World Development 186 (2025) 106834

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Regular Research Article

The data revolution in justice

Manuel Ramos-Maqueda [a,b,1], Daniel L. Chen [c,d,1,*]

[a] World Bank, USA
[b] Blavatnik School of Government, University of Oxford, United Kingdom
[c] Toulouse School of Economics, France
[d] Institute for Advanced Study in Toulouse, France

A B S T R A C T

This article explores the transformative potential of data science in enhancing justice systems globally. Leveraging the increasing availability of judicial data and the advancements of the digital revolution, this paper demonstrates how policymakers can significantly improve access, efficiency, and fairness within justice systems—crucial components of economic development as discussed in a companion paper (Ramos-Maqueda and Chen, 2024). We introduce a comprehensive framework for evaluating, diagnosing, and experimenting with judicial processes to deepen our understanding of judicial performance using data science methodologies. Key areas of focus include the application of machine learning and "text-as-data" techniques to enhance efficiency and identify disparities in judicial rulings. Through detailed case studies and empirical evidence, we illustrate how these technologies can address systemic shortcomings and drive meaningful reforms. By identifying specific areas where data science can bridge existing gaps, we aim to provide actionable insights for policymakers. Our findings highlight the profound impact of data-driven approaches on fostering a more just society and promoting sustainable economic growth. The paper concludes by suggesting future research directions and practical applications of data science in judicial contexts to ensure continuous improvement and innovation.

1. Introduction

Recent years have witnessed an increase of data availability in judicial systems around the world. Many countries have developed electronic case management and e-filing systems. This has led to the creation of massive databases that track every characteristic of each case. In fact, this is by definition an element of the fair administration of justice: everything that happens in a case needs to be tracked to demonstrate the evidence behind facts-based decisions. However, even though this data is readily available to policymakers, it has rarely been exploited to evaluate policies or improve the functioning of the judiciary.

The underutilization of data raises many questions: What if we could harness its power to understand what sectors of the population lack access to justice? Taking it one step further, what if we could use the text of decisions to reduce judicial biases, to revamp justice systems that unfairly target gender and racial minorities? What if we could use high-frequency data to understand and address why cases get backlogged, thereby removing a main barrier to justice in the developing world? In the era of COVID-19, what if we could use high-frequency data to predict the future congestion of courts, and plan ahead to be prepared for what is yet to come? Ultimately, what if we could use data to assess the specific impact of new laws on individuals and firms, and promote best practices among different countries?

By responding to these questions, the data revolution can help the rule of law reclaim its role in development planning. More importantly, by using data more effectively, judiciaries around the world, and particularly those in developing countries, will be able to improve their performance, address deficits in the quality and accessibility of justice, and contribute to prosperity. In this paper, we explore the opportunities that arise from the expanded data systems in justice. We discuss how data can help assess social prejudice and integrity in justice systems in ways that have not been possible until recently. As COVID-19 moved justice into virtual courtrooms and greater electronic processing of cases, this paper assesses the opportunities that arise to evaluate and improve crucial aspects of the efficiency, quality and integrity of courts.

2. The expansion of data availability

Judiciaries across the globe generate vast amounts of data on a daily basis, traditionally collected in paper form. However, this valuable resource has remained largely underutilized, with limited quantitative analysis and insights derived from the overall patterns within the data. According to the literature, the reasons are twofold: first, judicial

* Corresponding author.
E-mail addresses: mramosmaqueda@worldbank.org (M. Ramos-Maqueda), daniel.chen@iast.fr (D.L. Chen).
[1] The authors contributed equally.

https://doi.org/10.1016/j.worlddev.2024.106834
Accepted 2 November 2024
Available online 18 November 2024

officers have historically not been concerned with performance indicators; and second, legal scholars have prioritized qualitative over quantitative analysis of the functioning of the judiciary (Dakolias, 1999). Nonetheless, recent developments indicate a notable shift in this trend. Judiciaries worldwide are embracing technological advancements and implementing electronic systems such as e-filing and case management platforms, electronic calendars, virtual courts, and electronic jurisprudence. These systems generate vast amounts of data that can be harnessed to evaluate and enhance the functioning of the judiciary. Furthermore, the growing concern over case backlogs in developing countries has sparked an interest in performance metrics to improve judicial efficiency. Thus, the expanding availability of data in judiciaries worldwide presents valuable opportunities to measure overall performance, identify areas for improvement, and evaluate the impact of justice reforms.

The adoption of technological platforms that generate judicial data has happened in both developed and developing countries. The United States, Canada, European countries, and Australia are at the forefront of e-justice systems. Even so, electronic systems have also been implemented or are in the process of implementation in Latin America (Chile, Argentina, Uruguay, Peru, the República Bolivariana de Venezuela, and some Mexican states, among others), Africa (Kenya, Rwanda, Lesotho, etc.) and Asia (Malaysia, Singapore, and many others). The increased availability of judicial data can be used to improve the functioning of judiciaries, define what works in justice reforms, and produce rigorous evidence on how the judiciary may affect development outcomes. That said, this paper does not establish standard performance indicators. Indicators have already been discussed in areas such as access to justice; expedition and timelines; equality, fairness, and integrity; independence and accountability; and public trust and confidence (Reno et al., 1997). Moreover, we do not cover issues of transparency or accessibility of the data. The recommendations put forth are intended for all parties with access to judicial data. It is also recognized that not all countries may possess the institutional capacity or infrastructure to establish data ecosystems. Nonetheless, even countries without existing electronic systems can benefit from understanding the potential benefits that may arise from developing such data systems.

The extent to which each of these countries can utilize the approaches outlined in this paper will depend on the extent to which e-judiciaries are integrated, the sources of data available, as well as the quality and accuracy of this data. Concerns about representativeness and exhaustiveness may arise in contexts where not every case is included in the database, which can disproportionately impact those with limited digital access.

The main sources of data are automated registrars. Oftentimes, these cases can be merged to the judgment text. In an ideal scenario, this data will also be complemented by other sources of data to create a larger ecosystem. Case data can be linked to users or firms through tax or national IDs to evaluate the potential effect of policies or reforms on the affected parties. Case data can also be linked to other records, such as police and prison records, or to geographic areas, to better understand regional access. Finally, the data should be complemented by legal needs or court-user surveys to evaluate satisfaction with the functioning of courts. It is crucial not only to consider the perceptions of those who access courts, but also of those who have a legal need but resort to other methods to resolve it.

While creating a data ecosystem is ideal for analysis, there are at least three main limitations that may impede the utilization of this data by judiciaries and researchers. The first limitation is the difficulty in merging different databases with varying definitions, IDs, or relationships for each case. However, even in cases where data cannot be merged, valuable insights can still be gained by comparing the results of different sources. For instance, comparing the outcomes of legal need surveys in a region to case-level data from courts can provide insights into access to justice in remote areas.

The second limitation is related to human capital. Judicial officers are rarely experts in data analysis, and engineers generally lack the domain expertise necessary to understand the functioning of the law. In courts without sufficient human capital to take advantage of available data, training bureaucrats to learn the necessary skills or hiring those with the required skills may be a valuable long-term investment for improving the functioning of courts. In the meantime, non-governmental organizations, international organizations, or private companies may contribute to filling this gap.

