutely minimized. This can result in overfitting. Instead, data scientists employ techniques for balancing error minimization against generalizability. *Id.*

178. Kleinberg, *supra* note 171, at 136–37. This simplifies things slightly. In reality, training data is often partitioned into subsets, and the algorithm is trained using multiple combinations of

algorithm outputs the correct decisions for that set, this is evidence that its decision function is not “overfitted” to the training data.¹⁷⁹ That is, the algorithm is likely to make accurate determinations about *new* cases—those for which the correct answers are not already known.¹⁸⁰

Advanced machine learning differs from, for example, classic linear regression in two important ways. First, advanced machine learning algorithms’ decision functions can be radically nonlinear.¹⁸¹ That is, they do not attempt to fit the relations of inputs and outputs to any pre-defined, easy-to-interpret type of function.¹⁸² As a result, they have the ability to reduce the errors in their results much more than is possible with older-fashioned techniques.¹⁸³ Second, these algorithms can automate feature selection. Classic linear regressions depend on designers to make good guesses about which independent variables will best predict the dependent ones.¹⁸⁴ Advanced algorithms can analyze many potential independent variables, determine which—individually or in combination—are the best predictors, and ignore the others.¹⁸⁵

The arguments of this Article transcend the technical differences between different advanced-learning machines. Instead, the Article speaks generically of “algorithms,” referring to the whole family of methods that operate on the above-described principles. Particular cases may often raise disagreements about which of the available algorithms ought to be used and who should decide. But so long as there are reasonable methods for resolving such disagreements—and there are¹⁸⁶—any algorithm described above will do. Algorithmic ecumenicism lends the arguments here some quantum of future-proofness. Technical advances in machine learning are proceeding at a blinding pace. But so long as those advances preserve most of the basic cross-algorithmic features described above, the insights of this Article remain valid.

training data and hold-out data. This technique is called k-fold cross-validation. Sanjay Yadav & Sanyam Shukla, *Analysis of K-fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification*, in 2016 6TH IEEE INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING 78, 79 (2016).

179. HASTIE, *supra* note 161, at 398.

180. Yadav & Shukla, *supra* note 178, at 79; *see also* HASTIE, *supra* note 161, at 398.

181. *See* Gillis & Spiess, *supra* note 173, at 466–67, 469.

182. *See id.* at 469.

183. *See* HASTIE, *supra* note 161, at 11, 14–15 (showing error reductions achieved in simple model by moving from least-squares regression to k-nearest neighbors).

184. *Id.* at 11, 16.

185. V. Umayaparvathi & K. Iyakutti, *Automated Feature Selection and Churn Prediction Using Deep Learning Models*, INT’L RSCH. J. ENG’G & TECH., Mar. 2017, at 1846, 1846–87; Kleinberg, *supra* note 171, at 135–36.

186. *See infra* subpart II(D).

B. How Algorithms Can Facilitate Class Certification

Now we come to the heart of the proposal. Recall that the key to achieving the predominance of common questions is resolving individual ones without having to resort to resource-intensive procedures like mini-trials.¹⁸⁷ Advanced machine learning algorithms could be trained to efficiently resolve such questions and thus satisfy Rule 23's predominance requirement.

How would this be done? First and most important is the issue of the training data. Consider the individual medical-causation questions that preclude class certification in mass tort suits.¹⁸⁸ To train a cancer-causation algorithm, one needs a set of real claimants, each labeled with the correct answer as to whether they can show causation. Thus, the training data could comprise a sample set of determinations about causation made by the actual jury in the case.¹⁸⁹

This scheme should sound familiar. *Wal-Mart*-style sampling designs also relied on jury determinations of a subset of class members' individual questions.¹⁹⁰ AI class actions would similarly require the random selection of a statistically significant sample¹⁹¹ of class members. Those class members' claims would then be tried. Exactly what such a trial would entail would vary—as trials always do—with the theories presented and the evidence proffered. One can imagine that medical-causation claims might be largely decided on the basis of documentary evidence: work history, medical records, demographic information, and the like. Perhaps live testimony would be required from some of the sample class members. If the chemical exposure happened on the job, the jury might wish to know what exactly each job description entailed. At this stage, the judge should err on the side of admitting relevant information about the sample class members into evidence. This serves more than the normal evidentiary purpose of giving the jury a robust picture of the claimants. As discussed below, it also helps to produce a wealth of machine-digestible information for training the algorithm.¹⁹²

187. See *supra* notes 5, 98 and accompanying text; see also *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 622–24 (1997).

188. See *Amchem*, 521 U.S. at 624.

189. The sample set could also comprise decisions made by the judge in a bench trial, or a judge pursuant to a Federal Rule allowing judgment as a matter of law. See, e.g., FED. R. CIV. P. 56 (enabling motions for summary judgment).

190. See *supra* note 10 and accompanying text.

191. The sample size must be *at least* large enough to achieve statistical significance. Sometimes, larger samples may be required to adequately train an algorithm. This issue—along with the potential feasibility concerns it raises—is discussed in subpart II(C), *infra*.

192. See *infra* notes 195–97 and accompanying text. Judges should of course still be sensitive to admissibility considerations like confusion and prejudice. These factors affect algorithms less—and differently—than humans. See *infra* subparts II(D) and IV(A). Thus, judges might vary

At the end of the sample trial process, the litigants would possess a set of sample class members, a variety of evidence relating to their causation questions, and the jury's actual determination of causation as to each. This constitutes the training data. Then, training commences. A subset of the sample cases is set aside for later use in validation.¹⁹³ The remainder of them are fed to the algorithm.¹⁹⁴

To accurately mimic human decisions, algorithms do not need to be fed all of the information that the human jury saw. Likely, many features of the training data have little predictive power and will be weeded out in feature selection. But algorithms can also perform well without being fed certain features *with* significant predictive power. As Talia Gillis and Jann Spiess have recently shown, “when complex, highly nonlinear prediction functions are used, . . . one [omitted] input variable can be reconstructed jointly from the other input variables.”¹⁹⁵ Sometimes this is a bad thing, as when algorithms are able to reproduce humans' racially biased decisions despite not being fed racial data.¹⁹⁶ The algorithms are essentially able to *infer* people's race and make decisions based on race by uncovering complex racial correlates within the data the algorithms do have.¹⁹⁷

But other times—as with AI class actions—such inferences can be a very good thing. Algorithms designed to issue loans or screen job candidates would add little efficiency if, to make their decisions, they depended on inputs derived from lengthy in-person interviews.¹⁹⁸ Luckily, they do not.¹⁹⁹ Instead, they consume easily collected information—resumé, job history, income, credit score, etc.—and use just that to mimic interview-informed human decisions.²⁰⁰

admissibility decisions based on whether evidence would be reviewed by a jury, an algorithm, or both. These two evidentiary sets need not be the same. *See infra* notes 195–97 and accompanying text.

193. *See supra* notes 174–80 and accompanying text. This again simplifies things. In reality, some form of k-fold cross-validation might be used, perhaps with an *additional* set of training data completely held out from the training process. *See supra* note 178.

194. *See supra* notes 174–80 and accompanying text.

195. Gillis & Spiess, *supra* note 173, at 469.

196. *See id.*

197. *See id.* (“[I]f there are other variables that are correlated with race, then predictions may strongly vary by race even when race is excluded . . .”).

198. *See id.* at 467–68.

199. *See id.* at 467.

200. *See id.* at 467–68. Again, such mimicry is only as good as the human decisions being mimicked. If those decisions are, for example, racially biased, the algorithm will be similarly biased. But if the human decisions are, say, unbiased jury determinations, the algorithm will also be unbiased.

For these same reasons, it would make little sense to train the causation algorithm to rely on, for example, video of live questioning.²⁰¹ The whole point of AI class actions is to efficiently resolve individual questions across the entire class. The trial plan therefore cannot call for every individual class member to be deposed so that the algorithm can interpret their depositions. Instead, the algorithm should be trained using only data that can be procured with relative ease as to each of the class members. In the case of medical causation, this would likely include things like medical and employment records. But it could also include, for example, questionnaire-based testimony—made under oath—from plaintiffs, which is already used in class litigation.²⁰²

So, the algorithm would be trained using the jury’s sample decisions paired with the subset of easily collected and machine-digestible evidence relating to those decisions. As described above, the algorithm would generate a set of fantastically complex and semi-random relationships between the evidentiary inputs and use them to guess at some outputs.²⁰³ It would measure errors in its first set of guesses.²⁰⁴ Then, using its optimization function, it would update the complex set of relations between inputs and outputs and guess again.²⁰⁵ This process continues until error is minimized, signaling that the algorithm has generated a decision that matches the jury’s as closely as possible.²⁰⁶

Finally, the trained algorithm would be tested against the hold-out set of training data. The causation evidence of those held-out members would be fed to the machine, and the machine would output determinations.²⁰⁷ If the

201. Training the algorithm with video may even be computationally practicable. AIs can be taught to interpret video but doing so is computationally expensive. *See generally* Daniel Kang, John Emmons, Firas Abuzaid, Peter Bailis & Matei Zaharia, *Noscope: Optimizing Neural Network Queries over Video at Scale*, 10 PROC. VLDB ENDOWMENT 1586 (2017).

202. *See* *Vinole v. Countrywide Home Loans, Inc.*, 571 F.3d 935, 947 (9th Cir. 2009) (holding that “the decision to use” tools “such as questionnaires” is “within the discretion of the district court”); *see also In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 227 F.R.D. 553, 621 (W.D. Wash. 2004) (using questionnaires); *Millsap v. McDonnell Douglas Corp.*, No. 94–CV–633–H(M), 2003 WL 21277124, at *2 (N.D. Okla. May 28, 2003) (same); *Schwartz v. Celestial Seasonings, Inc.*, 185 F.R.D. 313, 316 (D. Colo. 1999) (same); *Krueger v. N.Y. Tel. Co.*, 163 F.R.D. 446, 451 (S.D.N.Y. 1995) (same).

Machine learning could also be used to narrow the set of information that would eventually be collected about the non-sample class members. Once trained, advanced machine learning algorithms can indicate which input features generate the most predictive power as to outputs. Jeff Heaton, *Feature Importance in Supervised Training*, PREDICTIVE ANALYTICS & FUTURISM, Apr. 2018, at 22, 24. Thus, while discovery as to the sample class members’ claims might be broad, for the rest of the members, it could be limited to only the categories the algorithm determined necessary.

203. *See supra* subpart II(A).

204. *See supra* notes 171–75 and accompanying text.

205. *See supra* note 176 and accompanying text.

206. *See supra* note 177 and accompanying text.

207. *See supra* note 178 and accompanying text.

algorithm accurately resolved *those* causation questions—which it had never before seen—that would be good evidence that it would perform well across the *entire* class.²⁰⁸ That is, the algorithm would be able to simulate—with high accuracy—the jury’s determination about whether *each individual* class member could show medical causation.

