

Texas Law Review Online

Volume 100

Response

Personalized Class Actions

Omri Ben-Shahar[†]

Algorithms and artificial intelligence (AI) are changing our world in profound ways. They are introducing previously unthinkable products and services and affecting our daily lives in many ways. The AI revolution is taking center stage also in law, but among legal academics, the primary interest is not how to introduce robotic methods to law, especially where they can improve upon imperfect, biased, and discriminatory human decisions. Rather, much of the focus among legal academics is alarmist: how to pull the brakes on the looming “takeover” by machines—the so-called algorithms of oppression and weapons of math destruction—and how to reform the law so as to establish limits and guarantees against misuse of algorithms.¹

No doubt important, the cautionary agenda should not suffocate experimentation with AI methods in law, especially in areas where Big Data and machine-learning methods can replace highly imperfect rules and biased human discretion.² To that end, legal scholars could be at the forefront of a creative agenda, imagining and designing legal frameworks that would be fueled by data and advised by algorithms.³ Peter Salib’s article, *Artificially*

[†] University of Chicago Law School

1. See generally, CATHY O’NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); MEREDITH BROUSSARD, ARTIFICIAL UNINTELLIGENCE: HOW COMPUTERS MISUNDERSTAND THE WORLD (2018); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2017); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018).

2. See, e.g., Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 240–42 (2018).

3. See, e.g., Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 114 (2018); Crystal S. Yang

Intelligent Class Actions,⁴ is a striking exemplar of that creative agenda. Salib is offering a novel application of AI methods in law—and not a science fiction one, but rather a specific tool that would solve one of the biggest problems in protective law: how to allow people with similar but different claims to band together in class actions. Specifically, Salib considers how to use statistical tools within the parameters permitted by law so that the differences between the class members will not eclipse the similarities.

A fundamental tradeoff in civil litigation is between accuracy and litigation costs: how to design cost-justified procedures for effective vindication of uncertain legal rights. Almost every doctrine of civil procedure represents some balance between the two goals, but none with greater stakes than those governing class actions.⁵ When numerous parties have claims that are sufficiently similar, the cost of pursuing each claim individually can be saved by aggregating them into a single representative suit while applying a single one-size-fits-all result to all claims.

The one-size-fits-all outcome of group litigation sacrifices accuracy, since the joined claims are not identical. But it saves litigation costs in avoiding relitigating identical questions and prevents an even greater inaccuracy if otherwise the individual claims would be too costly to pursue. The more the cases are alike, the greater the benefit of aggregation and the lesser the accuracy sacrifice. Much of Rule 23 of the *Federal Rules of Civil Procedure* conceptualizes what it means for cases to be alike, guaranteeing that the upside of reduced litigation costs is not overwhelmed by the downside of inaccuracy.⁶

Using statistical tools in litigation heightens this accuracy-versus-cost-saving dilemma. Statistics help characterize the distribution of cases and treat each case along some measure of typicality. This can be done relatively cheaply and efficiently and then scaled up to a large plaintiff group at little cost. But it ordinarily draws out average characteristics, average injuries, or other synthetic midrange values, which means that it ignores (averages out) factors unique to each case, sometimes viewed as “noise.” With statistical tools, uncertainty over the unique merits of specific claims is not resolved but rather addressed by referring to that midrange value of the distribution. True, under the law of large numbers, this exercise guarantees that the expected

& Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 MICH. L. REV. 291, 297 (2021).

4. Peter N. Salib, *Artificially Intelligent Class Actions*, 100 TEXAS L. REV. 519 (2022).

5. See Hillel J. Bavli, *Aggregating for Accuracy: A Closer Look at Sampling and Accuracy in Class Action Litigation*, 14 L. PROBABILITY & RISK 67, 70 (2015).

6. FED. R. CIV. P. 23; see Salib, *supra* note 4, at 520–21, 521 n.7.

error for the entire class is small and could even come close to zero.⁷ Mismeasurement for one member of the class is offset by an opposite-direction mismeasurement for another member. The accuracy thus achieved is at the level of the group, guaranteeing the defendant pays the right aggregate amount, equal to the total overall harm caused to all victims. But accuracy is not achieved at the level of the individual. Some members of the plaintiffs' class receive more than their injuries merit; others receive less.