Finally, while these records may often be public, the scale, ease, and speed of access facilitated by digital technologies, coupled with advanced analytical capabilities, can lead to ethical issues that may not have been as salient in the past. Potential concerns include privacy, misinterpretation, consent, and disproportionate impacts.

1. Privacy: Even though these records are often public, their increased accessibility can potentially infringe on the privacy of the individuals involved in the cases. The aggregation and processing of these records could lead to revealing patterns or information that an individual might not wish to be publicly known.
2. Misinterpretation: The potential for misinterpretation of data or misuse of analytics can lead to incorrect or harmful conclusions. This is particularly relevant given the complex nature of judicial records, which often require specialized legal knowledge to interpret correctly.
3. Consent: In traditional research involving human subjects, consent is a fundamental requirement. However, individuals whose information is included in these judicial records might not have had the opportunity to consent to their data being used in this manner.
4. Disproportionate impacts: The availability of judicial records and their use might disproportionately impact certain groups of people, such as those from marginalized communities or those with more encounters with the legal system. This may further exacerbate societal inequities.

Given these potential concerns, it would indeed be important for judiciaries and researchers to consider the ethical implications of their work. This could include discussing potential mitigating strategies, such as anonymizing data, establishing guidelines for the interpretation and use of the data, or setting up ethical review processes. Collaborations between researchers and judiciaries along with data use agreements can also limit some of the risks. In the end, ethical evaluations need to be considered to understand if the benefits of the research outweigh the costs.

3. Measurement, diagnostics and experimentation

This section presents an initial framework for using data science methodologies to improve judicial systems. By leveraging machine learning and advanced data analytics, judicial systems can measure performance, diagnose issues, and experiment with policy options more effectively. These data-driven approaches complement traditional experimental and quasi-experimental methods, offering a more comprehensive toolkit for evaluating judicial performance.

Machine learning can significantly enhance the predictive analytics capabilities of judicial systems by forecasting case outcomes, identifying patterns in case backlogs, and predicting future court congestion. These predictive models enable more effective resource allocation and overall efficiency improvement. Additionally, Natural Language Processing (NLP) techniques can be employed to analyze judicial rulings, detecting biases based on race, gender, or other characteristics, which provides valuable insights into systemic issues and helps guide targeted interventions. Data visualization tools, such as dashboards and interactive charts, play a crucial role in presenting complex judicial data in an accessible format, allowing policymakers and stakeholders to track performance metrics and identify trends more easily.

Experimentation with data science methodologies, such as

algorithmic randomization, can simulate random assignment in quasi-experimental designs, creating control groups for evaluating the impact of judicial reforms. Furthermore, data-driven policy simulations can be used to analyze historical data and run simulations, enabling policymakers to predict the potential impact of policy changes before implementation. These approaches ensure that decisions are based on robust data analysis, thereby enhancing the effectiveness and fairness of judicial systems.

Courts should then provide diagnostics on those areas that are most valuable to their service. For example, databases can be used to identify the main causes for adjournments, understand what the barriers to justice are, or evaluate which cases would benefit the most from mediation. Finally, researchers and policymakers should take advantage of data systems to rigorously evaluate the impact of justice reforms. The random assignment of cases to judges behaves as an exogeneous shock to evaluate the causal impact of judicial performance on the well-being of citizens and firms. By keeping track of innovations in the judiciary and experimenting with potential solutions, data systems generate the opportunity to understand the impact of reforms on people's lives and the prosperity of nations.

3.1. Measurement and diagnostics

Despite the growing availability of case-level data, most courts around the world are underutilizing the opportunities created by these databases. Many courts are not even producing performance or management reports that summarize key indicators, such as the number of incoming cases, time to disposition, and case clearance rate. Correspondingly, most courts allocate resources without taking into account important efficiency measures that may be gleaned from stronger empirical data, for example by evaluating the backlog or productivity of specific courts. As data systems become available and more courts decide to catalog data systematically, greater opportunities to evaluate a court's efficiency will arise.

The very basic usage of case management systems consists in measuring what happens in courts. To begin with, judiciaries could create indicators that evaluate the efficiency, access, quality and integrity of each court.[2] Court-user surveys can complement this data by providing additional diagnostics of the quality, integrity, and public trust in the judiciary, based on the experience of the user. Databases can also be used to diagnose the main issues and problems that the judiciary faces. In many developing countries, courts are concerned about their large backlog of cases, which limits citizens' access to justice and erodes their trust in formal institutions. Databases can be used to identify the main causes of delay and adjournments in legal cases, allowing courts to take targeted action to address these issues.

Similar diagnostics can be used to evaluate an array of outcomes that are important to judiciaries, such as who uses formal institutions and how they make use of them, the level of user satisfaction, which legal needs in a region remain unfulfilled, and the main barriers to accessing justice and how to address them. Apart from diagnosing problems, data systems should also be used to coordinate and deploy resources in an efficient and effective manner. For example, the assignment of cases to judges can be automated so that they take into account the workload and backlog of cases for each judge, but also to ensure the random assignment of cases to judges. Databases can also be used to evaluate the capacity of judges and mediators. Then governments can train or incentivize those who are less capable or less willing to improve performance. Finally, databases could be used to determine which courts and cases may benefit more from ADR. Advanced systems can detect which cases are more likely to be successfully resolved through mediation and automatically redirect these cases to mediation. By using data to identify the most appropriate cases for ADR, judiciaries can improve efficiency, reduce backlogs, and enhance access to justice. The literature on bureaucratic management in judicial systems emphasizes the critical role of court management practices, such as clear timelines for processing cases and robust monitoring mechanisms, in enhancing court performance. The article "An Analysis of the Causes of Corruption in the Judiciary" by Buscaglia and Dakolias (1996), prepared by the World Bank's Legal and Judicial Reform Unit, scrutinizes the systemic corruption within judicial systems, particularly in Latin America, exploring its economic and structural roots. It emphasizes the severe social and economic repercussions of endemic corruption which obstructs reform and deepens inequality. Through an empirical model, it identifies low wages, poor monitoring, and insufficient punitive measures as principal catalysts of corruption, supported by data from surveys of first-instance judges across Latin America. The findings suggest that reducing rent-seeking opportunities, increasing transparency, and integrating technologies like computer systems and alternative dispute resolution mechanisms can significantly decrease corruption levels. The paper also addresses the challenges of institutional inertia in judicial reform, proposing strategies to align immediate benefits for court personnel with long-term societal improvements. Finally, it highlights the World Bank's role in global anti-corruption efforts, developing diagnostic tools, building consensus for reform, and implementing measures to combat corruption effectively.

The specific feature of the article was that it used a variable in their regressions: 'One of these variables measures the number of computer systems used by each court. The computer systems provided the following five functions: jurisprudence database; backlogand case tracking network; word processor for judgments; cash-flow accounting, as monitored by external auditors within the judiciary; and network of professional and financial information for each member of the court's personnel. The lack of computer systems is considered by many Latin American lawyers and judges to be the main cause of the inconsistencies found in the application of jurisprudence and of the lack of judicial monitoring of the courts. These inconsistencies support the perception that decisions are arbitrary: coupled with the lack of internal monitoring, they give the judge room to make substantive discretionary decisions, and create an environment that enables corrupt practices to emerge and become more feasible.'