This sounds like speculative fiction, but it is not. Algorithms already exist that can accurately resolve thorny legal questions. For example, a team of professors at the University of Toronto Faculty of Law have successfully used machine learning to predict tax law decisions.²⁰⁹ Their “Blue J” algorithm can determine, among other issues, whether a given individual is rightly classified as an employee or an independent contractor.²¹⁰ This is exactly the kind of commonplace but fact-intensive legal question that algorithms would be asked to answer in AI class actions.

How does Blue J perform? “Exceptionally well.”²¹¹ Given just twenty-one facts about a given worker, the algorithm can correctly classify well over ninety percent of workers as either employees or contractors.²¹² This high performance persists when Blue J is turned loose on “a variety of different questions in tax law.”²¹³ Other academics have applied algorithms to much harder legal problems—for example, predicting appellate and apex court decisions.²¹⁴

Like Blue J, a trained algorithm in an AI class action would efficiently resolve individual questions as to each individual class member. Assume for now that these algorithmic decisions would, like a jury determination, be the

208. As above, this is a slight simplification. In actuality, k-fold cross-validation—whereby multiple subsets of training data take turns validating trained algorithms—would produce the best results. *See supra* note 178.

209. Benjamin Alarie, Anthony Niblett & Albert H. Yoon, *Using Machine Learning to Predict Outcomes in Tax Law*, 58 CAN. BUS. L.J. 231, 233 (2016).

210. *Id.* at 240–43.

211. *Id.* at 242.

212. *Id.* at 242–43.

213. *Id.* at 242.

214. *See generally* Daniel Martin Katz, Michael J. Bommarito II & Josh Blackman, *Predicting the Behavior of the Supreme Court of the United States: A General Approach*, PLOS ONE (2017), <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0174698&type=printable> [<https://perma.cc/ZUY5-86N9>]; Bernhard Walil, Georg Bonczek, Elena Scepankova, Jörg Landthaler & Florian Matthes, *Predicting the Outcome of Appeal Decisions in Germany’s Tax Law*, in ELECTRONIC PARTICIPATION 9TH IFIP WG 8.5 INTERNATIONAL CONFERENCE, ePART 2017 ST. PETERSBURG, RUSSIA, SEPTEMBER 4–7, 2017, PROCEEDINGS 89 (Peter Parycek, Yannis Charalabidis, Andrei V. Chugunov, Panos Panagiotopoulos, Theresa A. Pardo, Øystein Sæbø & Efthimios Tambouris eds., 2017). The predictions here are less accurate, but that is almost certainly because the questions are much harder. Questions that get to the Supreme Court are likely to be the hardest and least settled. *See generally* Ryan W. Copus, *Statistical Precedent: Allocating Judicial Attention*, 73 VAND. L. REV. 605 (2020) (predicting how every combination of Ninth Circuit panel would rule on each case between 2011 and 2012).

final word on each individual issue.²¹⁵ This straightforwardly illustrates how machine learning would facilitate class certification where it is currently considered impossible. Issues—like medical causation, reliance, or intent—which presently require individual adjudications for every class member would instead be resolved with extreme efficiency. Only a small sample of class members’ individual questions would be tried to the jury. Everyone else’s questions would be permanently resolved by a computer in essentially no time at all. And this would be the end of the story. With individual questions thus minimized at trial, district courts could rule confidently that common questions predominated, satisfying Rule 23.

C. *Generating Sufficient Training Data*

Some readers may wonder whether conducting sample trials would really be a feasible way to produce enough data to adequately train an algorithm. Some machine learning questions are harder than others. Thus, when applying an algorithm to a totally novel problem, it is often difficult to know in advance how much training data will be needed.²¹⁶ And some problems are so difficult that *no* amount of training data currently suffices.²¹⁷ If the individual questions presented in AI class actions required tens of thousands or millions of instances of training data, then the proposal would be dead on arrival.

There are good reasons to think that, in AI class actions, reasonable amounts of data would suffice. First, the kinds of individual questions to be resolved in a class action are not unprecedented in the world of machine learning. As already noted, the Blue J algorithm can accurately apply settled legal tests to complex factual situations.²¹⁸ Common individual questions in class actions—medical causation, reliance, etc.—work the same way,

215. They could instead be treated as rebuttable presumptions—a possibility explored below. *See infra* subpart III(B).

216. *See* Andrew H. Sung, Bernardete Ribeiro & Qingzhong Liu, *Sampling and Evaluating the Big Data for Knowledge Discovery*, in PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON INTERNET OF THINGS AND BIG DATA 378, 378–80 (Muthu Ramachandran, Gary Wills, Robert Walters, Victor Méndez Muñoz & Victor Chang eds., 2016) (sketching a proof that “the problem of selecting an optimal sample for model building using learning machines is intractable”).

217. *See, e.g.*, Pedram Ataee, *If You Consider Using AI in Your Business, Read This.*, MEDIUM (May 20, 2020), <https://medium.com/swlh/if-you-consider-using-ai-in-your-business-read-this-5e666e6eca23> [<https://perma.cc/T8TA-K3EC>]; Daniel Faggella, *The Self-Driving Car Timeline – Predictions from the Top 11 Global Automakers*, EMERJ (Mar. 14, 2020), <https://emerj.com/ai-adoption-timelines/self-driving-car-timeline-themselves-top-11-automakers/> [<https://perma.cc/E3B7-VLD5>]; Lauren Kunze, *We’re Thinking About the Turing Test All Wrong*, QUARTZ (Dec. 7, 2018), <https://qz.com/1487101/the-turing-test-shows-how-chatbots-ultimate-goal-isnt-intelligence-its-language/> [<https://perma.cc/LX97-PS9H>].

218. *See* Alarie, *supra* note 209, at 242–43.

suggesting that similar amounts of training data might be required.²¹⁹ Blue J was trained using the entire corpus of relevant Tax Court of Canada decisions—numbering 600.²²⁰ This is well within the range of the feasible, as far as sample adjudications go.²²¹ Recall that pre-*Wal-Mart* courts approved as practicable as many as 1,150 hearing-based sample adjudications of individual questions.²²² That number jumps to 2,000 for adjudication based on documentary evidence.²²³

Even smaller datasets might often be sufficient. Blue J used all of the readily available data, but it is possible that it could have done as well with less. Among data science professionals, an oft-cited rule of thumb is that one needs roughly ten training instances for every “feature”—every category of input evidence.²²⁴ Blue J makes accurate contractor/employee assessments based on just twenty-one features about each worker.²²⁵ This suggests a minimum number of training cases in the low 200s. Empirical studies for other machine learning applications—classifying medical images, spotting ads—likewise show minimum datasets around the 200 range.²²⁶

More importantly, if AI class actions become more prevalent, there are strong reasons to expect more and more individual questions to be resolvable with less and less data. The problem of “small data”—training accurate algorithms using modest datasets—is on the bleeding edge of machine learning research. New techniques are being invented all the time. Of

219. See, e.g., Jason Brownlee, *How Much Training Data Is Required for Machine Learning?*, MACHINE LEARNING MASTERY (July 24, 2017), <https://machinelearningmastery.com/much-training-data-required-machine-learning/> [<https://perma.cc/7DEF-ZFL4>] (suggesting that to determine the amount of data that may be required, one could “look at studies on problems similar to yours as an estimate”).

220. Alarie, *supra* note 209, at 241.

221. Recall that *Wal-Mart* took no issue with the approved trial plans’ feasibility. The *Wal-Mart* problem, as described above, is about whether proposed trial plans violate some substantive legal commitment. See *supra* Part I.

222. *Scottsdale Mem’l Health Sys., Inc. v. Maricopa Cnty.*, 228 P.3d 117, 123 (Ariz. Ct. App. 2010) (approving 1,150 sample claims).

223. *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris, Inc.*, 178 F. Supp. 2d 198, 226, 250 (E.D.N.Y. 2001) (approving 2,000 survey-based sample claims).

224. See, e.g., Brownlee, *supra* note 219 (describing “statistical heuristic methods” for estimating training data needs).

225. Alarie, *supra* note 209, at 242–43.

226. See, e.g., Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy & Synho Do, *How Much Data Is Needed to Train a Medical Image Deep Learning System to Achieve Necessary High Accuracy?* 6 (Jan. 7, 2016), <https://arxiv.org/pdf/1511.06348.pdf> [<https://perma.cc/UC3F-Q7UW>]; José Silva, Bernardete Ribeiro & Andrew H. Sung, *Finding the Critical Sampling of Big Datasets*, in ACM INTERNATIONAL CONFERENCE ON COMPUTING FRONTIERS 2017, at 355, 358 (2017); see also Ammara Masood & Adel Ali Al-Jumaily, *Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms*, INT’L J. BIOMEDICAL IMAGING, Dec. 23, 2013, at 1, 13 tbl.5 (collecting publications between 1993 and 2012 of machine-learning-aided cancer diagnoses with training datasets range from 22 to 5363 instances).

particular relevance here is the technique of “transfer learning.”²²⁷ In transfer learning, an algorithm is trained to answer a question for which lots of data is available.²²⁸ Then, using a small dataset, it is partially retrained to answer a *similar* question.²²⁹ Transfer learning can reduce the amount of data required to answer the latter question by orders of magnitude.²³⁰

It is easy to imagine, in a world of widespread AI class actions, that a market would develop for transfer-learning-ready algorithms. Such a market already exists for third-party expert models in, for example, antitrust and securities cases.²³¹ Private entities—like litigation consulting firms—could similarly train general-purpose algorithms to answer questions of medical causation, reliance, or discrimination. They could rely on vast troves of existing data, both in case law and privately held—for example, by insurance companies that pay tort settlements.²³² Then, the task of adapting a *general-purpose* medical-causation algorithm to mimic a *particular jury’s* causation decisions would require very little new data.

D. *Litigating Algorithmic Design*

Even given sufficient training data, machine learning algorithms are not self-constructing. Talented human data scientists craft them and, in that process, make many granular choices about their design. If AI class actions are adopted, these choices will surely give rise to disputes between litigants.

David Lehr and Paul Ohm have helpfully catalogued the steps to construct a fully functioning machine learning algorithm.²³³ Not all of these are likely to raise trouble in class litigation. For example, in AI class actions, there will be no disputes about the “problem definition”—the question the algorithm should be trained to answer.²³⁴ Here, unlike in other machine

227. See generally Sinno Jialin Pan & Qiang Yang, *A Survey on Transfer Learning*, 22 IEEE TRANSACTIONS ON KNOWLEDGE & DATA ENG’G 1345 (2010).

228. See, e.g., Esteva, *supra* note 22, at 115 (describing an algorithm pre-trained to classify general images using millions of examples).