Accordingly, while statistical methods in aggregate litigation satisfy the "defendant-accuracy" criterion—setting the magnitude of aggregate liability equal to the magnitude of aggregate harm—they fail the "plaintiff-accuracy" criterion. This, according to the Supreme Court's *Wal-Mart v. Dukes*⁸ case, is a fatal flaw.⁹ The defendant Wal-Mart, the Court held, had a substantive right to insist on "individualized determinations of each employee's eligibility for backpay."¹⁰ Because the harm to individual members of the class could have varied store by store and because plaintiffs did not establish in their statistical methods the store-by-store disparity (or other fine-grained differential effects), the statistical method failed the plaintiff-accuracy test.¹¹

An aside: call me simple-minded, but I find it mystifying why plaintiff accuracy matters so critically to the Supreme Court (and in a class action context of all places). The defendant, it seems, is shedding crocodile tears—with surprising success—in complaining that while it will be charged with the accurate measure of aggregate liability, the statistical tools will miss out on plaintiff accuracy. Once an exacting standard of defendant accuracy is policed by the court, and if plaintiffs freely elect to forgo plaintiff accuracy when they band together for redress, it is a bit rich to stand up and righteously agonize that individual plaintiffs might not be accurately compensated. Particularly since the denial of class certification has the pragmatic effect of no recovery, which ends up aggravating both sides of the inaccuracy. Surely, it is better that plaintiffs will get a rough estimate of their harm than the wrong measure of zero.

Be that as it may, the *Wal-Mart* precedent requires class aggregation to satisfy plaintiff accuracy, and unfortunately statistical methods like regression analysis that look at national trends without sufficient granularity

7. See generally LARRY C. ANDREWS & RONALD L. PHILLIPS, FIELD GUIDE TO PROBABILITY, RANDOM PROCESSES, AND RANDOM DATA ANALYSIS 14 (2012) (defining the law of large numbers).

8. 564 U.S. 338 (2011).

9. *Id.* at 366.

10. *Id.*

11. See *id.* at 357 ("A regional pay disparity, for example, may be attributable to only a small set of Wal-Mart stores[] and cannot by itself establish the uniform, store-by-store disparity upon which the plaintiffs' theory of commonality depends.").

cannot assure the satisfaction of this test.¹² While individuals will be compensated correctly on average, it is quite possible and even likely that a great many individual awards will be misaligned with those individuals' actual injury.

Enter Peter Salib with an ingenious solution. Instead of statistical tools that aim at estimating average impact on a protected group (such as sex or race), an algorithm will be trained to *predict* the merits of each individual claim based on a sample set of substantively litigated cases. The goal is to award each class member an estimate of individual redress based not on the average of the overall group but instead on a multitude of factors that characterize each individual case. If this method could be operationalized, it would solve the *Wal-Mart* concern.

What we need, Salib recognizes, is a method to train an algorithm to predict the differential remedies each class member would have received in an individual suit. For example, in a wrongful death suit for lost income, each member of the class would not receive the average value of loss. Instead, each member would receive a pinpointed estimate of individualized predicted income based on a host of factors that a trained algorithm would recognize to correlate with an individual's future income prospects. In an employment discrimination suit, each member of the class will receive compensation that reflects a set of relevant variables that determine actual loss, telling us what each person with certain skills and attributes would have earned but for the discrimination.

How would an algorithm generate such individualized estimates? The key to fulfilling the plaintiff-accuracy criterion is to provide the algorithm sufficient training data. According to Salib's plan, the algorithm would be fed information about how a sample set of prior cases involving class members, each individually litigated in a pilot phase, were decided. The algorithm does not need to code the reasons for each decision in the sample set, but it can merely scan the information in the file of each. The information about each case would include a list of inputs—facts in the file that describe the plaintiff and the plaintiff's circumstances which can be reduced to a vector of quantitative measures—and outputs measuring the damages outcome. With a sufficiently large sample set, the algorithm would identify inputs of a case that are correlated with the magnitude of the individual compensatory award.

In an employment discrimination case, for example, the algorithm might detect those factors determining each plaintiff's size of recovery, which could include some obvious inputs like age, education, work history, or

12. See *id.* at 356–57; Salib, *supra* note 4, at 535–37 (discussing the *Wal-Mart* Court's disapproval of the respondents' statistical methods and arguing that this disapproval was based on accuracy concerns).

productivity and some less obvious ones such as height or family status.¹³ The method could also detect—and this would be key to the litigation—correlations between the compensatory award and prohibited factors. These correlations would no longer have the problems of statistical evidence à la *Wal-Mart*. The algorithm might “notice” that not all women plaintiffs were successful in their discrimination suits, identifying additional factors that identify the subset among those plaintiffs who prevailed. For example, only women in certain positions, or in certain geographic locations, or of certain age groups were successful. And even when these women were discriminated against, the wage reduction due to discrimination would be measured in a more accurate and attenuated manner, excluding the wage effect of other interacting individual factors.