Pioneers like Buscaglia and Dakolias (1996) underscored the importance of technology in this realm, specifically discussing how computer systems within courts could support data-driven decision-making. These systems improve case tracking and reduce inconsistencies in the application of jurisprudence, thereby streamlining case flow management and staffing adequacy. Further, Dakolias and Said (2000) research on a bottom-up approach to judicial reforms provides valuable insights. Findings from pilot projects in Colombia, Peru, Argentina, and Ukraine indicate that reforms are most effective when implemented at the lower court level. The success of these reforms is significantly enhanced by a supportive change in culture and management practices and by garnering backing from various stakeholders, including legal professionals, businesses, and NGOs. This multi-faceted support helps to broaden the impact of judicial reforms.

Databases may also be used to diagnose the overall functioning of the system and the capacity of different legal actors involved. For example, Carmignani and Giacomelli (2010) argue that the number of lawyers has a large positive effect on civil litigation across Italian provinces from 2000 to 2005. This conclusion is challenged by (Yeung and Azevedo, 2011), who find that human or material resources do not fully explain the differences in efficiency across Brazilian courts. Instead, it correlated more closely with court management practices. Mitsopoulos and Pelagidis (2010) suggest that the main problems in the Greek judicial system are actually failings in the system's own design: insufficient judicial organization and accountability, excessively burdensome procedures,

[2] [1]A list of indicators is available at "Court Performance Around the World: A Comparative Perspective," which includes, among others, the number of cases, the case clearance rate, the congestion rate, and the average duration of a case (Dakolias, 1999).

and lack of competition in the provision of legal services. Moreover, Coviello et al. (2015) show that judges carry out work in short, interrupted segments–what they define as task juggling. When judges juggle tasks, it lowers their productivity substantially. In particular, those judges who juggle more trials at once instead of working sequentially on a few of them over time, take longer to complete their portfolios of cases.

These diagnostics may be an essential tool to understand the main areas for improvement in the justice system. The establishment of comprehensive databases and performance indicators is the first step towards understanding what works, or what needs to be improved, in a country's justice system. Policymaking can then benefit from targeted goals and recommendations that are based on solid measurements.

3.2. Experimentation

Apart from measurement and diagnostics, data systems provide the ideal opportunity to experiment with policy options in order to rigorously evaluate the impact of justice reforms. Innovative policies can be implemented in a staggered fashion to provide the opportunity to rigorously evaluate them. Alternatively, performance incentives and other cost-effective measures can be tested via randomized controlled trials (RCTs). For policies that can have a relevant long-term impact, iteratively testing and evaluating their impact through RCTs can lead to a better understanding of the benefits and trade-offs of such policies.

Legal scholars and judges have long made arguments about laws and regulations and justified their arguments with theories about the effects of these legal rules. A particularly challenging dimension of studying the effects of legal rules is that many other aspects of society are correlated with the presence of legal rules, so it is difficult to determine cause or effect. Much like medicine a century ago, prior to clinical trials, theories most often lacked rigorous causal evidence.

Randomizing judicial decisions, however, violates our notion of justice and equal treatment before the law. On the other hand, randomizing case assignment to judges generates a retrospective "clinical trial", the first example of which was only published as recently as 2006 (Kling, 2006), where the policy question was whether longer sentence lengths affected subsequent labor outcomes and earnings. Randomizing sentences is unethical, but randomizing cases to judges who are systematically harsher or more lenient than others generated the inference on the long-run impacts of sentence lengths. Thus, in countries where cases are randomly assigned, the random assignment itself can be used as an exogenous source of variation to evaluate the impact of judicial decisions. This method has become widely used since the credibility revolution.

For instance, Dobbie and Song (2015) investigate the causal effect of consumer bankruptcy–one of the largest social insurance programs in the United States–on debtors by exploiting the random assignment of bankruptcy filings to judges. The authors find that debt relief increases individuals' earnings and employment and decreases mortality and foreclosure rates. In Norway, Dahl et al. (2014)) evaluate the existence and importance of family welfare cultures, where the receipt of a welfare program by one generation causes increased participation in the next generation. The authors exploit the random assignment of judges to applicants for disability insurance whose cases are initially denied, by comparing the implications of being assigned an appeal judge who is systematically more lenient than a harsher one. They find strong evidence for a causal link across generations, where the adult child's participation increases if the parent also participated in the insurance program.

Another policy question that has been answered by leveraging the random assignment of cases to judges is the impact of pre-trial detention on defendants. Arnold et al. (2018) use the detention tendencies of quasi-randomly assigned bail judges to evaluate the impact of pre-trial detention on subsequent defendant outcomes. They find that pre-trial detention–due to comparably harsher judges–decreases formal sector employment and the receipt of government benefits.

This "judge-leniency" design may be applied to other characteristics of judicial decisions. For example, what is the impact of the speed of justice? These kinds of questions can be studied in administrative data where random case assignments exist. Judges predicted to be fast also tend to be fast in other cases. The causal effect of faster case resolution can be studied by linking it to long-term outcomes where the data infrastructure permits.

Overall, experimentation brings the opportunity to implement rigorous causal evidence to the legal realm. Whether it is via RCTs or leveraging the random assignment of cases to judges, "clinical trials" can bring important lessons on why the rule of law matters for development outcomes, and perhaps more importantly, on what is the impact of specific policies and reforms on the litigants' well-being.

As a final note, while researchers obtain rigorous empirical evidence from incremental experimentation, carrying out effective justice reforms may require efforts on a much broader level. Literature has shown that only comprehensive judicial reforms that seek to improve quality, speed, and access all at once, but not limited reforms, increase firm productivity and economic growth on a national level with lasting effect (Chemin, 2020).

4. Machine learning applications

In this section, we discuss avenues to leverage machine learning to improve judicial decision-making and increase the effectiveness of justice. Moreover, we propose a multi-step approach to leverage administrative data to analyze disparities in judicial decisions. We also explore the potential of machine learning and natural language processing (NLP) techniques to process large quantities of text to improve the knowledge and efficiency of justice systems. These techniques currently allow researchers and policymakers to analyze texts to an extent that was not humanly possible before. Such opportunity has arisen not only due to the increase in the large volumes of available data, but also thanks to the recent development of computational tools that make it possible to process and analyze such large and complex data.

4.1. Applications of administrative data

Even though the adoption of machine learning tools in the justice system is in its early stages, the preliminary results showcase the potential of these methods while also highlighting the risks that their inappropriate use may present. Recent research examines the potential of machine learning techniques to "improve human decision", in specific by evaluating the potential of machine learning to decide whether defendants should await trial at home or in jail in different US urban cities (Kleinberg et al., 2018). The authors focus on bail decisions, in which judges have to assess whether the defendant will flee or commit a new crime if released, and trade off these risks against the cost of incarceration. In brief, the judges have to decide, based on their prediction, what a defendant would do if released. This specific prediction task makes an algorithm particularly promising for this job.

The authors use a large data set of cases heard in New York City from 2008 to 2013 to build an algorithm that assesses the risk of fleeing or recidivism. The algorithm then makes a prediction based only on data available to the judges at the time of the bail hearing. After comparing results, the authors find that the algorithm improves judicial decisions, and has the potential to reduce crime by as much as 25 % without changing the number of people waiting in jail. By evaluating the results, they realize that judges release many defendants the algorithm ex-ante identifies as very high risk, and that stricter judges do not jail the riskiest defendants first. These results are not restricted to New York City, as the authors report similar findings in a national data set as well. Moreover, the authors show the potential of machine learning tools to diagnose why judges mispredict.