229. See *id.* (using a general image classifier trained to classify images of lesions).

230. See *id.* (training data for target question roughly one order of magnitude smaller than pre-training data).

231. See, e.g., *Practice Areas*, COMPASS LEXECON, <https://www.compasslexecon.com/practice-areas/> [<https://perma.cc/6LPU-VZA5>].

232. Insurance companies—which often pay companies’ damages awards—possess massive troves of data cataloging the features and results of real-life court cases already. Lahav, *supra* note 2, at 579. Although that data is currently treated as proprietary—used to set competitive rates—AI class actions would create a financial incentive to share it.

233. See generally David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653 (2017).

234. See *id.* at 672–73.

learning contexts, the answer is clear: The machine will mimic the jury's answers to individual questions in the sample set.²³⁵

Disputes over algorithmic design choice are likely to fall into two categories: data collection and preparation, and model training.²³⁶ As to the first, parties may disagree about whether certain data—like work history, medical records, or survey responses, discussed above²³⁷—is accurate. “Accurate” here has a few distinct meanings. Data may be inaccurate if it was recorded in error;²³⁸ think of a medical record where a doctor has accidentally checked a box denoting hypertension. Data also may be inaccurate if it is a poor proxy for the information that truly matters;²³⁹ obesity is correlated with diabetes, but the former certainly does not imply the latter. Or data may be accurate along both of those dimensions but suffer from statistical deficiencies like lack of randomness or representativeness.²⁴⁰ Moreover, when data *does* suffer from some accuracy problem, there are multiple ways to fix it. Inaccurate instances might be “corrected” using approximate values, dropped entirely, or supplemented with additional data.²⁴¹ As for conflicts about model training, there are many contestable options for which machine learning algorithm to use, how to partition training and testing data, and how (or how much) to fine-tune algorithms after an initial round of training.²⁴²

Problems like these may sound daunting, but they are not foreign to complex civil litigation. Some could be resolved by applying basic evidentiary rules. Rule 401 of the Federal Rules of Evidence requires that evidence be relevant.²⁴³ Data that has been systematically misrecorded or that has *no* correlation to operative facts fails that test. And data with only a weak correlation to the facts that matter might be excluded under Rule 403 for its potential to confuse or prejudice the jury.²⁴⁴

Other disputes would likely be resolved via “battles of the experts.” In such battles, both parties submit their own expert opinions about proper algorithmic design. The competing opinions are evaluated by some

235. These variables are usually well-defined. For elements of liability, the jury supplies a simple “yes” or “no” as to whether the element is satisfied. For damages, the output is a dollar figure.

236. These correspond to Lehr and Ohm’s steps 2–4 and 5–7, respectively. *See* Lehr & Ohm, *supra* note 233, at 677–701.

237. *See supra* notes 190–92 and accompanying text.

238. *See* Lehr & Ohm, *supra* note 233, at 679.

239. *See id.*

240. *See id.* at 680.

241. *Id.* at 681–83.

242. *See generally id.* at 684–701.

243. FED. R. EVID. 401.

244. *See* FED. R. EVID. 403 (“The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of . . . unfair prejudice, confusing the issues, [or] misleading the jury . . .”).

combination of the court—applying the *Daubert* test²⁴⁵—and the factfinder—choosing the best argument that survives *Daubert*. These battles are not new, either. Indeed, in complex litigation—including class actions—they are regularly fought specifically *about model design*. Older-fashioned statistical models—like traditional regressions—are mainstays of securities, antitrust, and disparate impact litigation.²⁴⁶ These models, like advanced machine learning algorithms, can only be as accurate as the underlying data. They therefore also raise opportunities for expert disagreement about whether faulty data has been adequately corrected. Likewise, as with training a machine learning algorithm, traditional models rely on their designers’ decisions about how best to make the model fit the data. Different regression techniques, for example, fit the data to different types of curves.²⁴⁷ Which technique to use is a matter of expert discretion.

Some readers may worry that courts and factfinders are ill-equipped to referee expert debates over technical issues of algorithmic design. But the reality is that they *already* do this sort of thing all the time. The world is a technical place. And governing it increasingly involves technical thinking to be embedded in legal rules and decisions. Certainly, if AI class actions become widespread, there will be an adjustment period as courts develop doctrine about what constitutes credible scientific practice in algorithmic design. Such bumps in the road, however, are the cost of admission if generalist judges are to continue playing any major role governing our increasingly complex world.

There is also another, machine-learning-specific, mechanism by which some disputes over algorithmic design might be resolved. Call this “battle of the algorithms.” Many of the above-described disputes can be boiled down to the following conflict: Which design approach—the plaintiff’s or the defendant’s—will produce a final algorithm that most accurately mimics the jury’s decision procedure? This is a testable proposition. To find the answer, courts could borrow the approach of Kaggle—a Google subsidiary—and conduct a competition.²⁴⁸ Both parties would train their preferred version of the algorithm. Then, both final versions could attempt to predict the jury’s decisions about the hold-out set of training data.²⁴⁹ The party with the most

245. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993). Under *Daubert*, the court acts as a gatekeeper, excluding expert opinions that lack sufficient scientific credibility. *Id.* at 597. The *Daubert* factors are now codified in section 702 of the Federal Rules of Evidence. *See* FED. R. EVID. 702.

246. *See supra* notes 60, 80, 85 and accompanying text.

247. *See, e.g.*, Oleksii Kharkovyna, *Key Types of Regressions: Which One to Use?*, DATASOURCE (Aug. 20, 2020), <https://www.datasource.ai/en/data-science-articles/key-types-of-regressions-which-one-to-use> [<https://perma.cc/N39L-EGCR>].

248. *See Competitions*, KAGGLE, <https://www.kaggle.com/competitions> [<https://perma.cc/NC9J-FDVR>].

249. *See supra* notes 207–08 and accompanying text.

accurate results²⁵⁰ would win, and that algorithm be adopted for the remainder of the AI class action.²⁵¹

E. *Settlement Dynamics*

For all the talk above about how machine learning could be used at trial, actual examples of its use—if AI class actions were adopted—would likely be rare. Trials, of course, are rare in civil lawsuits generally—and vanishingly so in class actions.²⁵² Instead, nearly all cases settle. Thus, rather than becoming a mainstay of civil adjudication, the primary effect of introducing algorithmic decision-making into class actions would be a shift in settlement dynamics.

Settlements reflect parties' best estimates of the litigation's value.²⁵³ Under the prevailing post-*Wal-Mart* paradigm, those values are, for many categories of cases, pushed below the efficient level. Again, this is because, for many small but meritorious claims, expected recoveries exceed litigation costs only if litigable as a class action.²⁵⁴ Thus, when class-wide adjudication is impossible, the value of those claims is arbitrarily pushed to zero. And even for claims with positive value in individual litigation, losing the efficiencies of the class form significantly reduces their value. Thus, the settlement value of a collection of claims depends significantly on the probability of certification.

If machine learning is introduced into class actions, the class vehicle will become newly available across a wide swath of claims in a number of doctrinal areas.²⁵⁵ This eliminates the extra costs of individual litigation and pushes settlement values up. Defendants, obviously, will not be pleased with that effect.

But there are good reasons to believe that such an increase in settlement values would be good for society as a whole. Presumably, when legislators²⁵⁶ create legal liability, it is in response to socially costly activity. That is, they

250. There is more than one way to measure accuracy, leading to the possibility of contested results. See *infra* note 273. But in most cases, when one algorithm clearly outperforms the other, these differences in evaluation methodology will not matter. Additionally, a court might specify the accuracy measure at the outset of the competition, before either side knew which measure would most benefit them.

251. For a specific example about how a dispute over data accuracy might be resolved via battle of the algorithms, see *infra* notes 343–46 and accompanying text.

252. See Nagareda, *supra* note 3, at 99 (“With vanishingly rare exception, class certification sets the litigation on a path toward resolution by way of settlement, not full-fledged testing of the plaintiffs’ case by trial.”).

253. Frank H. Easterbrook, *Justice and Contract in Consent Judgments*, 1987 U. CHI. LEGAL F. 19, 19 (1987).

254. See *supra* note 7.

255. See *supra* notes 4–6 and accompanying text.

256. Or, for common law claims, judges.

judge the effects of some activity to be harmful, and they give those affected the right to recover for that harm. When plaintiffs recover their actual damages, this generates efficient deterrence and also serves other values like corrective justice.²⁵⁷ Thus, when litigation costs artificially push down recoveries—sometimes to zero—that counts as suboptimal.

But what about claims by which plaintiffs can recover *more* than their actual damages?²⁵⁸ Is pushing up settlement values for those claims a good thing? One reason for imposing damages above actual harm is in an attempt to achieve optimal deterrence, even in the face of otherwise-suboptimal enforcement.²⁵⁹ And if the main reason legislators originally anticipated suboptimal enforcement was because of hurdles to class certification, then removing those barriers might result in *over-enforcement*. However, there are many other possible sources of underenforcement, most importantly the under-detection of violations.²⁶⁰ Moreover, there are other reasons to increase damages above actual harm. Doing so might deal justice for particularly onerous wrongs or promote the moral repair of hard-to-value harms.²⁶¹ We therefore ought not worry very much that by pushing settlement values up, AI class actions would inadvertently work some social ill.

Increased settlement values would be a gift to plaintiffs. But AI class actions could also benefit defendants in certain settlement situations. Occasionally, defendants argue, classes are certified before the parties have determined which putative class members were injured and which were not.²⁶² And an order granting certification can lead to an immediate “‘in *terrorem*’ settlement[.]”²⁶³ These factors combine to produce the possibility that when some classes are certified, defendants overpay.

Machine learning algorithms can help parties negotiating settlement avoid *both* underpayment *and* overpayment. Just as in an AI-driven trial, where algorithms would sort valid claims from invalid ones and estimate plaintiff-by-plaintiff damages, they could do so during settlement. In fact, the process in settlement would be even more straightforward. Freed from the need to emulate a *particular* jury’s decision function, the parties would have no need for sample adjudications. Instead, they could directly deploy commercially produced, off-the-shelf algorithms designed to answer the

257. See *supra* notes 102–08 and accompanying text.

258. See, e.g., 15 U.S.C. § 15(a) (providing for treble damages in antitrust suits).

259. See, e.g., A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 889 (1998).

260. See *id.*

261. See Carleen M. Zubrzycki, *Punitive Damages in an Era of Consolidated Power*, 98 N.C.L. REV. 315, 317 n.3 (2020).

262. See, e.g., *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 452 (2016).

263. *Amgen Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 477 (2013).

individual questions at hand.²⁶⁴ Such algorithms would provide quick and cheap answers to the question, “What would an average jury say about the validity and value of each class member’s claim?” Those answers could be transposed directly into settlement agreements, since settling parties have little reason to believe that *their* jury would be different from the *average* one.