Salib argues that using an algorithm thus trained via a sample set of pilot lawsuits would resolve the plaintiff-accuracy concern of *Wal-Mart* and that this method of estimation can be practically designed.¹⁴ I agree on the first claim and can be further persuaded on the second. The potential accuracy of predictive algorithms is a premise on which many collective activities dealing with pools of people already rely—e.g., insurance, credit markets, employment hiring, and, of course, digital services—and Salib is more than entitled to assume that it could potentially be harnessed in the litigation pool. While each class action would no doubt require an ambitious training phase whereby a sufficiently large sample set of cases would have to go through individual litigation generating sufficient raw data on which an algorithm would be trained and tested for accuracy, there are enough big disputes with massive class sizes to justify such sample sets and to make this tool get off the ground.¹⁵

If Salib is right that his algorithmic prediction method passes the *Wal-Mart* accuracy test—namely, that courts would accept algorithmic predictions of individual case merit based on data from prior cases—I want to think out loud whether the method can be further improved. Is a sample set that includes dozens (if not more) of individually litigated cases necessary? Can we do without this cumbersome detour? I want to suggest

13. *Cf. id.* at 545 (explaining that algorithms designed for loan issuance or candidate screening might use “resumé, job history, income, credit score,” among other factors, “to mimic interview-informed human decisions”). Similarly, in medical-causation cases, inputs might include medical and employment records as well as “questionnaire-based testimony—made under oath—from plaintiffs.” *Id.* at 546.

14. *Id.* at 524, 547–48.

15. *See, e.g.,* Edward Segal, *Workplace Class Action Settlements Set New Record In 2021: Report*, FORBES (Jan. 4, 2022, 5:00 AM), <https://www.forbes.com/sites/edwardsegal/2022/01/04/workplace-class-action-settlements-set-new-record-in-2021-report/?sh=7197d57040b8> [https://perma.cc/GR5C-4SFW] (discussing a record number of 1,607 class action rulings in 2021 and a record amount of settlements at \$3.62 billion).

that the accuracy achieved via litigation by prediction in the individual level could be obtained without Salib's pilot phase. Instead, a class action could rely on statistical analysis based on the very same factors that are substantively relevant in the individual cases and that become the inputs of Salib's algorithm. Every class action could accordingly be resolved by aggregating the procedure but disaggregating the remedies members of the class are entitled to, each according to their idiosyncratic traits as identified by the already trained algorithm. Indeed, the approach can be applied to all litigation—not only class actions. If we trust the accuracy of algorithms or regression methods in estimating and predicting individual merits, why not put this to use in single-plaintiff cases?

But let us focus on class actions and think whether they can reach plaintiff accuracy without Salib's pilot phase of individual lawsuits. Remember, the goal is to move away from averages—from giving all members of the class a uniform, midvalue redress. Some deserve more; others deserve less. In Salib's sample set, separate courts will figure out who deserves what, and their findings will be used to train an algorithm to figure out who gets how much when they individually litigate and then replicate this formula for other class members. But on what basis will the separate courts in the sample set decide each individual case? In Title VII litigation involving employment sex discrimination, particularly when the claim is for discrimination in pay, statistical proof is allowed to show disparate impact so long as the impact is correlated with the employer's policy.¹⁶ In principle, each case would have to establish a case-specific benchmark—the wage paid by this employer to other (male) workers with attributes similar to the plaintiff's (e.g., education, experience, performance scores). The court would then compare such a case-specific benchmark with the wage actually paid to the individual female claimant in the case and award redress equal to the measured gap. The case-specific wage benchmark would of course be different across plaintiffs in the sample set—this variation was the premise of the *Wal-Mart* Court in rejecting class certification and requiring individual litigation.¹⁷ The case-specific wage benchmark would accurately reflect variation in pay metrics across the different geographic regions and locations in which the defendant/employer operates, across the different departments of operation, across different periods of time, or across other circumstances relevant to the outcome in the labor market. In the individual lawsuits, taking such variations into account in setting the case-specific benchmark guarantees that the resulting damages would be plaintiff accurate.