The results from the aforementioned study suggest a promising avenue for future research on the application of machine learning to

predict the risk of recidivism and improve the precision of judicial decisions. According to the United Nations (2018), the proportion of prisoners held in detention without being sentenced for a crime has remained almost constant between 2003–2005 and 2014–2016, and still affects almost one-third of all prisoners. Thus, the importance of potential developments in this field may be particularly relevant all over the world to better identify those who should stay in prison–or at home–while expecting a trial.

In addition, clear predictions about the length of a trial or expected gains from it may be helpful to improve litigants' decisions on their cases. In Mexico, the rate of settlement in mediation in specialized labor courts remains low because of overconfidence on the client-side and low-quality legal representation (Sadka et al., 2017). Many plaintiffs spend more on legal fees than what they recover in court. Quality of legal aid and predictions about the case are both effective in correcting these erroneous perceptions regarding their entitlement and the importance of particular types of evidence. Sadka et al. (2017) find that providing information to litigants on their predicted outcomes in courts reduces the level of overconfidence of litigants, and nearly doubles the overall settlement rate. Nonetheless, the increase in settlement rates only occurs for the subset of cases for which the plaintiff is present to receive the information, as the information provided to lawyers does not have an effect. Thus, administrative data may also be able to improve litigants' expectations of a case, identify the optimal path to resolve a specific case and—as in the case of Mexican labor courts—lead more litigants to settle, freeing up space in court proceedings.

"Data Science for Justice: Evidence from a Nationwide Randomized Experiment in Kenya" investigates whether data science can improve court efficiency and contribute to economic development (Chemin, 2020). This study explores the causal relationship between judicial efficiency and economic outcomes through a controlled experiment in Kenya. In this experiment, Kenyan courts were randomly assigned to three groups: one received no new information (control), another received detailed analyses of court delays (information group), and the third received the same analyses along with accountability measures where the information was also shared with the public (information and accountability group). The core innovation of the study was the development and use of an algorithm to analyze the first digitized daily court records in Kenya to identify the main causes of court delays. The courts in the information and accountability group showed a 22 % reduction in case duration, suggesting that transparency and accountability significantly enhance judicial efficiency. Further, the study found that in regions with treated courts, there was an increase in formal employment contracts and higher wages, particularly in industries that depend heavily on contracts. This indicates that improving court efficiency can have substantial economic benefits, enhancing both the enforcement of contracts and general economic conditions. This research not only demonstrates the practical impacts of improved judicial efficiency on economic development but also illustrates the potential of data science in transforming public institutions and promoting growth.

Nonetheless, the implementation of machine learning tools in the judicial sector entails relevant risks that may not be overlooked. Inappropriate application of these tools can lead to decisions that may violate due process or that may discriminate based on race, gender or other characteristics. This may be of particularly high risk when the tools represent a black box with no clear explanation of how it works (Rudin and Ustun, 2018). Without appropriate safeguards, letting machines make judicial decisions could amplify existing biases and discrimination.

4.2. Analyzing disparities in justice systems

Equal treatment before the law is an essential feature of democratic societies. According to the Universal Declaration of Human Rights, "All are equal before the law and are entitled without any discrimination to equal protection of the law." As a consequence, everyone must be treated equally under the law regardless of race, gender, color, ethnicity, religion, disability, or other characteristics, without privilege, discrimination or bias. Judges are responsible for following and interpreting the law. Thus, they play a crucial role in ensuring de facto equality of treatment under the law. De jure equality—that is, equality in laws and regulations—will not lead to de facto equality under the law if those responsible for applying the law are still affected by biases along gender, race, or other dimensions of discrimination. On a global scale, despite equal laws, social discrimination, stereotypes, biases, and even ignorance and reluctance continue to undermine the implementation of equal justice (Hyland et al., 2020).

Thus, it is crucial for the formal legal system to enforce equal rights, turning the law into real outcomes. For example, in India, improvement in women's legal rights to land and enforcement by courts increased women's chances to inherit land and leave a violent spouse (Agarwal, 2003; Deininger et al., 2013). In Rwanda, likewise, reforms to the Successions Law increased women's ability to resist polygamy while keeping permanent rights to land (Daley et al., 2010). Despite this principle, there is ample social scientific evidence documenting arbitrariness, unfairness, and discrimination in judicial decision-making. Judges become more politicized before elections and more unified during war (Chen, 2016; Berdejo and Chen, 2017). Politics and race also appear to influence judicial outcomes (Schanzenbach, 2005; Mustard, 2001; Steffensmeier and Demuth, 2000; Albonetti, 1997; Zingraff and Thomson, 1984; Abrams et al., 2012; Boyd et al., 2010), as does masculinity (Chen et al., 2017), defendant birthdays (Chen and Philippe, 2020), football game outcomes (Eren and Mocan, 2018; Chen, 2017), time of day (Chen and Eagel, 2017), name (Chen et al., 2016), and shared biographies (Chen, 2016) or dialects (Chen et al., 2015). These biases affect the quality of decisions and may undermine citizens' confidence in the judiciary.

The overwhelming majority of existing literature on judicial in-group biases concerns the United States (Ash et al., 2021a), but the literature is now extending to other countries, including the Global South. Analyzing the decision of Israeli Arab and Jewish judges during and in the aftermath of the Second Intifada, Shayo and Zussman (2017) found that ethnic biases in local small claims court decisions are positively associated with the intensity of conflict across different localities, and the adverse effects persisted despite the decline of violence. Gazal-Ayal and Sulitzeanu-Kenan (2010) found similar ethnic biases in detention decisions in first bail hearings in Israel. In Kenya, judges hearing criminal appeals are 3–5 percentage points more likely to grant appeals to defendants of the same ethnicity than to other defendants (Choi et al., 2022). In contrast, in India where women, Muslims, and lower castes are underrepresented in the judiciary, Ash et al. (2021a) found no judicial biases in decisions along any one of the three demographic factors.

There are also various papers showing clear judicial biases in laboratory environments. Judges, as humans, are also affected by behavioral biases such as anchoring, framing, hindsight bias, representative heuristics, egocentric bias, snap judgments, and inattention (Guthrie et al., 2001; Rachlinski et al., 2009, 2015). Thus, the primary question is not whether these problematic features of the legal system exist. Rather, the dilemma facing policy-makers is what, if anything, can be done.

Predictive judicial analytics holds the promise of increasing the fairness of the judiciary. Much empirical work observes inconsistencies in judicial behavior. By predicting judicial decisions—with more or less accuracy depending on judicial attributes or case characteristics—staistical analysis and machine learning offer opportunities to detect when judges are most likely to allow extralegal biases to influence their decision-making. Thus, data may be used to understand, and address, judicial biases in decision-making.