The introduction of AI class actions could also help facilitate *earlier* settlement, helping parties avoid wasteful litigation. Under the current paradigm, certification is often uncertain, making it difficult for parties to agree at a lawsuit’s outset on a settlement value. Putative class actions are therefore often litigated to the point of a certification ruling, with settlement becoming a near-certainty immediately thereafter.²⁶⁵ And because class certification decisions often involve a hard look at the merits, such litigation can be extremely costly.²⁶⁶ If AI class actions render some set of questionable cases clearly certifiable, then parties in those cases will be able to settle confidently at the outset, avoiding significant waste.

III. Can AI Class Actions Overcome the *Wal-Mart* Problem?

Assume that, for some significant set of currently uncertifiable classes, machine learning could feasibly be employed in litigation to efficiently resolve individual questions, allowing common ones to predominate. Would such a system pass muster under *Wal-Mart* and its progeny? This Part argues that the answer is yes; the proposal would survive in either a strong or a slightly weakened form.

A. *AI Class Actions Are More Than Accurate Enough to Satisfy Wal-Mart*

If, as argued above, values relating to individual accuracy animate *Wal-Mart* and its progeny, then the question for AI class actions is how accurate they are. Specifically, would AI class actions be at least as accurate—at the individual level—as the methods of statistical proof the Supreme Court allows?

Assume, for the moment, that algorithmic decisions in AI class actions would be, like jury determinations, the final word about individual questions. This is the simplest case to evaluate under *Wal-Mart*, since it offers no opportunity to correct potential errors in algorithmic outputs. A more complex version of AI class actions—where algorithmic determinations operate as mere presumptions that are challengeable by either party—is analyzed below.²⁶⁷

264. See *supra* notes 19–23, 231–32 and accompanying text.

265. See Nagareda, *supra* note 3, at 99.

266. See *Comcast Co. v. Behrend*, 569 U.S. 27, 33–34 (2013).

267. See *infra* subpart III(B).

Even in this simple version of the proposal, AI class actions would be very accurate. Accuracy is the whole point of advanced machine learning algorithms. One way to understand the difference between cutting-edge algorithms and older-fashioned linear regression is that the former increases accuracy—often dramatically—while sacrificing interpretability.²⁶⁸

Just how accurate would machine-learning-driven resolutions of individual questions be? There is no way to give a precise figure, in advance, as to every conceivable individual question. But we have guideposts. The designers of the Blue J algorithm reported in 2017 that “[i]n out-of-sample testing, in a variety of different questions in tax law, Blue J consistently gets more than 90% of predictions correct.”²⁶⁹ Since many individual questions in class actions are similar in nature to those that Blue J predicts,²⁷⁰ we might expect similar performance.²⁷¹ Similarly, classification algorithms regularly achieve accuracy approaching 100% across a range of at least moderately difficult non-legal questions.²⁷² It seems likely, then, that for many run-of-the-mill individual questions in class litigation, algorithms will be able to produce correct results nearly all the time.²⁷³

Would an algorithm that could correctly predict a jury’s decisions about individual questions over ninety percent of the time be good enough?

268. See Toshiki Mori & Naoshi Uchihira, *Balancing the Trade-Off Between Accuracy and Interpretability in Software Defect Prediction*, 24 EMPIRICAL SOFTWARE ENG’G 779, 780 (2019).

For more on this “black box” problem of interpreting AI decisions, see *infra* subpart IV(B).

269. Alarie, *supra* note 209, at 242.

270. See *supra* notes 209–14 and accompanying text.

271. See Alarie, *supra* note 209, at 241. Even legal algorithms trained to predict the absolute least-certain legal issues—the outcomes of Supreme Court cases—can do so with roughly seventy percent accuracy. See Katz, *supra* note 214, at 8. One of the main jobs of the Supreme Court is to resolve the hardest cases, many of which have split the Circuits. Moreover, Supreme Court cases often involve multiple difficult issues, each of which can independently change the disposition. We therefore ought to expect these results to be the floor, not the ceiling, of AI performance answering legal questions.

272. See, e.g., Mariusz Mlynarczuk & Marta Skiba, *The Application of Artificial Intelligence for the Identification of the Maceral Groups and Mineral Components of Coal*, COMPUTS. & GEOSCIENCES, June 2017, at 133, 136 (identifying the success of AI in mineral classification); Sanjay Sharma, C. Rama Krishna & Sanjay K. Saha, *Detection of Advanced Malware by Machine Learning Techniques*, in ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING 333, 335 (Janusz Kacprzyk ed., 2019) (identifying the success of AI in malware identification); Shuai Ye, Cuixia Li, Ruoyan Zhao & Weidong Wu, *NOAA-LSTM: A New Method of Dialect Identification*, in ARTIFICIAL INTELLIGENCE AND SECURITY 5TH INTERNATIONAL CONFERENCE, ICAIS 2019 NEW YORK, NY, USA, JULY 26–28, 2019 PROCEEDINGS, PART I 16, 21 (Xingming Sun, Zhaoqing Pan & Elisa Bertino eds., 2019) (identifying the success of AI in dialect classification); Eric Chan, *Predicting Fake News with 99% Accuracy*, GITHUB (Nov. 23, 2018), <https://echanclarityinsights.github.io/posts/2018/Nov/23/predicting-fake-news-with-99-accuracy/> [<https://perma.cc/5AKY-ND7L>] (identifying the success of AI in sorting real from fake news).

273. Technical measures of algorithm accuracy—like, for example, “F1 scores”—account for both false positive and false negative rates. See, e.g., Shumeet Baluja, Vibhu O. Mittal & Rahul Sukthankar, *Applying Machine Learning for High-Performance Named-Entity Extraction*, 16 COMPUTATIONAL INTEL. 586, 587 n.3 (2000).

Obviously, neither the *Wal-Mart* Court nor any other has articulated a clear cutoff. But examining the differences between disallowed statistical proof—like the regressions in *Wal-Mart*—and allowed proof—like the market analysis in *Halliburton*—provides substantial guidance.

The *Wal-Mart* Court believed that the accuracy of the plaintiffs' regression analysis might be very low, indeed. Again, the Court understood the analysis to show only aggregate gender disparities at the level of forty-one U.S. corporate regions—each containing, on average, tens of thousands of employees.²⁷⁴ This, the Court thought, left opportunity for high aggregation bias: “A regional pay disparity, for example, may be attributable to only a small set of Wal-Mart stores” or managers.²⁷⁵ And absent a unified system for determining pay, “almost all of [the managers] will claim to have been applying some sex-neutral, performance-based criteria[.]”²⁷⁶ Ultimately, the Court thought that such arguments, made by “almost all” of the managers would be highly plausible: “[L]eft to their own devices most managers in any corporation . . . would select sex-neutral, performance-based criteria for hiring and promotion that produce no actionable disparity at all.”²⁷⁷ In sum, the Court thought that the statistical evidence left open the possibility of getting most liability determinations wrong. And the proposed causal story, the Court determined, made that possibility *likely*.²⁷⁸

Contrast this with the fraud-on-the-market proof in *Halliburton*. There, the Court credited recent economic scholarship showing that markets are not perfectly efficient.²⁷⁹ It likewise acknowledged that not all investors rely on the efficiency of market prices to transmit information—including alleged misrepresentations—to them.²⁸⁰ Yet, despite these potential sources of error, the Court determined that *actual* error rates would be quite low. This was

274. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 342, 356–57 (2011).

275. *Id.* at 357.

276. *Id.* (emphasis added).

277. *Id.* at 355 (emphasis added).

278. It is, of course, possible that the majority misunderstood the evidence, as the concurrence charged. *See id.* at 372 n.5 (Ginsburg, J., concurring in part and dissenting in part). Moreover, there are ways to check regression analyses for aggregation bias. *See, e.g.*, Uzi Landau, *Aggregate Prediction with Dis-Aggregate Models: Behavior of the Aggregation Bias*, TRANSP. RSCH. REC., 1978, at 100, 102; *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137, 156–57 (N.D. Cal. 2004).

The sampling design in *Wal-Mart* presents a more complicated picture. Since that design treats *all* of the non-sample class members as having a valid claim, its error rate depends on the true rate of invalid claims in the class. That rate would be estimated, to a high degree of statistical certainty, by the sample trial process itself. It therefore would have been possible to conduct the sample trials and then determine whether the resulting error rate would be acceptably low. Nothing in *Wal-Mart* indicates that any party advocated for this approach. But as the discussion above shows, the Court evidently believed that the actual rate of invalid claims—and thus sampling's error rate—would likely be quite high.

279. *Halliburton Co. v. Erica P. John Fund, Inc.*, 573 U.S. 258, 271–72 (2014).

280. *Id.*

because, the Court thought, “market professionals *generally* consider *most* publicly announced material statements about companies, thereby affecting stock market prices.”²⁸¹ Likewise, “it is reasonable to presume that *most* investors—knowing that they have little hope of outperforming the market . . . —will rely on the security’s market price.”²⁸² Thus, the Court judged the fraud-on-the-market presumption largely accurate, with defendants needing only to “pick off the occasional class member here or there through individualized rebuttal.”²⁸³

Compared against these two examples, AI class actions look quite good. If algorithms can resolve class members’ individual questions with accuracy rates in the high nineties, they look much more like *Halliburton* than *Wal-Mart*. Their decisions would not be “most[ly]” wrong,²⁸⁴ but—by a significant margin—“most[ly]” right.²⁸⁵ Indeed, they would be much better than that. Their low error rates would produce wrong results only for “the occasional class member here or there.”²⁸⁶

Taking a more theoretical view, we should think about cutting-edge algorithms as being *at least* as accurate as the very best traditional statistical models. Again, the whole point of them is to significantly outperform, for example, traditional regression analyses²⁸⁷ at classification and quantification in individual cases.²⁸⁸ And again, algorithms do this by enabling extreme nonlinearity in their decision functions.²⁸⁹ That is, they can account for substantially more complicated relationships between inputs and outputs, generating individual answers that, in general, lie significantly closer to the truth.²⁹⁰ Thus, if the Supreme Court endorses, for example, well-calibrated regression analysis, so too should it endorse machine-learning-based proof.

281. *Id.* at 272 (emphasis added) (quoting *Basic Inc. v. Levinson*, 485 U.S. 224, 247 n.24 (1988)).

282. *Id.* at 273 (quoting *Amgen Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 462 (2013) (emphasis added)).

283. *Id.* at 276.

284. *Wal-Mart Store, Inc. v. Dukes*, 564 U.S. 338, 355 (2011).

285. *Halliburton*, 573 U.S. at 272–73.

286. *Id.* at 276.

287. Highly nonlinear algorithms like neural networks and random forests do not invariably outperform simpler regression analysis. For some problems, changes in independent variables may be explained almost perfectly using a linear function from regressors. But usually, the selected regressors leave much of the variation between individuals unexplained, producing relatively high error. See Bone, *supra* note 45, at 584–85.

