Exploring Mutable Characteristics and Discriminatory Perceptions in Justice Systems

Both sophisticated analyses of historical evidence and simple real-world observation indicate that judicial decisions demonstrate systematic racial and gender bias. For example, Republican-appointed federal judges sentence black defendants more severely and female defendants more leniently[1]. Federal judges behave more politically before presidential elections[2], especially for judges residing in states with close races in presidential elections[3]. A judge's political party of appointment can be predicted by the citations they choose to motivate their decisions[4].

In their examination of these issues, Naomi and June's *Fair Shake* probes similar patterns of bias across diverse sectors. They explore how deeply ingrained societal prejudices shape outcomes for marginalized groups and argue that systems often fail to provide equitable treatment to all, even when impartiality is purportedly a guiding principle. Like the judges whose behaviors fluctuate with political pressures, institutions and individuals, as described in *Fair Shake*, often operate on biases rooted in history, culture, and deeply internalized norms. The book presents a tapestry of narratives demonstrating how systemic inequities manifest not only in legal contexts but also in everyday life, perpetuating cycles of disadvantage.

The data on the justice system reveals sustained bias, both in arbitrary decisions that correspond with the identity of the assigned judge[5] (e.g., a judge's racial identity is predictive of the disparities in their sentencing decisions) and also those that correspond with trivialities like whether or not a judge's hometown football team[6] won or lost[7], whether a defendant shares with a judge the same first letter in their name[8], whether it is the litigant's birthday[9], etc. Furthermore, minoritized groups consistently bear the punitive brunt of these deviations[6] from objectivity.

As we grapple with bias in our societies, one controversial yet crucial battleground has emerged: artificial intelligence (AI). The digital world mirrors the analog, and our systems of AI are susceptible to our inherent biases. However, there is hope; unlike humans, AI is fundamentally flexible. It can be reprogrammed and adjusted to mitigate biases, a process far more direct than attempting to reshape deeply ingrained human prejudices.

---

[1] Cohen A., Yang C. Judicial Politics and Sentencing Decisions. https://www.aeaweb.org/articles?id=10.1257/pol.20170329
[2] Berdejo C. and Chen D. Electoral Cycles among US Courts of Appeals Judges. https://www.journals.uchicago.edu/doi/abs/10.1086/696237
[3] Chen D. Priming Ideology? Why do Presidential elections affect US Judges https://users.nber.org/~dlchen/papers/Priming_Ideology_withmodel.pdf
[4] Machine Prediction of Political Polarization from Circuit Court Judgements. Lu W., Reinhart E and Chen D. https://users.nber.org/~dlchen/papers/Machine_Prediction_of_Political_Polarization_from_Circuit_Court_Judgments.pdf
[5] Abrams D., Bertrand M. and Mullainathan S. Do Judges Vary in Their Treatment of Race?. https://www.journals.uchicago.edu/doi/abs/10.1086/666006
[6] Eren O. and Mocan N. Emotional Judges and Unlucky Juveniles. https://www.aeaweb.org/articles?id=10.1257/app.20160390
[7] Loecher M. and Chen D. Mood and the malleability of moral reasoning: the impact of irrelevant factors on judicial decisions. https://users.nber.org/~dlchen/papers/Mood_and_the_Malleability_of_Moral_Reasoning.pdf
[8] Chen D. The judicial superego: Implicit egoism, internalized racism, and prejudice in three million sentencing decisions. https://users.nber.org/~dlchen/papers/Implicit_Egoism_and_Prejudice_Kyklos.pdf
[9] Philippe A. and Chen D. Clash of norms judicial leniency on defendant birthdays. https://users.nber.org/~dlchen/papers/Clash_of_Norms_JEBO.pdf

Using a study on gender attitudes in U.S. Circuit Courts as a case in point, I will illustrate how AI holds the potential to counter biases more efficiently than humans. It can do so by diagnosing bias in a manner that humans cannot.

The study in question uses Natural Language Processing (NLP), a branch of AI, to detect judges' attitudes toward females. The researchers developed a "gender slant" measure to gauge how judges associate men with careers and women with families in their written opinions. This nuanced approach brought to light subtle gender biases not captured by direct ruling analysis.

The unique strength of AI here is the ability to aggregate and analyze enormous amounts of data objectively, far beyond human capacity. NLP provided a quantitative, unambiguous measure of gender slant by examining 380,000 published opinions, something that would be impractical if not impossible for humans to achieve in a comparable timeframe. AI, therefore, holds an exceptional potential for bias detection that goes beyond human capabilities.