As one of the key institutions that enforces social contracts and resolves conflicts, the judicial system plays a vital role in addressing disparities. Based on previous literature, we propose the following framework to measure and analyze disparities in a systematic manner:

1. **Gathering relevant data:** Begin by collecting comprehensive data from multiple sources related to the justice system. Divide the data into three categories: pre-decision data, decision data, and post-decision data. Each category will inform different steps in the framework. To the extent possible, collect data on external factors that may influence judicial decisions, implicit biases of judges and court actors, and the economic and social outcomes of litigants after their judicial case.
2. **Examining Pre-Decision data and initial interactions**: Examine pre-decision data, focusing on the initial encounters with law enforcement. Analyze whether adding control variables, such as race or ethnicity, influences the disparity in these interactions. This analysis identifies potential biases at the earliest stages of the criminal justice process. *Example:* Fryer (2019) assesses whether racial differences in police use of force arise from omitted variables associated with race in the US, showing that Blacks and Hispanics are more likely to experience non-lethal use of force by the police compared to whites, even after taking into account the context and civilian characteristics.
3. **Step 3: Identifying and Analyzing bias in Decision-Makers**
   (a) **Identifying Biased Decision-Makers:** Analyze past behavior of decision-makers to understand if bias is concentrated among specific individuals or widespread. The goal is to identify the aggregate disparity driven by which and how many of the decision-makers. *Example:* Goncalves and Mello (2021) identified racial bias at the police officer–level and found that the entire discrepancy within a unit could be explained by the behavior of just 40 % of the force. In contrast, in Kenya, Chen et al. (2022) find that gender and ethnic in-group biases are mild and widespread across the entire body of judges.
   (b) Understanding Internal Drivers of Biases: Acknowledge and analyze the susceptibility of decision-makers to subconscious stereotypes. Employ Implicit Association Tests (IATs), if available, or Natural Language Processing (NLP) techniques to understand these biases in the written decisions of court actors. Given the textual nature of the law, and the importance of argumentation and reason giving to legal decision-making, there is a substantial amount of textual data that can be used to examine how legally relevant and legally irrelevant factors affect legal outcomes. Recordings of online hearings represent another method to evaluate and understand biases in judicial decisions by, for example, using court recordings and 3D Virtual Reality technology to alter the defendants' race to simulate alternative environments. Examples: Ash and Chen (2018) uses judges' writings to predict the average harshness and racial and sex disparities in sentencing decisions. That work finds that the information contained in written opinions can improve significantly on naive prediction of punitiveness and disparity. Bielen et al. (2021) conducted a study to understand courtroom biases in virtual reality courtrooms. They find that, while evaluators are harsher towards defendants of their own race during the guilt-innocence decision, in the sentencing phase they tend to be more lenient towards defendants of their own race. This pattern leads to significant bias against minorities at all stages. Bielen et al. (2021) did not employ complex data science methodologies but rather relied on straightforward statistical comparisons to analyze the data. Our proposal is that virtual reality courtrooms can be programmed, using generative AI, to debias from the perspective of the perceiver. Multi-modal models already exist for generating life-like movies. We suggest that such approaches can be used to reduce judicial bias.
   (c) **Evaluating the Impact of External Stressors on Decisions:** Examine how unrelated environmental stressors affect judgment. Understand whether certain groups bear a disproportionate impact of these stressors, which could introduce another layer of bias into the decisions. These environmental factors range from weather and time of day to outcomes of football games. The question then is not whether such factors affect judgment, but whether certain groups bear a disproportionate brunt of judges' inattention or their psychological stressors. For example, if disgruntled or tired judges are more likely to punish blacks more harshly than whites because they are more likely to rely on heuristics, it inadvertently adds another layer of bias.
4. **Sequential Analysis of Disparities:** Conduct an in-depth sequential analysis of sentencing decisions, taking into account the defendant's criminal history and other legally permissible characteristics. Aim to pinpoint sources of discrimination that may be embedded in overlooked legal aspects of the case. *Example:* Using linked data across stages on federal cases, Rehavi and Starr (2014) show that blacks receive sentences that are 10 % longer than those of whites with similar history and charged with similar crimes. The primary source of this disparity is the prosecutor's initial decision to file charges carrying mandatory minimum sentences. Blacks are charged with such offenses more often than whites.
5. Evaluating Post-Decision Data: Observe the post-judgment behavior of defendants using post-decision data. Focus on outcomes like bail decisions, recidivism rates, and employment status post-incarceration to illuminate bias and its potential sources. Example: If the bail decisions are not biased, we should expect to see more or less similar levels of pretrial misconduct among different racial groups. Arnold et al. (2018) shows that this is not the case. Blacks are systematically perceived to be higher risk and are less likely to be granted bail. However, estimates suggest that marginally released white defendants are about 22 percentage points more likely to be re-arrested prior to disposition than comparable black defendants.
6. **Assessing Long-Term Consequences of Disparities:** Just as important as ensuring that the justice system works efficiently and without bias is understanding how it impacts the lives of those that pass through it. It is possible that the long-term consequences of being in prison, for example, are different for different racial groups, even if they were convicted of the same crime. Linking judicial data to administrative data such as social security information, employment status, health information and so on, can provide an informative picture of the impact both of bias within the judicial system and the independent differential impact of court decisions on defendants of different races.
7. **Comparing Human and Machine Learning Decisions:** Deploy machine learning algorithms to study disparities and biases. Compare human decisions with algorithmic predictions to quantify the extent of human bias without necessarily replacing human judges with algorithms. *Example:* An algorithm, when built with rich data for clearly defined objectives, can be less biased than the human judge. As shown by Kleinberg et al. (2018), the algorithm can achieve the same crime rate as that achieved by human judges but by jailing 38.8 % fewer blacks and 44.6 % fewer Hispanics, suggesting that machine learning can also be used to identify biases and understand how to improve human decisions.
8. **Implementing and Evaluating Interventions:** Based on the analysis, implement appropriate interventions to reduce the biases identified. Include mechanisms to evaluate the effectiveness of these interventions and adapt as needed. This step would ensure continuous improvement and real-world effectiveness of the strategies implemented.

Overall, the results from this analysis could be used to aid decision-makers in ways that reduce bias in the system. For example, training programs could be targeted toward biased judges, either with the goal of debiasing or to help them learn how to use the hearing process to better advantage. Simply alerting judges to the fact that their behavior is highly predictable in ways that may indicate unfairness may also be sufficient to change their behavior. Informing judges about the predictions made by a model decision-maker could help reduce judgelevel

variation and arbitrariness. Potential biases that have been identified in prior decisions or writing could be brought to a judge's attention, where they could be subjected to higher-order cognitive scrutiny.

Thus, data not only makes it possible to measure and understand judicial bias, but also provides different avenues to address these biases and reduce existing discrimination. Leveraging such opportunities may be essential to ensure that not only de jure equality exists, but also that there is de facto equality in the interpretation and application of the law.

4.3. Applications of text as data

Recent innovations have generated new opportunities for empirical research on the delivery of justice. Court proceedings and rulings are now increasingly digitized, allowing the construction of large-scale data sets. The increase in text availability is particularly promising in the legal field, where legal documents are meticulously documented and play an essential role in judicial decisions. Additionally, computer scientists have developed a slate of machine learning tools that can produce interpretable data from unstructured text–including written judicial opinions–making it possible to analyze a quantity of text that would be far too large for humans to read.