288. See Mori & Uchihira, *supra* note 268, at 780.

289. See Gillis & Spiess, *supra* note 173, at 469.

290. *Id.*

B. Weak AI Class Actions Provide Even More Bang-for-the-Buck

The preceding examination of AI class actions' legality assumed that algorithmic answers to individual questions would be the final word on those questions. It argued that such a system would survive scrutiny under *Wal-Mart*. Nevertheless, some readers may have lingering doubts. And even those who accept that the above-described version of the proposal would survive *Wal-Mart* scrutiny might nevertheless wonder whether it could be further improved.

This subpart explores a modified, weaker version of the AI class action proposal designed to allay any lingering concerns. The weaker version promotes marginal improvements in individual accuracy, further shoring up the argument under *Wal-Mart*. And at the same time, the weakened proposal makes improvements in terms of pure process values like legitimacy and autonomy. Although those values are not what drives acceptances and rejections of statistical proof under *Wal-Mart*, they are nevertheless important. Thus, it makes sense to promote them when doing so requires little sacrifice elsewhere.

The weakened proposal borrows a design feature from a few of the cases in *Wal-Mart*'s lineage: rebuttable individual presumptions. Observe that the methods of statistical proof allowed after *Wal-Mart* fall into two categories. Sometimes, statistically derived answers to individual questions are the final word. *Tyson* authorized the use of average donning and doffing times because no evidence of individual times existed.²⁹¹ Thus, the only options were to apply the average answer to every class member or to reject it wholesale.²⁹² Similarly, in antitrust, most courts do not permit defendants to rebut statistical proof of antitrust injury on a plaintiff-by-plaintiff basis.²⁹³ But other permissible methods of statistical proof create presumptions that may be individually rebutted. Fraud-on-the-market and Title VII disparate impact work this way.²⁹⁴ Defendants are entitled to present evidence that any

291. *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 422, 454, 456–47 (2016).

292. *See id.* at 454. The *Tyson* Court did speak about the defendant's opportunity to present its own evidence about the "amount of work performed." *Id.* at 456 (quoting *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687–88 (1945)). But since *Tyson* allows statistical proof only when individual evidence is unavailable, defendants' arguments here are logically limited to showing why the plaintiffs' average figure is wrong. *See id.* at 443, 454, 456–57.

293. *See* Joshua P. Davis & Eric L. Cramer, *Of Vulnerable Monopolists: Questionable Innovation in the Standard for Class Certification in Antitrust Cases*, 41 RUTGERS L.J. 355, 394 & n.124 (2009) (collecting cases). *But cf. In re Hydrogen Peroxide Antitrust Litig.*, 552 F.3d 305, 311–13 (3d Cir. 2009) (questioning whether statistical evidence could adequately prove injury as to all class members).

294. *See* *Amgen Inc. v. Conn. Ret. Plans & Tr. Funds*, 568 U.S. 455, 463 (securities fraud); *Tex. Dep't of Cmty. Affs. v. Burdine*, 450 U.S. 248, 254 (1981) (Title VII).

individual plaintiff did not, respectively, actually rely on efficient market prices or suffer an injury from a discriminatory policy.²⁹⁵

Similarly, the weak version of AI class actions would treat algorithmic answers to individual questions as rebuttable presumptions.²⁹⁶ Either party might challenge them. Defendants might challenge an individual finding of liability or argue that damages should be lower, and plaintiffs would do the opposite. Parties would mount their challenges in precisely the same manner as they would in, say, a Title VII disparate impact case. The challenging party would bear a burden of production, requiring a showing sufficient to persuade a reasonable jury that the algorithm's decision was wrong.²⁹⁷ Such a showing might include the introduction of new evidence. Or it could amount to a particularized argument that the algorithm misinterpreted the existing evidence. Defendants could also raise affirmative defenses applicable only to individual class members. Whenever a challenger carried its burden of production, the challenge would be litigated to judgment via normal procedures, potentially terminating in a jury decision.

How would moving to this presumption-based model promote pure process values? Under the strong version of the AI class action proposal, the only parties who would participate in the adjudicatory process would be the defendants, the named plaintiffs, and the sample class members. This means that non-sample members who felt, for example, that dignitary considerations required that they have a say in court would be denied that say. Defendants likewise would lack the opportunity to air their side of the story about individual machine-derived results.

Weak AI class actions can do better. In a weak AI class action, any defendant or plaintiff who wished, in service of process values, to be heard on an individual question could do so. Most of the time, as discussed below, this would be a money-losing proposition. But for a few, the cost of litigation would be worth the satisfaction of participation. Common sense and experience teach that such litigants—those who place an outsized value on

295. See *Amgen*, 568 U.S. at 463; *Burdine*, 450 U.S. at 254.

296. In addition to the case law, this proposal is inspired in part by Jay Tidmarsh's proposal that *Wal-Mart*-style sampling could be improved by making the resulting awards presumptive. See Tidmarsh, *supra* note 2, at 1464.

297. See, e.g., *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 658–59 (1989) (describing challenges to Title VII disparate impact presumptions); Eric R. Sunde, Case Note, *Title VII Disparate Impact Theory—Employer's Burden of Rebutting Prima Facie Case Under Disparate Impact Theory Is One of Production While Ultimate Burden of Persuasion Remains with Complainant*, 21 ST. MARY'S L.J. 1081, 1090–92 (1990) (same). Note here that, even when a burden of production shifts to the defendant, the burden of proof may remain with the plaintiff. *Id.* at 1092. Structuring rebuttals of algorithmic decisions this way avoids any possible criticism, rooted in the Rules Enabling Act, that the scheme changes the parties' substantive rights. See *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011).

process—are rare.²⁹⁸ Thus, few would challenge algorithmic presumptions for the sake of pure participation. But this is little detriment to legitimacy or other process values. What matters to them is the bona fide *opportunity* to participate, even if waivable.²⁹⁹

A shift to presumptive algorithmic determinations would also further improve accuracy. Almost always, some small set of algorithmic outputs in an AI class action will be wrong. When those wrong answers are obvious to either the plaintiff or the defendant, then challenging them—and successfully obtaining the correct result—increases accuracy on the margin.

Some readers may be concerned that this shift in the proposal gives away the whole game. If defendants maintain the right to challenge the algorithm's results as to *every* class member, won't they do it? And if they do, demanding that those individual questions be resolved in innumerable mini-trials, won't predominance be lost and certification be defeated?

Likely not. Begin with a simple model³⁰⁰ of litigation incentives. Under this model, parties initiate an action only if the payoff (P) times the likelihood of success (L) is larger than the action's costs (C): $LP > C$.³⁰¹ In a weak AI class action, the potential challenger to an algorithmic determination must weigh these factors. The payoff is the difference in damages paid (for the defendant) or recovered (for the plaintiff) that a successful challenge would net. For challenges to an individual question that forms a requisite element of liability, P is the entire damages award owed—or potentially owed—to the relevant class member. For challenges to an algorithmic calculation of damages, P is the increase or reduction in damages that results from a successful challenge. C is the total cost of litigating the challenge, including the upfront cost of sustaining the burden of production necessary to initiate it.³⁰²

The crucial point here is that the more accurate the algorithm's initial determinations are, the lower PL is. As PL approaches \$0, it becomes—at a first cut—irrational to challenge algorithmic determinations at any cost. And

298. If it were otherwise, then we would have little need for class actions in the first place. Even negative-value claims would commonly be individually litigated in service of process values.

299. Solum, *supra* note 100, at 280.

300. For a discussion of more complicated models incorporating strategic cost-sinking, see *infra* notes 307–19.

301. See D. Rosenberg & S. Shavell, *A Model in Which Suits Are Brought for Their Nuisance Value*, 5 INT'L REV. L. & ECON. 3, 3–4, 11 nn.3–5 (1985).

302. Note that none of these payoff figures will be identical to what those plaintiffs would face if litigating their claims from the start in individual suits. Total costs to litigate *just one* individual question—like an element of liability or damages—will likely be much smaller than costs to litigate the whole claim. Likewise, the likely payoff from challenging an existing presumptive decision will often be quite different than the likely payoff of initiating a lawsuit from scratch. Finally, the existence of a presumption often inverts the parties' positions—when the defendant, not the plaintiff, must mount a challenge to secure a payoff.

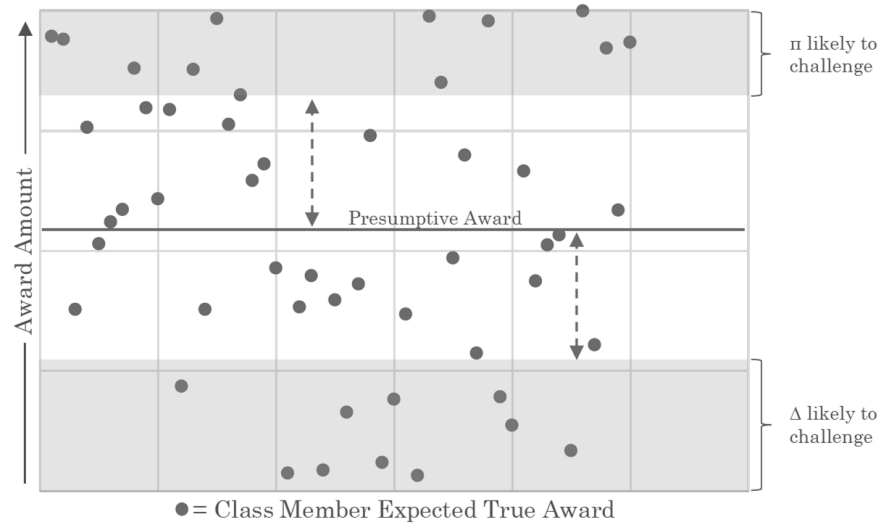
because, as described above, advanced machine learning algorithms are highly accurate, there will rarely be any facial incentive to mount *strategic* challenges. Thus, under the simple model, only those rare algorithmic decisions that are plainly wrong will be worth challenging. And these are the very cases that, in service of accuracy, *ought* to be challenged.

This point bears illustration. Jay Tidmarsh has recently advocated a resurrection of the sampling design that *Wal-Mart* rejected, but with a modified design that treats the statistical awards as merely presumptive.³⁰³ The discussion below shows why challenges to presumptions would be much rarer in weak AI class actions than under Tidmarsh's system. The analysis is the same for any statistical approach that, like Tidmarsh's, awards a kind of "average" amount to multiple individuals when their true jury awards would vary. Thus, the advantages of AI determination generally persist even when compared with statistical proof that includes subclass averaging or linear regression.³⁰⁴

303. See Tidmarsh, *supra* note 2, at 1464.

304. Common methods of regression analyses can significantly reduce individual error, as compared with sampling. But because regression analysis is highly parametric, it usually generates much more error than advanced machine learning. There are two reasons for this. First, regressions are usually run using a constrained set of pre-selected dependent variables. They therefore treat class members as identical whenever they bear the same handful of traits. Hence, like Tidmarsh's sampling design, they impute a kind of "average" answer to class members who may actually be relevantly different. Second, regressions try to fit a pre-determined type of function—simple linear, polynomial, logistic, etc.—to the data. Thus, to the extent that the data do not perfectly match the selected function, error cannot be minimized. Advanced machine learning algorithms, by contrast, are highly nonlinear, allowing outputs that lie very close to true values, without the constraint of a pre-selected function.