Gender slant is a convincing proxy for bias. Female and younger judges display lower gender slant. Having a daughter reduces gender slant. Lower gender slant is also associated with more frequent use of gender-neutral pronoun constructions, such as "he or she". Finally, judges with higher slant tend to express less empathy towards women in their writing.

Judges not only systematically differed in the way they write about gender, these differences are predictive of how they decide gender rights cases and how they treat their female colleagues[10]. The study examines how judges with different gender slant levels interact with female judges in three areas: reversals of lower court decisions, opinion assignments, and citations. Results show that **judges with higher gender slant are more likely to reverse decisions by female district judges, less likely to assign opinion authorship to female judges, and less likely to cite female judges' opinions.** These judges also tend to vote more conservatively in gender-related cases. Findings suggest that gender bias could hinder female judges' career progression and reinforce gender disparity in the judiciary.

The underrepresentation of women at the top of the legal profession is an issue that has received considerable attention in the United States. It's troubling that although nearly 45% of law school graduates since the 1990s have been women, females still account for only 20% of equity partners in large law firms and 30% of state and federal judgeships. The disparities in these numbers speak to a systemic problem: the differential treatment of female judges, possibly due to gender attitudes among their colleagues.

Gender attitudes, or the biases and preconceptions one holds about social groups, notably women and racial minorities, are known to significantly influence judgments and choices. These biases affect decisions in a range of contexts, from physician treatments and hiring decisions to employer-employee interactions and even the effectiveness of teachers. If these attitudes imply

---

[10] Ash E., Chen D. and Ornaghi A. Gender Attitudes in the Judiciary: Evidence from US Circuit Courts.
https://users.nber.org/~dlchen/papers/Gender_Attitudes_in_the_Judiciary_AEJApplied.pdf

differential treatment of female judges, they could be a contributing factor to the underrepresentation of women in the judiciary.

It's challenging to examine these issues among justice actors due to the lack of traditional measures of gender attitudes for judges. However, researchers have innovatively used recent developments in natural language processing (NLP) to propose a novel measure of gender attitudes. By analyzing a large corpus of written text from appellate judges, researchers have developed a measure of gender bias based on how strongly judges associate men with careers and women with families in their writing. Using a technological tool called word embeddings, the researchers calculated a judge-specific gender bias measure.

Unlike humans, where bias can become deeply entrenched over time, AI can be swiftly and effectively adjusted once a bias is detected. There's no need for lengthy educational or societal campaigns to alter AI behavior – a programmer can do it with a few lines of code.

Besides the text, there is voice. Voice, long considered a neutral channel of expression, is revealed by Chen, Halberstam, and Yu's study to be a surprisingly potent driver of biased decision-making within the U.S. Supreme Court bar. Their work, which delves into the influence of perceived masculinity and femininity in advocates' voices on case outcomes, raises critical questions about implicit bias, systemic inequality, and mutable characteristics that Naomi and June explore with equal vigor in *Fair Shake*. In *Fair Shake*, Naomi and June argue for a reimagined vision of justice and equality—one that moves beyond traditional markers of equity and embraces deeper, more nuanced understandings of how individuals navigate the world through mutable and immutable aspects of their identities.

This essay reflects on the ways vocal traits serve as battlegrounds for broader social biases and how such subtle but pervasive discrimination often denies a "fair shake" to those who diverge from traditional norms. Chen and colleagues' study provides an examination of how voice-based judgments influence Supreme Court outcomes. Their analysis of 1,901 oral arguments from 1998 to 2012 finds that male petitioners perceived as having less masculine voices are significantly more likely to succeed. In contrast, women with more feminine-sounding voices fare better. These patterns indicate that voice, a mutable characteristic, carries weighty consequences in high-stakes legal contexts—ones that ostensibly should be free from prejudicial influence.

One of the most poignant intersections between Chen et al.'s study and *Fair Shake* lies in their shared exploration of "covering"—the demand that individuals downplay aspects of their identity to gain social acceptance. Chen and colleagues note that lawyers may adopt vocal strategies to "cover" their natural speaking patterns to conform to expectations. Naomi and June emphasize that covering extends far beyond the legal arena, shaping how individuals navigate every facet of their lives. The authors recount examples of professionals who consciously alter their speech, behavior, or appearance to fit within rigid social norms, often at great personal cost.