The question of how to analyze texts has gained importance in social science research in recent years (Gentzkow et al., 2019). In the past, the most common approach was qualitative, with either a deep reading of the text or a subjective coding of important themes (see Glaser et al. (2017) for an example of the latter approach). However, these approaches lack a rigorous method of replication (Ricoeur, 1981; DiMaggio, 1997) and more formal methods to analyze texts have been developed (Andrade, 1995; Mohr, 1998). Topic modeling discovers underlying topics and themes through an inductive method (Blei et al., 2003; Blei, 2012; Mohr and Bogdanov, 2013). Another family of models learns the features of text that are predictive of some outcome, such as political ideology (Gentzkow et al., 2019; Jelveh et al., 2018; Ash, 2015; Ash et al., 2017; Osnabrügge et al., 2020).

Recent approaches have gone beyond the traditional network or topic methods by mapping word relations into a high-dimensional vector space (Mikolov et al., 2013; Pennington et al., 2014). This approach positions connected words close to each other in the space and can be used to recover relevant dimensions in language. A rich literature in computational social science has begun to apply these methods in many contexts (Rodriguez and Spirling, 2022).

More specifically, an active literature has begun to apply Natural Language Processing methods to legal documents (Carlson et al., 2016; Leibon et al., 2018). Ganglmair and Wardlaw (2017) apply a topic model to debt contracts; Ash et al. (2019) implement a syntactic parser to extract legal commitments and entitlements from union contracts. Ash and Chen (2019) construct document embeddings for federal courts and show they recover differences between courts, over time, and across topics. "Mapping the Geometry of Law Using Natural Language Processing," proposes NLP to understand judicial reasoning more broadly. The study employs Doc2Vec models to analyze judicial documents, revealing distinct patterns and tendencies in judicial decision-making across different courts and times. Their methodological innovation not only enhances our understanding of legal texts through the application of machine learning and NLP techniques but also pioneers predicting legal outcomes, such as the likelihood of cases being appealed to the Supreme Court.

As online collections of court decisions grow, many options become available to use text for legal research. For instance, legal scholars may be able to trace evolving interpretations of legal concepts such as fault, causation, or damages, or to examine patterns in how courts handle specific types of cases or parties (Liebman et al., 2020). Given that word embeddings measure correlations between words, they can be used to detect biases in language. Ash et al. (2021b) analyze gender bias in the language of US Circuit Court judges, finding that slanted judges vote more conservatively in gender-related cases.

Even further, NLP may be used not only to identify and understand human biases, but also to mitigate them. For instance, a possibility may be to prevent prosecutors from seeing irrelevant information about a case (such as race or ethnicity) when making an initial decision on whether to charge someone. The San Francisco District Attorney's Office has begun to use an algorithm that automatically redacts race-related information from free-text case narratives (Chohlas-Wood et al., 2021). This approach has been described as "blind charging", and may enable prosecutors' offices to reduce the potential racial bias in prosecutors' initial decisions on whether to charge someone.

The promise of NLP may also be of particular relevance to support legal actors' knowledge of jurisprudence. Legal search engines already compile jurisprudence and identify decisions that are similar to the one at hand; decisions that apply similar laws and regulations; or decisions with similar case patterns, among other characteristics. Thus, these tools bring opportunities to improve the quality and consistency of legal arguments and judicial decisions, as well as to improve the training of lawyers, judges and prosecutors. Summarization tools may also be able to extract patterns of interest, reducing the amount of work spent on inputting case information into case management systems. Even further, NLP may be able to automatize basic judicial decisions, such as those that only require human revision of documents. For instance, according to a report submitted to the Administrative Conference of the United States, NLP is already being used to improve the accuracy and efficiency of formal adjudication in the United States (Engstrom et al., 2020). A tool called Insight is used to parse text in draft decisions to flag potential errors. The tool identifies weaknesses in draft opinions that are suggestive of policy noncompliance or internal inconsistencies in the decision. Thus, it aims to ensure that adjudicators properly go through the analysis required by regulations.

5. Future directions: data for better judiciaries

In a companion paper, we discussed opportunities for researchers to expand the empirical evidence regarding the role of justice in development (Ramos-Maqueda and Chen, 2024). In this paper, we discuss how the data revolution in justice systems may bring ample opportunities for policymakers and researchers alike to improve the performance of the justice sector.

In addressing the challenges of data collection and standardization in judicial systems, particularly within varying economic contexts, this study acknowledges significant disparities. Several initiatives, including by our own research team, are spearheading a global initiative to enhance and standardize judicial data. This initiative focuses on the development and implementation of advanced data collection frameworks and consistent data entry protocols across jurisdictions. With an emphasis on leveraging cutting-edge data technologies and extensive training programs for court staff, these initiatives aim to ensure uniform data quality and reliability. Such efforts are crucial for maintaining the integrity and comparability of judicial data, thus facilitating more accurate and impactful comparative legal research.

To exploit the potential of the data revolution, countries should start by creating a data ecosystem that relates different sources of data to each other, thereby improving the universe of analyses that becomes possible. This requires planning ahead and making an up-front investment in data infrastructure whose benefits may not be realized until the long term. Realizing the promise of data requires another essential asset: human capital. In the legal field, engineers generally lack domain knowledge whereas lawyers tend to lack the technical skills to wrangle and analyze data. Thus, staff training and hiring may be essential to ensuring that data is used responsibly to develop effective innovations. It is through investment in human capital that countries may leverage the promise of data in justice systems for the greater good.

Once the appropriate data ecosystem and the human capital are established, we propose the following paths for the use of data, each of which is explored in greater detail below: 1) access to e-justice for

citizens; 2) data for better court performance; and 3) data for better knowledge.

First, the data revolution can bring better access to e-justice for citizens and companies. One of the most salient problems facing legal systems around the world is the lack of access to formal justice mechanisms. Thus, identifying ways to make justice more accessible for vulnerable populations is on top of the policy agenda (Steven et al., 2020). Data may be used to identify gaps in access to justice, for example, by comparing legal need surveys to the actual disputes that are brought to courts. This may help policymakers better understand if there are specific disputes that elude the justice system, where better access to legal means of resolving disputes might be the main gap. For instance, if surveys indicate that a district is deeply affected by domestic violence, but few cases of domestic violence are resolved in courts, this will provide evidence of a gap in domestic violence cases being brought to courts. Data systems may also be used to inform citizens of their prospective outcomes in mediation as compared to courts, allowing them to decide the most appropriate mechanism to resolve their dispute (Sadka et al., 2017). Furthermore, data and technology may create new avenues to access justice. For instance, in response to the COVID-19 pandemic, the Chilean judiciary developed a digital platform, Conecta, which allows citizens to access courts' services through a video call, chat, or WhatsApp. By bringing different data sources together in one platform, Conecta allows courts to effectively respond to citizens' queries, share updates on their cases, and provide a more efficient service to litigants. Going further, a comprehensive data infrastructure may introduce the ability to build chatbots that improve institutional capacity to respond to citizens' queries and demands.