Figure 1



Each datapoint represents the award a class member would likely receive if a party were to challenge the presumptive award.³⁰⁵ Call this the expected true award. The solid line represents the presumptive damages award when sampling is used as a method of statistical proof. Recall that, under that system, every plaintiff gets the sample's average award. The absolute vertical distance from the solid line to each datapoint is thus LP—the likely payoff for a challenge. The vertical lines with arrows at each end represent C—the cost to mount a challenge. The top and bottom shaded zones are thus where a financial incentive exists for plaintiffs and defendants, respectively, to mount challenges.

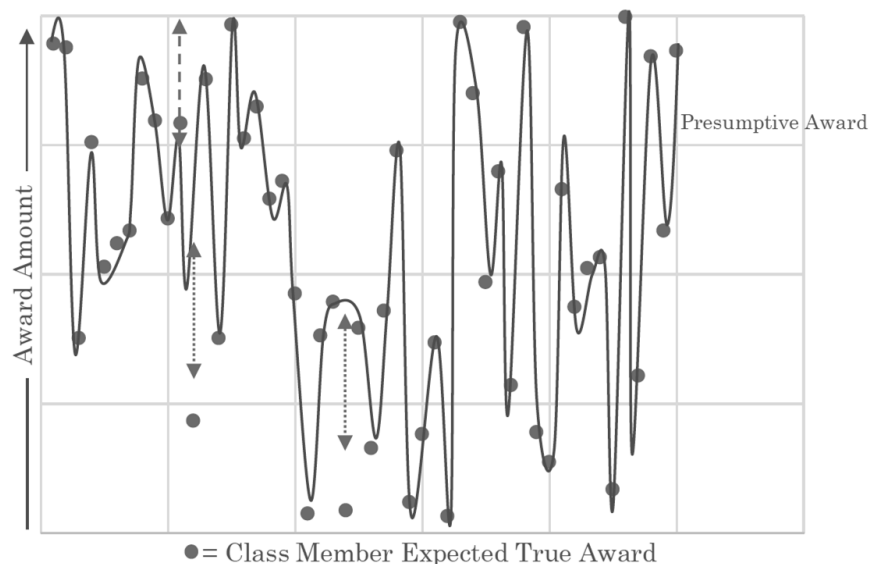
The shaded zones become smaller as either C rises or as variance in expected true awards falls. And they disappear altogether if C happens to exceed the maximum distance between the average award and the biggest outlier. Nevertheless, it is clear that for many distributions of the true awards, and for many values of C, there are incentives to challenge large numbers of individual presumptions.³⁰⁶

Figure 2 shows what the situation might look like in an AI class action:

305. This is the *ex ante* expected award resulting from a challenge. That is, it is the award that results from a successful challenge, adjusted by the likelihood of failure.

306. If true awards are normally distributed, one can determine precisely how many challenges are viable by considering how many standard deviations C represents.

Figure 2



Data points are the same true awards as in Figure 1. And the solid line again represents presumptive awards, but this time determined algorithmically—varying individual by individual. The vertical distance between any given dot and the solid line is therefore again LP. Because the algorithm can produce radically nonlinear results, falling very close to true awards, most LPs are extraordinarily small. Vertical, arrowed lines again illustrate C—the cost to challenge a presumptive award. Costs nearly always swamp LP, as illustrated by the dashed line. Only in a few cases—denoted with dotted lines—does it make financial sense for a party to challenge an award. Of course, some values of C and some sets of algorithmic results will yield higher challenge rates. But for high-accuracy algorithmic results, like the ones described in subpart III(A), financial incentives to challenge should remain rare.

A robust law-and-economics literature on so-called “nuisance suits”³⁰⁷ complicates the simple model under which parties will not challenge presumptions unless $LP > C$. These scholars show that, under certain conditions, even a plaintiff with a negative-expected-value lawsuit may

307. The term “nuisance,” applied in the literature to denote suits for which litigation costs exceed the small, expected damages awards, can be inapt. It implies that the plaintiff’s claim has little legal merit. In fact, there are two situations in which costs may, in expectation, exceed likely payoffs. First, a plaintiff may demand high damages but have almost no legal chance of success. Second, as emphasized in this Article, plaintiffs may have strong claims for real harms but small damages. The latter category is not a nuisance and finding a way to vindicate such claims at low social cost—perhaps via collective litigation—is a social good.

extract a positive settlement from the defendant.³⁰⁸ Likewise, a defendant whose costs to defend a suit are less than a defense's expected reduction in damages can sometimes mount it anyway and extract an even lower settlement.³⁰⁹ Equivalently, parties in AI class actions might sometimes be able to extract a "nuisance" settlement for dropping a challenge to an algorithmic output, even though $LP < C$.

However, such opportunities arise only under particular sets of conditions. Take, as examples, *asymmetries* in parties' costs or their ability to strategically time *when* costs are paid.³¹⁰ Suppose a party could file a challenge to an algorithmic presumption at low C —say \$25. Suppose further that C for the other party to defend the presumption would be \$200. Suppose finally that LP is only \$10, because the challenger has a 1% chance of winning on liability and thus reducing a presumptive \$1,000 award to \$0. As David Rosenberg and Steven Shavell show, to avoid incurring the cost of defense, the defender will pay the challenger up to \$200 to drop such a challenge.³¹¹ The analysis is the same if the challenger's *total* cost to litigate is also \$200, but it can sequence costs so it pays only \$25 before a defendant must ante up its full \$200 to avoid defaulting.³¹² In these circumstances, parties might strategically challenge presumptions, even when $LP < C$.

These conditions are likely to be rare in AI class actions. For one thing, it is unlikely that the upfront costs of challenging an algorithmic presumption would be trivial, as compared with defending one. Challengers would bear the burden of producing new evidence or legal argumentation sufficient to convince a reasonable jury that the presumption was wrong.³¹³ Conclusory or otherwise insubstantial challenges would fail on their face. They might even

308. See, e.g., Lucian Arye Bebchuk, *A New Theory Concerning the Credibility and Success of Threats to Sue*, 25 J. LEGAL STUD. 1, 4–5 (1996); Robert G. Bone, *Modeling Frivolous Suits*, 145 U. PA. L. REV. 519, 524, 529, 534, 537, 542 (1997); William H.J. Hubbard, *Sinking Costs to Force or Deter Settlement*, 32 J.L. ECON. & ORG. 545, 547, 578 (2016); Rosenberg & Shavell, *supra* note 301, at 4.

309. See Warren F. Schwartz & Abraham L. Wickelgren, *Advantage Defendant: Why Sinking Litigation Costs Makes Negative-Expected-Value Defenses but Not Negative-Expected-Value Suits Credible*, 38 J. LEGAL STUD. 235, 235–36 (2009).

310. See Rosenberg & Shavell, *supra* note 301, at 3–4.

311. See *id.*

312. See *id.*

313. Here, my proposal diverges from Tidmarsh's proposed presumptive approach. Under his proposal, the cost of challenging a presumption appears to be *de minimis*. It could perhaps be as simple as filing a document that says something like, "Defendant hereby challenges algorithmic outputs numbered 248 through 1,023." See Tidmarsh, *supra* note 2, at 1478 & n.70 (explaining that "[a]s soon as either party challenges the award, the presumption collapses," so that the presumption is, contra Rule 301 of the Federal Rules of Evidence, a "bursting bubble"). But it is not clear why presumptions should be so fragile. As described here, the other types of presumptions used in class actions are not. And Rule 301 imposes a similarly high general standard for challenging presumptions, requiring the challenging party to carry a substantial burden of production. See *supra* note 297.

invite sanctions.³¹⁴ Moreover, since the challenger bears its burden at the outset, the possibility of strategically delaying costs is minimized.

More recent game-theoretical models of “nuisance” claims explore additional circumstances in which incentives for strategic challenges could arise.³¹⁵ But the conditions quickly become quite complicated. Whether such incentives exist varies simultaneously with: the ratio of challengers’ costs to defenders’ costs, the ratio of LP to challengers’ C, the number of stages of cost expenditure, challengers’ ability to preemptively sink costs, defenders’ ability to preemptively pay a large retainer to its attorney, and the effect of default.³¹⁶ Moreover, these models assume perfect knowledge by both parties of all these facts—both as to themselves and their opponents.³¹⁷

These considerations combine to suggest that it would be a tricky proposition to pervasively challenge algorithmic presumptions for which $LP < C$. The financial viability of such challenges would certainly vary litigation by litigation, and likely even class member by class member. Thus, while the risk of strategic behavior to defeat class certification is not zero, it is also probably not high enough to threaten the general viability of weak AI class actions.³¹⁸

This theoretical argument for optimism is borne out by common sense and experience. Again, defendants in certain class actions *already* have the right to challenge presumptions set by statistical proof. Yet in the realms where this is allowed—say, securities fraud or Title VII disparate impact—such challenges have not been common enough to threaten class certification. As described above, the Supreme Court approves class certification when statistical proof is highly accurate, since the Court anticipates few individual challenges in those cases.³¹⁹ And the Court’s expectations seem to be right.

Thus, although the strong version of AI class actions might—like certain other non-presumptive statistical methods of proof—survive *Wal-Mart*, weakening the proposal generates additional benefits. Challenges to presumptions could correct any significant algorithmic errors. This increased accuracy is worthwhile in and of itself, but it also shores up the proposal’s legality under *Wal-Mart*. Moreover, giving litigants the option of airing their views serves additional values like legitimacy, dignity, and autonomy.

314. See FED. R. CIV. P. 11; 28 U.S.C. § 1927.

315. See generally Hubbard, *supra* note 308, at 551–84.

316. *Id.* at 551–55, 563, 567.

317. See *id.* at 547.

318. Tidmarsh has also suggested that courts overseeing class actions containing presumptions might wield an inherent power to shift fees. See Tidmarsh, *supra* note 2, at 1500. This solution could further reduce any lingering issues regarding negative-expected-value challenges.

319. See, e.g., *Halliburton Co. v. Erica P. John Fund, Inc.*, 573 U.S. 258, 276 (2014) (anticipating that defendants would need only to “pick off the occasional class member” after fraud-on-the-market presumption applied).

Finally, since challenges would be rare, these benefits are achieved without much additional expenditure of judicial resources. This is again intrinsically good—promoting cost-related values. But it is also an assurance that, in weak AI class actions, common questions would continue to predominate over individual ones, satisfying Rule 23’s demand.

