Algorithmic bias in justice

American courts have, in recent years, turned to algorithmic systems to predict whether a defendant is likely to appear at a subsequent court date, recommending the level at which bail should be set. The idea is to train these systems based on public data such as whether historical defendants who were let out on bail actually showed up at their court date. These systems take into account dimensions such as the charges being levied, the defendant's history, their income level, and much more, in the hopes of yielding outcomes that increase public welfare, for example by reducing jailing rates by 40% with no change in resulting crime rates [5], all while being less biased and more empirically grounded.

Instead, observers have identified patterns of racial bias that in some cases reflect and in others amplify well–documented prejudices in the criminal justice system. Researchers have found bail recommendation systems replicating and exacerbating racial and gender biases — recommending against offering bail to black men disproportionately more than for white men, for example [1]. In some cases, it seems that problems stem from the data that informs models [2]; in others, recommendation systems are merely, as AI researchers say, reflecting a mirror back at our own society, itself steeped in racially prejudicial bias [6, 9].

In this case, the analogical street–level bureaucrat is probably clear: it is the person whose work the algorithm seeks to replicate: the judge. These algorithms are often even trained on judges' prior decisions [5]. However, as street-level bureaucrats, judges have struggled to answer the question of "which defendants secure release before trial?" for most of the 20th century [11]. While constitutional law protects people from "excessive bail", Walker points out that ultimately this decision is left to the discretion of the judge [11]. A judge hears preliminary information about the case, reviews information about the defendant (such as past criminal record, assets, and access to means of travel), and sets bail that should be sufficiently high that a defendant will appear for their court date without being inaccessible.

In this third case study, we observe something new: a street-level bureaucrat interacting with a street-level algorithm. This interaction can be fraught: bureaucrats in the judicial system resist, buffer, and circumvent the algorithmic recommendations, especially as those algorithms attempt to subsume the work of those bureaucrats. Indeed, Christin explores some of the tensions that emerge when algorithms begin to absorb bureaucrats' responsibilities and shift the latitude that bureaucrats enjoyed, finding that bureaucrats work around and subvert these systems as a way of keeping their autonomy through foot-dragging, gaming, and open critique [3]. Veale et al. go further to illustrate some of the ways that designers of algorithmic systems can better support street–level bureaucrats given these and other tensions [10].

Researchers have contributed many valuable insights about bail recommendation algorithms from the perspective of fairness, accountability and transparency (reviewed in [4]); the literature of street–level bureaucracies adds a reminder that each case may involve novel circumstances and deserves thoughtful consideration about which humans in particular are well–equipped to reason. As Lipsky writes, "street–level bureaucrats … *at least [have] to be open to the possibility* that each client presents special circumstances and opportunities that may require fresh thinking and flexible action." [7]. Otherwise, why bother having judges or trials at all? Why not articulate the consequences directly in the law, feed the circumstances of the crime into a predefined legal ruleset (e.g., `{crime: murder}`, `{eyewitness: true}`, `{fingerprints: true}`), and assign whatever conclusion the law's prescriptions yield? Largely the reason that our society insists on the right to a trial is that there may be relevant characteristics that cannot be readily encoded or have not been foreseen in advance.

If street–level algorithms are liable to make errors in marginal and novel situations, it suggests that the problem is not just how to handle biased data, but also how to recognize and handle missing data. Increased training data is insufficient: for important cases at the margin, there will be few or no prior cases. Intersectionality is growing as an area of focus within HCI [8]; intersectionality fundamentally calls attention to the fact that combinations of traits (e.g., being a woman and a person of color) need to be treated as a holistically unique constellation of traits, rather than as some sort of sum of the individual traits. As a matter of probability, each additional dimension in the intersection makes that constellation less likely. While similar cases in the mainstream may well have been seen before by the algorithm, when the case is at the margin, its particular intersection of traits may be completely novel. Adding training data is almost a waste of resources here, as the combination may be so rare that even increasing dataset size tenfold or one hundredfold may only add a single additional instance of that combination.

In practice, this intersectional challenge is one reason why many democracies use a form of case law, allowing an individual to argue to a judge that their circumstances are unique and should be examined uniquely, and with discretion. Many cases are straightforward; however, when they're not, the court system must re–examine the case and the law in this new light. How could an algorithm identify a situation that needs to be treated as a novel interpretation of a policy, as opposed to one that is only a small variation on a theme that has been seen before?

Much of the discussion of judicial bail recommendation algorithms today is focused on the goals of fairness, accountability and transparency, or *FAT*. We argue that this is a necessary, but not sufficient, goal. Even a perfectly fair, transparent, and accountable algorithm will make errors of generalization in marginal or new cases. Algorithmic systems don't reflect on the overall role that they play in the criminal justice system when they make bail recommendations.

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*. 77–91.

[3] Angèle Christin. 2017. Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4, 2 (2017), 2053951717718855. https://doi.org/10.1177/2053951717718855

[4] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv preprint arXiv:1808.00023* (2018).

[5] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. Human decisions and machine predictions. *The quarterly journal of economics* 133, 1 (2017), 237–293.

[6] Anja Lambrecht and Catherine E Tucker. 2018. Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads. (2018).

[7] Michael Lipsky. 1983. *Street-Level Bureaucracy: The Dilemmas of the Individual in Public Service.* Russell Sage Foundation.

[8] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5412–5427.

[9] Jacob Thebault-Spieker, Loren G Terveen, and Brent Hecht. 2015. Avoiding the south side and the suburbs: The geography of mobile crowdsourcing markets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 265–275.

[10] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 440, 14 pages. https://doi.org/10.1145/3173574.3174014

[11] Samuel Walker. 1993. *Taming the system: The control of discretion in criminal justice, 1950-1990.* Oxford University Press on Demand.