In addition to this example, other tools such as e-arbitration or electronic hearings may represent important ways in which data and technology can bring justice closer to citizens and improve the effectiveness of the legal system while enhancing people's trust in institutions. For instance, various countries have introduced virtual courts to resolve fast or urgent matters. First, Canada introduced the Civil Resolution Tribunal (CRT)—an online tribunal that allows online dispute resolution, mostly related to small claims, property disputes, and traffic accidents. Other countries such as India, Pakistan, and South Africa rapidly introduced virtual courts to resolve urgent matters as a result of the COVID-19 pandemic. These technologies filled in a crucial gap in access to justice as courts closed, yet they may also bring risks to effective dispute resolution going forward. For example, a study in the United States finds that the switch to video in bail hearings in Cook County led to an average increase of 51 % in overall bail amounts (Diamond et al., 2010). A rigorous evaluation of the trade-offs between faster and–potentially–more accessible justice, on the one hand, and the risks associated with online communication, cyber-security, and disparities in access, on the other, should be undertaken before adopting these technologies going forward.

Second, the data revolution may improve court performance. Data has the promise of improving the efficiency, fairness, and effectiveness of justice systems. It may be used to identify the main bottlenecks of courts, i.e., the reasons why cases get backlogged or why adjournments take place (Chemin et al., 2022). Data systems have the potential to facilitate a better deployment of resources by diagnosing and forecasting future needs in courts and litigants' needs. In addition to providing diagnoses, there are data-driven solutions that may improve court performance. An example is dashboard management tools and nudges to court actors. These apps allow for a wide range of options: from providing personalized feedback on best and worst-performing metrics, to motivating better performance via behavioral nudges (Cooke et al., 2018). These applications can be enhanced by bringing AI to act as a support tool for court actors. AI can help predict future congestion in courts, of particular importance due to the pandemic's effect on adjournments; detect early and prioritize cases at high risk of gender-based violence; or potentially help predict recidivism (Kleinberg et al., 2018). These innovations may come with risks and trade-offs that need to be evaluated and calibrated, yet the promise of the applications makes them worthy of a proper assessment. Finally, data may also be used to identify and address judicial biases and inconsistencies, including gender and racial discrimination or snap judgments, among others. We have presented in this paper an 8-step approach by which data integration may help identify such biases.

Third, the data revolution may be used to create better knowledge within and across justice systems. Data has the promise to improve the quality and accessibility of judicial decisions and promote best practices among legal actors. As a diagnostic tool, data may help evaluate the text of judicial decisions, identify inconsistencies in writings, and create tools for better training of judges. Specific innovations may include, for example, legal search engines that allow judges and court actors to find cases with similar fact patterns or legal citations, thus speeding up the process–and ideally the quality–of legal decisions. This data may also be used to train judges based on the history of their past decisions in comparison to their peers. In addition, data may not only help legal actors develop a knowledge base for their own decisions, but may also help to create and share knowledge with peers. For example, the sharing of advice in a systematic manner across mediators or judges may create a set of best practices that others in the profession can use to improve the quality and efficiency of their services. This knowledge can scale up beyond cases in a country and actually refer to laws across countries. The data revolution may bring the possibility to identify best legal practices or"missing laws" in any specific country by developing an empirically-based method to analyze and compare existing regulations to enhance a country's legal system.

In recognizing the significant influence of political-economic factors on judicial data reporting, we acknowledge that the environment in which data is generated can profoundly affect its availability and reliability. We suggest that variables from existing global databases like the World Justice Project can be used to assess the impact of political-economic factors on the quality of judicial data across countries, providing a broader context to any specific study. By recognizing the incentives for data manipulation—ranging from political pressures to institutional corruption—we hope future researchers have a nuanced understanding of the potential biases and limitations within existing datasets. Future research can guide developing the mechanisms for enhancing transparency and accountability in judicial reporting processes.

Overall, the efforts to collect data, measure indicators, evaluate performance, and test innovations in judiciaries around the world can improve the efficiency, access to, and quality of justice. However, despite the extensive survey efforts on legal needs and access to justice, such as those offered by the World Justice Project and the World Bank, it is evident that a comprehensive, systematized repository of judicial data across diverse legal domains remains elusive, much less, the knowledge of which countries have these repositories available for actionable analysis. This limitation not only hinders the potential for comparative legal studies but also underscores the critical need for enhanced data collection, standardization, and transparency within judicial systems worldwide. Recognizing these challenges, we advocate for international cooperation and innovation in legal data analytics, which are essential for supporting robust, empirical research in law and justice. Data systems and technological innovations also bring the opportunity to improve the functioning of justice systems while also evaluating the impact of justice reforms. Such opportunities should not be overlooked, particularly in developing countries, given the crucial role that justice and legal institutions play in people's lives and the prosperity of nations.

CRediT authorship contribution statement

Manuel Ramos-Maqueda: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Daniel**

L. Chen: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Chen acknowledges financial support from the IAST, TSE-Partnership, and Artificial and Natural Intelligence Toulouse Institute (ANITI) funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EUR-0010. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily reflect the views of the World Bank and its affiliated organizations, nor those of the Executive Directors of the World Bank or the governments they represent.

Data availability

No data was used for the research described in the article.