IV. Can We Trust the Machines?

This Part turns to normative questions about AI class actions that resonate beyond the realm of civil procedure, and even law. Specifically, this Part addresses two families of concerns that have animated much of the broader scholarship on algorithmic decision-making. First is the potential problem of discrimination. As scholars and journalists have documented, machine learning can sometimes produce or entrench bias, disadvantaging people of color and other marginalized groups. Would introducing AI class actions risk creating such discriminatory results?

Second is the “black-box” problem. Even when advanced algorithms’ outputs are extremely accurate, it is often impossible for humans to interpret their decision functions. To what extent is this a concern for algorithmic decision-making in the class action context? Ought we demand *both* that algorithms provide highly accurate answers to individual questions *and* that they be able to “explain” their decisions?

This Part argues that, on both issues, there is less reason to worry about AI class actions than there is to worry about other applications of algorithmic decision-making.

A. Algorithmic Discrimination

A vast and growing academic literature explores the problem of potential discrimination by algorithms. Machine learning already guides decisions across a wide range of domains—college acceptance, pre-trial incarceration, hiring decisions, and more.³²⁰ And scholars have argued that the use of algorithms in such domains causes or entrenches discriminatory results.³²¹

Yet the problem is rarely algorithms, *per se*. Advanced machine learning algorithms give us the results we ask for, based on parameters we choose, by

320. See, e.g., Kleinberg, *supra* note 171, at 115; Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 237–38 (2018) [hereinafter Kleinberg, *Human Decisions*].

321. See, e.g., Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CAL. L. REV. 671, 673 (2016); Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold & Richard Zemel, *Fairness Through Awareness*, in ITCS ’12: PROCEEDINGS OF THE 3D INNOVATIONS IN THEORETICAL COMPUT. SCI. CONFERENCE 214, 214 (2012). For an archive of dozens of such studies, see *Scholarship, FAIRNESS, ACCOUNTABILITY & TRANSPARENCY MACH. LEARNING*, <https://www.fatml.org/resources/relevant-scholarship> [https://perma.cc/U6DG-WTSK].

mimicking a set of training data that we generate.³²² Thus, at a conceptual level, it is almost never the *algorithms themselves* that discriminate. Instead, it is *humans* who discriminate—either intentionally or unintentionally—as we design and implement the algorithms.³²³

Recent scholarship by leading legal academics ably catalogues the most likely avenues by which discrimination can invade algorithmic decision-making.³²⁴ First, the algorithm’s designers may set the algorithm to answering the wrong questions.³²⁵ Consider a firm that wishes to predict job performance among applicants for a data-entry job. To do so, the firm needs to decide what it means by “performance.” Perhaps the firm thinks of an idle computer as a wasted resource, so it chooses “hours worked” as its measure of performance.³²⁶ This selection is likely to disadvantage women who apply to the job, “since home-life differences may lead to gender differences in hours.”³²⁷ Better to choose another interpretation of “performance”—like, for example, accuracy or speed—that is less likely to systematically disadvantage people of one gender.

Second, even when an algorithm’s designers pick a facially reasonable question for an algorithm to answer, the training data may be discriminatorily mislabeled.³²⁸ This is one potentially serious problem with, for example, algorithms designed to aid in decisions about incarceration. Suppose a criminal justice system uses an algorithm to determine which arrestees will be released pending trial.³²⁹ To decide, it asks the algorithm to predict who is most likely to be rearrested upon release.³³⁰ The training data, then, must be labelled with “correct” answers about who was actually rearrested. But if, for example, African Americans are arrested at higher rates than whites, *even when their actual rates of offense are the same*, the data itself reflects a discriminatory bias. So, too, will an algorithm that decides based on such data. Similar problems arise if, although arrest rates are unbiased, the process of *recording* or *collecting* arrest data systematically over- or under-records some group.³³¹

322. See, e.g., Barocas & Selbst, *supra* note 321, at 677–93; Kleinberg, *supra* note 171, at 139–43.

323. See, e.g., Barocas & Selbst, *supra* note 321, at 678; Kleinberg, *supra* note 171, at 142–43.

324. See, e.g., Barocas & Selbst, *supra* note 321, at 677–93; Kleinberg, *supra* note 171, at 139.

325. See Kleinberg, *supra* note 171, at 139; see also Barocas & Selbst, *supra* note 321, at 679.

326. Kleinberg, *supra* note 171, at 139.

327. *Id.*

328. See, e.g., Barocas & Selbst, *supra* note 321, at 681.

329. See generally Kleinberg, *Human Decisions*, *supra* note 320.

330. *Id.* at 239.

331. See Barocas & Selbst, *supra* note 321, at 684; Kleinberg, *supra* note 171, at 140; see also Latanya Sweeney, *Discrimination in Online Ad Delivery*, COMM. ACM, May 2013, at 44, 47. Note that, even in this situation, where the data encodes some bias, algorithmic decisions may still be

A slight variation on these stories is discrimination that arises during “feature selection”—the determination of what inputs an algorithm will be given to use in predicting the target variable.³³² Suppose again that a firm wishes to train an algorithm to assist with hiring decisions. And suppose the firm correctly thinks that academic performance will correlate with job performance. The firm has relatively easy access to data about where high-performing employees went to high school and which high schools produce, on average, students with strong academic acumen. The firm, however, does not know its employees’ high school grades. The firm might then be tempted to use high-performing high schools as a proxy for high academic performance.³³³ But persistent racial segregation in the United States forces minority students to attend, on average, lower quality high schools.³³⁴ Thus, an algorithm trained solely on data about which high school employees attended would likely end up disfavoring minority candidates.³³⁵ Training instead with a richer set of features—including, perhaps, high school grades—could avoid the imbalance.³³⁶

It is certainly *possible* that bias of each of these kinds could infect AI class actions. But there are also reasons to believe that, having identified the possible sources of trouble, litigants and judges could most often avoid it.

Consider that, in the examples above, discrimination arises only because the algorithm’s designers lack access to what statisticians call “ground truth.” Ground truth can be thought of as the fact of the matter, as contrasted with data that may deviate from the reality.³³⁷ In the first case of the hiring firm, the real desideratum—“performance”—is nebulous, encompassing many factors in hard-to-quantify proportions. The algorithm is instead asked to predict something concrete and measurable—hours worked—despite that measure failing to reflect ground truth about performance. Insofar as those who work the most are not actually those who perform the best, discrimination—intentional or inadvertent—can creep into the gap.

AI class actions would have an important advantage in the realm of ground truth. In them, algorithms are asked to predict the decisions of the jury in the case. And in the context of civil litigation, we treat jury determinations as ground truth. That is, it is the jury’s job to resolve factual

significantly less biased than human decisions based on the same data. See Kleinberg, *Human Decisions*, *supra* note 320, at 277 tbl.7.

332. See Barocas & Selbst, *supra* note 321, at 688; Kleinberg, *supra* note 171, at 140–41; Ke Wang & Suman Sundaresh, *Selecting Features by Vertical Compactness of Data*, in FEATURE EXTRACTION, CONSTRUCTION AND SELECTION: A DATA MINING PERSPECTIVE 71, 71–72 (Huan Liu & Hiroshi Motoda eds., 1998).

333. Kleinberg, *supra* note 171, at 140.

334. *Id.*

335. *Id.*

336. *Id.*

337. See Barocas & Selbst, *supra* note 321, at 682 & n.37.

questions and determine what actually happened. With few exceptions,³³⁸ the jury's determination is the final word. Doubtless, juries make mistakes. But as long as the case was not beset with serious procedural defects, and the jury's determination was within the realm of the reasonable, that does not matter.³³⁹ Civil jury verdicts are not revocable *ex post* simply because, as things turn out, the jury got things wrong.³⁴⁰ Thus, unlike with the hiring firm, algorithm designers in AI class actions do not need to rely on proxies for the real question they care about. Rather, they know the true answer to the right question. Armed with this ground truth, AI class actions are immunized from bias that occurs when an algorithm's designers ask it to answer the wrong question.

Discriminatory mislabeling of data is likewise a problem of ground truth. In the case of bail decisions, the problem is that the data do not accurately reflect the thing we actually care about. What ought to matter is who will commit *offenses* upon release. But when policing is racially unequal, patterns of arrest will fail to match that underlying reality.

The remaining source of potential bias—poor feature selection—presents a slightly subtler mismatch between data and reality. It is not so much that the selected data fail to reveal something relevant. Nor are the data inaccurate, given what they purport to record. It is just that the thing they record fails to represent the whole picture.

The structure of AI class actions helps to guard against these shortcomings in the training data. Almost uniquely among applications of machine learning, AI class actions permit the party to whom the algorithm will be *applied* to have a say in how it will be *designed*. The job applicant has no voice in the construction of the computerized hiring screener. Nor does the pretrial detainee help design the release-risk algorithm. But the class members, via their attorney, have *primary responsibility* for assembling training data—in the form of their evidence. They therefore have every opportunity to screen data for biased mislabeling or inadequate richness of features. Class attorneys also have an *incentive* to do this work. Insofar as the training data unfairly disadvantages some subgroup of their class, the result is a lower recovery and lower attorneys' fees.³⁴¹ Defendants, conversely, have an incentive to police any data selection that produces biased results by unfairly advantaging some subgroup. Disagreements about data bias could be resolved via normal evidentiary rulings, as described above.³⁴²

338. See FED. R. CIV. P. 50(a).

339. See FED. R. CIV. P. 50(a)(1)(B).

340. See FED. R. CIV. P. 50(a).

341. Assuming a contingency fee arrangement.

342. See *supra* notes 243–44 and accompanying text.

Disputes about what data should be used for training might also be resolved using a “battle of the algorithms.”³⁴³ Suppose the plaintiffs maintain that minority class members’ medical records are discriminatorily mislabeled, failing to reflect their true medical histories.³⁴⁴ They argue that, because of this, the medical records of all minority class members should be excluded from the training data. The defendants disagree on both counts. The dispute can be resolved by allowing the parties to each train their own preferred version of the algorithm. These two versions can then be tested against the hold-out set of training data. That training data reflects the jury’s ground-truth determinations, made after full and fair litigation about the accuracy of the sample members’ medical records. If the defendants’ algorithm more accurately reproduces the jury’s decisions about minority class members, then training using medical records causes no harm. In that case, the jury apparently credited the records’ accuracy. But if the plaintiffs’ version is more accurate, this suggests the jury credited the plaintiffs’ arguments that medical records misrepresented the sample members’ true histories.³⁴⁵ Then, the minority members’ records should be excluded from training.³⁴⁶

Of course, none of these solutions are a cure for structural racism, misogyny, or other entrenched power imbalances. For example, what if the *jury* is biased—either consciously or subconsciously—against some subgroup? Then, a perfectly performing algorithm would reproduce that bias. And even if the jury is perfectly fair-minded, the world itself might be unfair. For example, all of the available evidence in an employment discrimination class action might *accurately* show that black class members had, on average, fewer traditional qualifications. Such disparities are, of course, intimately connected with this country’s history of slavery and white supremacy. But an algorithm might nevertheless follow a jury in correctly determining that black applicants were less likely to be hired by the defendant than their white counterparts, even under a “fair” system.