Does this methodology have to be performed anew in every case? Do courts have to distill the plaintiff-specific benchmark case by case? I suspect

16. *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 356–59 (2011).

17. *Id.* at 357.

that Salib would agree that it does not. In the sample set, the first few courts would develop a general substantive method to ascertain the wage benchmark for the plaintiff in front of them, but soon courts that follow will borrow the method from earlier cases and apply it to subsequent individual plaintiffs. They will do so because the *method* is general even if its application is case specific. The method tells us which factors determine the pay an employee is entitled to.

Taking this logic to the limit, a sample set would not be needed at all. The substantive inquiry early courts in this set are expected to make for subsequent courts to follow could be pursued in a class action from the get-go without a sample set. For each plaintiff in a class of all women employees, a personalized plaintiff-specific wage benchmark would be computed—not a single uniform benchmark for all, but a *formula* that calculates how the benchmark depends on workers' characteristics. From this formula a personalized wage figure would be derived for each member of the class. This personalized benchmark figure would be identical to the case-specific benchmark tailored to each claimant based on case-by-case sample set litigation because it would be computed from the same data answering the same question: what wage is typically paid by the employer to other male workers with attributes like those of the plaintiff?

Notice that while this method focuses on the difference in pay between men and women, it does not average it out. Instead, for each woman in the class, a different pay gap would be estimated. For some, there will be no gap *vis-à-vis* equally qualified men; for others, the gap might be substantial. The distribution of gaps across women in the class will be the same distribution that would come out from separate individual actions.

Accordingly, when the merits of each case are determined by statistical evidence, as often happens in wage-discrimination suits, there really is no need for a sample set of individually litigated cases, and therefore there is also no need to train an algorithm to predict case outcomes based on such a sample. The *Wal-Mart* Court's plaintiff-accuracy concern can be resolved by a shortcut: a statistical method that is careful not to average the award class members receive. In the manner described, each member of the class is awarded redress in relation to her personalized, algorithm-derived, target-wage benchmark figure.

Specifically, the plaintiffs in the *Wal-Mart* case failed because they were equipped with a statistical method that was too crude. They controlled for some factors that explain wage variation, but not all. Had they controlled for a more comprehensive list of regressors (e.g., different Wal-Mart stores)—or, ideally, for all correlated inputs—the objection by the Court would have likely subsided. When you run the regression model in such a manner, plaintiff accuracy increases.

This example of a wage-discrimination suit and how it can dispense with the need for a sample set of individual cases may be misleading because it is a special case. Its logic is specific to claims that already are natural candidates for statistical evidence at the individual dispute level. Discrimination lawsuits are often such cases—a persuasive way to show that a complainant was discriminated against is by statistically comparing that complainant to others. Because the evidence in any individual case involves data on comparables, my intervention here is to suggest that such data could also be used in the class proceedings, and if analyzed with sufficient depth, it could yield the same results that Salib sought to generate by a sample set of individual litigation and a subsequent algorithmic mining of it.

But what about the multitude of other cases in which the merits are adjudicated on the basis of nonstatistical types of evidence? Consider Salib's example of a mass tort class action where plaintiffs who used a product allege that it caused cancer.¹⁸ These are usually unsuitable for class certification because individual questions dominate.¹⁹ Here, too, Salib proposes a pilot phase in which a sample set of cases would be individually litigated, and subsequently an algorithm would analyze the results, identify case features that correlate with plaintiff success, and replicate the results for the entire class, differentiating the outcome plaintiff by plaintiff.²⁰ Can we implement the differentiation without this process and without the sample set phase? If the merits of individual claims depend on factors that vary across class members (like duration and intensity of exposure to the carcinogenic product, the type of warning they received, or the presence of other intervening causes of cancer), and if the measure of damages each successful plaintiff receives varies by the gravity of injury (and it too depends on idiosyncratic inputs), an aggregated lawsuit could in principle account for all of these factors. The outcome of the class litigation would not treat all members of the class as an average case and would not reach a uniform result across all plaintiffs. Instead, it would have to generate a formula to specify which inputs of those characterizing individual plaintiffs affect defendant's liability and how to weigh those factors. The formula will then be applied to each class member individually, granting each a personalized remedy.

No doubt, asking a court to follow such a formula is a high order, and the difficulty in constructing one probably explains the reluctance to certify

18. See Salib, *supra* note 4, at 521, 521 n.4, 544; *cf.*, *e.g.*, *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591 (1997).