The concept of covering underscores a tension at the heart of both works: assimilation often serves as a prerequisite for success in systems purportedly based on merit. Yet, as Chen et al.

show, lawyers who don't conform to the masculine norm can be punished for being different: individuals misperceive more masculine-sounding lawyers as winners or have a taste for being around masculine-sounding lawyers. Less masculine-sounding men and women are less likely to be hired by the elite Supreme Court bar firms. This pressure to conform can become dehumanizing, stripping individuals of their authenticity and perpetuating a cycle of marginalization.

Workplace practices that prioritize "cultural fit," an often-coded term used to enforce homogeneity and limit true diversity, can limit the opportunities for a fair shake. Just as law firms and judicial actors may inadvertently (or purposefully) penalize lawyers with voices that deviate from perceived norms, so too do workplaces often judge individuals harshly for failing to assimilate fully into dominant workplace cultures. By drawing attention to the stakes involved—whether in the form of a legal ruling or job security—both works highlight how mutable characteristics become barriers to equitable outcomes.

The correlation between voice-based perceptions and Supreme Court outcomes mirrors the discriminatory practices Naomi and June critique in *Fair Shake*. This subtle form of prejudice serves as a reminder that discrimination has evolved into more insidious and less overt forms. In both cases, individuals are judged based on attributes that deviate from the perceived "norm." In the courtroom, as on other societal stages, such judgments create a new class of outsiders and insiders—those who fit and those who do not.

A key reflection emerging from this integration is that bias does not operate solely at the level of conscious prejudice. It can manifest as statistical discrimination—where judgments are made based on assumed characteristics or misperceptions. Chen and colleagues' experiment using incentives and information illustrates that voice-based perceptions can be mitigated when individuals are made more aware of their biases or incentivized to make accurate judgments. Naomi and June similarly call for educational and structural reforms to promote empathy, challenge stereotypes, and create spaces where marginalized voices are heard on their own terms.

Chen et al.'s experimental efforts to debias perceptions through information and incentives offer a glimmer of hope. By making biases explicit and offering incentives to counteract prejudice, they demonstrate that it is possible to attenuate the influence of voice-based judgments. This finding resonates with Naomi and June's call for systemic change through awareness-building, education, and policy reforms aimed at creating more inclusive institutions.

However, both works caution against over-reliance on individual efforts to address structural issues. As Naomi and June emphasize, achieving a "fair shake" requires not just debiasing individuals but transforming the structures that enable discrimination to persist. The study by Chen and colleagues provides a starting point, revealing how mutable characteristics shape outcomes in high-stakes environments. But it also highlights the limits of individual adaptation in the face of entrenched systemic biases.

The intersection of Chen et al.'s study and *Fair Shake* reveals a complex tapestry of identity, perception, and justice. Mutable characteristics like voice may seem superficial, but they carry

profound implications for how individuals are perceived and treated in society's most powerful institutions. For lawyers arguing before the Supreme Court, as for professionals in every field, success often hinges on the ability to navigate a web of biases that privilege conformity over authenticity.

The call to action from both works is clear: true justice requires dismantling the systems that perpetuate inequality, challenging the biases that shape perception, and creating environments where all individuals—regardless of their voices, gender, or other mutable traits—can receive a fair shake. By integrating empirical evidence with personal narratives, Chen et al. and Naomi and June invite us to reflect deeply on the kind of society we wish to build—one where equity is not contingent on conformity but is grounded in respect for each individual's unique voice.

This vision of equity is particularly relevant in the context of emerging technologies like AI, which both mirror, magnify, and mute human intent. The dichotomy of AI and bias highlights a crucial truth: AI is neither inherently good nor bad, but a tool whose impact depends on how we choose to wield it. While it has the potential to amplify existing prejudices, it also offers an unprecedented opportunity to reduce biases and foster greater fairness. The challenge lies in ensuring that AI aligns with the equitable society we aspire to create.

For instance, as we advance from the analysis phase to the application phase in the study on gender attitudes in U.S. Circuit Courts, AI could be utilized to counter the detected gender biases. Systems can be programmed to prompt human judges to self-reflect and reconsider potential implicit biases.

Ultimately, the goal of leveraging AI isn't to replace human judgment but to enhance it, to make us more aware of our inherent biases, and aid us in countering them. Just as a spellchecker alerts us to a misspelled word, AI could alert us to potential bias, pushing us toward a more fair and just society.

AI is only as good or as bad as we allow it to be. As researchers, programmers, and users, we must remain vigilant to the biases we could be unknowingly encoding into our AI systems. Only then can we truly unlock the potential of AI to reduce bias and discrimination, making strides toward a more equitable world and a more inclusive and unbiased environment.