These are serious problems. But they are broad, systemic ones. They do not arise from the introduction of machine learning into class adjudication. Class members from vulnerable groups are just as likely to encounter biased juries in individual litigation as in a class action. And accurate records of

343. See *supra* notes 248–50 and accompanying text.

344. Health law scholarship suggests that race can influence doctors’ interpretations of symptoms and their subsequent diagnoses. See Barbara A. Noah, *A Prescription for Racial Equality in Medicine*, 40 CONN. L. REV. 675, 677–78 (2008).

345. Note that a similar process can be used for resolving disagreements about *which* machine learning algorithms to use. The parties can simply train their preferred algorithms, and whichever one best mimics the jury’s decision procedure wins.

346. They might, mimicking *Tyson*, be replaced with dummy variables reflecting the class’s average medical history. Here again, testing an algorithm trained on such data against one trained without any medical record information would reveal which produced higher accuracy.

employment qualifications, infected with racism as they might be, are the same whether an algorithm or a human examines them.

Even if AI class actions would not *create* or *worsen* social injustices, some might worry that it could entrench them. Imagine we trained an algorithm to adjudicate, for example, medical causation using the entire existing corpus of such decisions. And suppose we then abandoned litigation entirely, relying exclusively on the algorithm to resolve all future issues of medical causation. Then, we would be stuck indefinitely with our imperfect, present-day decision function for determining medical causation. If among that function's imperfections was, say, a bias against women, the flaw would persist forever, regardless of broader social progress.

AI class actions do not suffer from this potential pitfall. Their algorithmic decision procedures are not frozen in time. Instead, every algorithm in an AI class action would be judged against the decisions of a *new*, present-day jury. Thus, assuming juries evolved over time to become less biased, the algorithms would evolve with them.

B. *The “Black Box” Problem*

The move from traditional modes of statistical proof—like sampling and linear regression—to the advanced machine learning techniques advocated here involves a tradeoff. Such algorithms can provide huge advances in terms of the accuracy of their determinations. But generally, as accuracy goes up, interpretability goes down.³⁴⁷ Often, the decision functions on which algorithms settle are so complex, and the correlations between inputs and outputs so unexpected, as to beggar human reason.³⁴⁸ To be sure, it is possible—sometimes trivial—to look under the hood to see *how* an algorithm has linked inputs to outputs.³⁴⁹ But in many cases, understanding *why*—the causal theory about the world that drives the machine-derived correlations—will be impossible.³⁵⁰ This is called the “black box” problem.³⁵¹

There are two main reasons to worry about the black-box nature of algorithmic decision-making. First, the legitimacy of certain kinds of legal decisions may demand public reason-giving. And a set of machine-generated “reasons” that defy human understanding is little better than no reasons at all. Second, ceding empirical discovery to computers has the potential to stymie human progress. Suppose a Twitter-trained algorithm learned to predict users' crimes before they happened, relying on incomprehensible shifts in

347. See Mori & Uchihira, *supra* note 268, at 780.

348. See Jenna Burrell, *How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms*, 3 *BIG DATA & SOC’Y* 1, 4–9 (Jan.–June 2016), <https://journals.sagepub.com/doi/pdf/10.1177/2053951715622512> [<https://perma.cc/DRM8-KL9Y>].

349. See, e.g., *id.* at 5–6.

350. See *id.* at 4–9.

351. See *id.* at 10.

their patterns of likes and retweets.³⁵² If we relied on such insights blindly and indefinitely, we would do so to our detriment.³⁵³ The Twitter bot might help us to prevent many crimes, but it would shed no light whatsoever on the underlying social ills that produced them. Insofar as algorithms allow us to understand the “what,” but not the “why,” we incur an “intellectual debt,” which will inevitably come due.³⁵⁴

The structure of AI class actions again minimizes these concerns, as compared with other machine learning applications. Begin with reason-giving and legitimacy. In certain contexts, we demand reasons from legal decision-makers. Police must explain the set of facts that, in their minds, gave rise to probable cause and authorized a search.³⁵⁵ Administrative agencies must publish proposed rules, review comments on them, respond to those comments, and provide reasons justifying their final rules.³⁵⁶ Mandatory reason-giving of this kind serves several values. It enables oversight of decision-makers, allowing us to catch and remedy errors—unintentional or malicious—even in otherwise-accurate systems.³⁵⁷ It also enables another kind of oversight, ensuring that decision-makers had valid reasons *at the time*, rather than concocting post-hoc rationalizations.³⁵⁸

AI class actions can protect these values at least as well as the presumptive alternative of individual litigation. Again, algorithms would be trained using actual jury decisions in the case. Those jury decisions would be subject to all of the normal checks of civil litigation. Evidence would be admitted or excluded.³⁵⁹ Juries would be instructed.³⁶⁰ Pre- and post-verdict motions for judgment as a matter of law would be heard.³⁶¹ Such procedures reduce error and narrow the evidence—and thus the contemporaneous reasons—on which the jury can rely.

352. Cf. Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> [<https://perma.cc/GR5C-4SFW>] (describing Target’s ability to predict pregnancy when shoppers started buying certain vitamins and baby-friendly products). The hypothetical Twitter variation makes the inference more abstract. Twitter is surely on the lookout for inferences like these, valuable as they would be to the company’s ad-revenue-based business model.

353. See Kiel Brennan-Marquez, “*Plausible Cause*”: *Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1296–97 (2017).

354. Jonathan Zittrain, *The Hidden Costs of Automated Thinking*, NEW YORKER (July 23, 2019), <https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking> [<https://perma.cc/45ZR-7U3U>].

355. See Brennan-Marquez, *supra* note 353, at 1253.

356. Administrative Procedure Act § 4, 5 U.S.C. § 553.

357. See Brennan-Marquez, *supra* note 353, at 1255 & n.18.

358. See *id.* at 1291–92.

359. See FED. R. EVID. 103(a).

360. See FED. R. CIV. P. 51(b)(3).

361. See FED. R. CIV. P. 50(a)(1)(B).

Such ordinary procedural safeguards of litigation would likewise apply to the other information that the algorithms would consume. Data constituting input features would be subject to ordinary discovery and evidentiary rules.³⁶² And, as described above, the adversarial process of algorithmic design would present opportunities for the parties to stamp out nefarious inaccuracies in both training data and the inputs used to make class-wide decisions.³⁶³ Moreover, in weak AI class actions, every algorithmic determination would be subject to challenge. And if challenged, the presumption would go to ordinary litigation. There again, all of the normal rules would apply.

Yet, as already discussed, there is no guarantee that juries will not act with secret, forbidden intentions.³⁶⁴ Judicial oversight usually cannot, for example, police a jury's undisclosed racist motives, assuming the evidence puts its ultimate decision within the realm of the "reasonable."³⁶⁵ And if juries are motivated by undisclosed invidious animus, algorithms trained on their decisions will mimic it.

But here we find a limit to our demands for reason-giving, even in normal, non-algorithmic litigation. Jury deliberations are generally protected as secret, at least until the end of the case.³⁶⁶ And even when they *can* be disclosed, the law almost never requires that they *must* be.³⁶⁷ Thus, jury decisions represent an area of legal decision-making affirmatively protected from mandatory reason-giving. This likely represents a trade-off, sacrificing some oversight for the sake of free-thinking candor in the deliberation room. This trade-off is not without its critics—including some who have questioned whether it threatens the legitimacy of certain jury decisions.³⁶⁸ But it is nevertheless the law of the land.

Thus, from the perspective of error correction and oversight, AI class actions—especially in their weak form—fare no worse than the alternative. They are subject to at least the same protections as ordinary litigation. In fact, AI class actions might fare somewhat better. As described above, the only part of an algorithmic decision that constitutes a black box is the causal story underpinning the links between inputs and outputs. The inputs themselves—along with the design of the algorithm—are open to inspection and challenge in the adversarial process. And unlike a jury that hears inadmissible

362. See FED. R. CIV. P. 26(b)(1); FED. R. EVID. 402.

363. See *supra* subpart IV(A).

364. See *supra* text accompanying notes 346–47.

365. See *supra* text accompanying notes 346–47.

366. See FED. R. EVID. 606(b).

367. See *id.*

368. See generally Robert I. Correales, *Is Peña-Rodriguez v. Colorado Just a Drop in the Bucket or a Catalyst for Improving a Jury System Still Plagued by Racial Bias, and Still Badly in Need of Repairs?*, 21 HARV. LATINX L. REV. 1 (2018).

testimony and is instructed to ignore it, algorithms can be trusted to consider only information that the court ultimately approves.

What about the second aspect of the black box problem—intellectual debt? The answer here is straightforward. Unlike in the example of the Twitter crime predictor, AI class actions do not employ algorithms to generate inferences that humans cannot. Just the opposite. The purpose of an algorithm in an AI class action is to replicate a jury's inexorably *human* decisions. The algorithm's results are thus in thrall to human judgment, not the other way around.³⁶⁹

Thus, while AI class actions would, in some sense, outsource human thinking to robots, they would not fundamentally outsource human *understanding*. In contrast with other Big Data applications, the goal of AI class actions is not to uncover—and blindly rely on—new, incomprehensible connections in underlying data. It is instead to automate the process of drawing the connections at which human minds are adept.

Conclusion

By most accounts, *Wal-Mart* threw a serious roadblock in the path of those seeking to achieve class certification in traditionally hard-to-certify cases by using creative statistical proof. But as this Article has shown, *Wal-Mart* and its progeny do not flatly forbid such new, ambitious approaches. Instead, they demand only that new methods of statistical proof are accurate at the level of individual class members. AI class actions can satisfy that criterion. Advanced machine learning algorithms can determine, with high accuracy, whether *individual* class members can establish crucial elements of liability and, if so, what their damages should be. Thus, machine learning represents a path to class certification across a wide swath of important, but currently uncertifiable, categories of claims. Using it, countless individual litigants with valid, but currently unvindicable, claims could access the court system and, ultimately, secure the justice they are presently denied.

369. It is—in principle—possible that algorithms in AI class actions could be trained using apparently irrelevant evidence, resulting in incomprehensible connections. But at a first cut, evidentiary and discovery rules requiring relevance would keep such information out of the litigation and, thus, out of the training data. More importantly, the parties litigating algorithmic design would have incentives to avoid inducing inferences from apparently irrelevant data. Under the weak AI class action proposal, algorithmic determinations that were incomprehensible, given the input evidence, would be ripe for challenge. They would certainly be reversed in ordinary litigation.