19. MANUAL FOR COMPLEX LITIGATION (FOURTH) § 22.7 (2004) (“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) (citing *In re Initial Pub. Offerings Sec. Litig.*, 471 F.3d 24, 43 (2d Cir. 2006)).

20. Salib, *supra* note 4, at 544–47.

such personal injury classes.²¹ A court would have to account for factors that are observed and weighed by judges and juries in individual proceedings based on specific evidence without seeing that evidence. If these factors are to be relied upon in a class action decision, the court will have to develop an explicit formula, and courts are not accustomed to doing this. They will have to allow the plaintiffs to develop and introduce algorithms that analyze prior case law and derive the formula that reflects precedent. Indeed, existing algorithms developed to predict court outcomes and offer legal advice (which Salib invokes to motivate his own AI method) do just that: they identify the factors that are correlated with prior case outcomes.²² Even with such a formula, applying it to individual plaintiffs would be challenging, as the features that characterize each are not as handy as in Title VII suits. But if these features can be distilled from a sample set of prior suits, they could perhaps be predicted based on prior case law.

Consider another example—a class action in which consumers seek damages for a wrongful act, either tort or breach of contract, but vary widely in their consequential harms. A class action procedure that awards each member of the class a uniform sum of money would violate plaintiff accuracy. Salib’s solution would involve a sample set of individual cases, each seeking proof of the individual plaintiff’s harm and awarding plaintiff-accurate damages. Here too, the distribution of damage remedies could be created in a class proceeding so long as the factors that determine individuals’ entitlement to remedies are systematic. For example, harm may be correlated with age, income, education, occupation, or specific physical attributes. These factors will have to be proven in the class action and then applied to all members.

In sum, the key to Salib’s approach is the idea that the merits of individual cases and the magnitude of damages each plaintiff is likely to receive depend on factors that can be teased out *systematically* from prior court decisions. If so, they could also be teased out in a class proceeding. Instead of asking different courts or juries in the sample set time and again to redo this exercise of weighing the relevant inputs and identifying those that matter for liability, the effort can be concentrated with more statistical effort in the class action.

21. See *Amchem*, 521 U.S. at 609–10.

22. See, e.g., Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1331–32 (2019); Rebecca Crootoof, “Cyborg Justice” and the Risk of Technological-Legal Lock-In, 119 COLUM. L. REV. F. 233, 233–34 (2019); Bernard Marr, *How AI and Machine Learning are Transforming Law Firms and the Legal Sector*, FORBES (May 23, 2018, 12:29 AM), <https://www.forbes.com/sites/bernardmarr/2018/05/23/how-ai-and-machine-learning-are-transforming-law-firms-and-the-legal-sector/?sh=4a5e7232c388> [https://perma.cc/6VRG-TUR7]; ANTHONY J. CASEY & ANTHONY NIBLETT, MIT COMPUTATIONAL L. REP., WILL ROBOT JUDGES CHANGE LITIGATION AND SETTLEMENT OUTCOMES? 2 (2020).

Salib recognizes a version of this critique. In his discussion of settlements, he notes the following:

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 individual questions at hand. 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.²³

True, in settlement negotiations plaintiff accuracy can be waived. Using predictive tools from prior case law may be instructive, but it does not give a sufficiently accurate assessment of the varying strengths of the class members' claims. If this estimation is compounded with evidence specific to the pending liability issues, greater accuracy would be achieved. This is why predictions based on prior case law may not be useful to overcome *Wal-Mart's* accuracy problem. A new algorithm involving the practices of each defendant is needed for each case to determine the weights of the various inputs. In developing this litigation-specific algorithm, we could hopefully do without a sample set of individual cases.

In the end, Salib opens our eyes to a novel way to overcome *Wal-Mart's* objection to statistics. If *Wal-Mart's* command is to stay away from the averaging of claims—namely, to reach a sufficient level of plaintiff accuracy—we need statistical methods that are trained to predict *individual* merits. The specific algorithm Salib has in mind, trained by a sample set of individual lawsuits, may not be the only (or even the most elegant) statistical tool for this estimation. But this is a technicality, and it can be refined over time. The breakthrough in Salib's article is the qualitative insight: a new understanding of the potential use of statistics in class actions.

23. Salib, *supra* note 4, at 554–55 (footnote omitted).