References

Abrams, S., Bertrand, M., & Mullainathan, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies, 41*(2), 347–383.
Agarwal, B. (2003). Gender and land rights revisited: Exploring new prospects via the state, family and market. *Journal of Agrarian Change, 3*(1–2), 184–224.
Albonetti, C. A. (1997). Sentencing under the federal sentencing guidelines: Effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991-1992. *Law & Society Review, 31*(4), 789–822.
Andrade, J. (1995). Learning during anaesthesia: A review. *British Journal of Psychology, 86*(4), 479–506.
Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics, 133*(4), 1885–1932.
E. Ash. The political economy of tax laws in the U.S. states. 2015. URL: http://www.barcelona-ipeg.eu/wp-content/uploads/2015/09/Ash-tax-laws-current.pdf.
Ash, E., & Chen, D. L. (2018). Mapping the geometry of law using document embeddings. Technical Report TSE Working Paper No. 18-935, Toulouse School of Economics, Toulouse, France, July 2018. URL https://papers.ssrn.com/abstract=3305761.
Ash, E. & Chen, D. L. (2019). Case vectors: Spatial representations of the law using document embeddings. In M. A. Livermore and D.N. Rockmore, editors, Law as data, 313–337. Santa Fe Institute Press.
Ash, E., Morelli, M., & Van Weelden, R. (2017). Elections and divisiveness: Theory and evidence. *The Journal of Politics, 79*(4), 1268–1285.
Ash, E., MacLeod, B., & Naidu, S. (2019). The language of contract: Promises and power in union collective bargaining agreements. Stockholm, Sweden.
Ash, E., Asher, S., Bhowmick, A., Chen, D., Devi, T., Goessman, C., Novosad, P., & Siddiqi, B. (2021). In-group bias in the Indian judiciary: Evidence from 5.5 million criminal cases. Technical report, Oxford Policy Management.
Ash, E., Chen, D. L., & Ornaghi, A. (2021). Gender Attitudes in the judiciary: Evidence from U.S. Circuit Courts. Working Paper No. 02/2019, ETH Zurich.
Berdejo, C., & Chen, D. L. (2017). Electoral cycles among US Courts of Appeals judges. *The Journal of Law and Economics, 60*(3), 479–496.
Bielen, S., Marneffe, W., & Mocan, N. (2021). Racial bias and in-group bias in virtual reality courtrooms. *The Journal of Law and Economics, 64*(2), 269–300.
Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.
Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science, 54*(2), 389–411.
Buscaglia, E., & Dakolias, M. (1996). *Judicial reform in Latin American courts: The experience in Argentina and Ecuador. Technical Report World Bank Technical Paper No. 350*. Washington, D.C.: The World Bank.
Carlson, K., Livermore, M. A., & Rockmore, D. (2016). A quantitative analysis of writing style on the U.S Supreme Court. *Washington University Law Review, 93*(6), 1461–1510.
Carmignani, A. & Giacomelli, S. (2010). Too many lawyers? Litigation in Italian civil courts. Working Paper 745, Bank of Italy. URL https://papers.ssrn.com/abstract=1669988.
Chemin, M. (2020). Judicial efficiency and firm productivity: Evidence from a world database of judicial reforms. *The Review of Economics and Statistics, 102*(1), 49–64.
Chemin, M., Chen, D. L., Di Maro, V., Kimalu, P., Mokaya, M., & Ramos Maqueda, M. (2022). Data Science for Justice: The Short-Term Effects of a Randomized Judicial Reform in Kenya. TSE Working Papers. URL https://ideas.repec.org//p/tse/wpaper/127593.html. Number: 22-1391 Publisher: Toulouse School of Economics (TSE).
Chen, D.L. (2016). Priming ideology: Why presidential elections affect U.S. Judges. Technical Report TSE Working Paper No. 16-681, Toulouse School of Economics, Toulouse, France. URL https://econpapers.repec.org/paper/tsewpaper/30576.htm.
Chen, D. L. (2017). Mood and the malleability of moral reasoning. Technical Report TSE Working Paper No. 16-707, Toulouse School of Economics, Toulouse, France, 2017. URL: https://www.tse-fr.eu/publications/mood-and-malleability-moral-reasoning. Publication Title: TSE Working Papers.
Chen, D. L., & Eagel, J. (2017). Can machine learning help predict the outcome of asylum adjudications?. *In Proceedings of the ACM Conference on AI and the Law*.
Chen, D. L. & Philippe, A. (2020). Clash of norms: Judicial leniency on defendant birthdays. Working Paper, 2020. URL https://papers.ssrn.com/abstract=3203624.
Chen, D.L., Kozbur, D., & Yu, A. (2015). Pandering vs. persuasion? Phonetic accommodation in the U.S. Supreme Court. Working Paper. URL http://users.nber.org/ dlchen/papers/Mimicry.pdf.
Chen, D. L., Frankenreiter, J., & Yeh, S. (2016). Judicial compliance in district courts. Technical Report TSE Working Paper No. 16-715, Toulouse School of Economics, Toulouse, France. URL: https://www.tse-fr.eu/publications/judicial-compliance-district-courts.
Chen, D. L., Halberstam, Y., & Yu, A. (2017). Covering: Mutable characteristics and perceptions of voice in the U.S. Supreme Court. Working Paper. URL https://papers.ssrn.com/abstract=2816241.
Chen, D. L., Graham, J., Ramos Maqueda, M., & Singh, S. (2022). Do judges favor their own ethnicity and gender? World Bank Policy Research Working Paper, 9956.
Chohlas-Wood, A., Nudell, J., Yao, K.,. Lin, Z. J, Nyarko, J., & Goel, S. (2021). Blind Justice: Algorithmically Masking Race in Charging Decisions. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, AIES '21, pages 35–45, New York, NY, USA, July 2021. Association for Computing Machinery.
Choi, D. D., Harris, J. A., & Shen-Bayh, F. (2022). Ethnic bias in judicial decision making: Evidence from criminal appeals in Kenya. *American Political Science Review, First View*, 1–14.
Cooke, B., Diop, B. Z., Fishbane, A., Hayes, J., Ouss, A., & Shah, A. (2018). Using behavioral science to improve criminal justice outcomes: Preventing failures to appear in court. Technical report, ideas42 and the University of Chicago Crime Lab.
Coviello, D., Ichino, A., & Persico, N. (2015). The inefficiency of worker time use. *Journal of the European Economic Association, 13*(5), 906–947.
Dahl, G. B., Kostøl, A. R., & Mogstad, M. (2014). Family welfare cultures. *The Quarterly Journal of Economics, 129*(4), 1711–1752.
Dakolias, M. (1999). Court performance around the world: A comparative perspective. Technical Report World Bank Technical Paper No. 430, World Bank Group, Washington, D.C.
Dakolias, M., & Said, J. (2000). Judicial reform: A process of change through pilot courts. *European Journal of Law Reform, 2*(1), 95–117.
Daley, E., Dore-Weeks, R., & Umuhoza, C. (2010). Ahead of the game: Land tenure reform in Rwanda and the process of securing women's land rights. *Journal of Eastern African Studies, 4*(1), 131–152.
Deininger, K., Goyal, A., & Nagarajan, H. (2013). Women's inheritance rights and intergenerational transmission of resources in India. *Journal of Human Resources, 48*(1), 114–141.
Diamond, S. S., Bowman, L. E., Wong, M., & Patton, M. M. (2010). Efficiency and cost: The impact of video conferenced hearings on bail decisions. *Journal of Criminal Law & Criminology, 100*(3), 869–902.
DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology, 23*(1), 263–287.
Dobbie, W., & Song, J. (2015). Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *American Economic Review, 105*(3), 1272–1311.
Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cúellar, M.-F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. Technical report, Administrative Conference of the United States, Washington, D.C.
Eren, O., & Mocan, N. (2018). Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics, 10*(3), 171–205.
Fryer, R. G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy, 127*(3), 1210–1261.
Ganglmair, B. & Wardlaw, M. (2017). Complexity, standardization, and the design of loan agreements. Working Paper, Apr. 2017. URL https://papers.ssrn.com/abstract=2952567.
Gazal-Ayal, O., & Sulitzeanu-Kenan, R. (2010). Let my people go: Ethnic in-group bias in judicial decisions—Evidence from a randomized natural experiment. *Journal of Empirical Legal Studies, 7*(3), 403–428.
Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature, 57*(3), 535–574.
Glaser, B. G., Strauss, A. L., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. London and New York: Routledge.
Goncalves, F., & Mello, S. (2021). A few bad apples? Racial bias in policing. *American Economic Review, 111*(5), 1406–1441.
Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *Cornell Law Review, 86*(4), 777–830.
Hyland, M., Djankov, S., & Goldberg, P. K. (2020). Gendered laws and women in the workforce. *American Economic Review: Insights, 2*(4), 475–490.
Jelveh, Z., Kogut, B., & Naidu, S. (2018). Political language in economics. Technical Report Columbia Business School Research Paper No. 14-57, Sept. 2018. URL http://www.ssrn.com/abstract=2535453.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics, 133*(1), 237–293.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review, 96*(3), 863–876.

Leibon, G., Livermore, M., Harder, R., Riddell, A., & Rockmore, D. (2018). Bending the law: Geometric tools for quantifying influence in the multinetwork of legal opinions. *Artificial Intelligence and Law, 26*, 145–167.

Liebman, B. L., Roberts, M. E., Stern, R. E., & Wang, A. Z. (2020). Mass digitization of Chinese court decisions: How to use text as data in the field of Chinese law. *Journal of Law and Courts, 8*(2), 177–201.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space, Sept. 2013. URL http://arxiv.org/abs/1301.3781.

Mitsopoulos, M., & Pelagidis, T. (2010). Greek appeals courts' quality analysis and performance. *European Journal of Law and Economics, 30*(1), 17–39.

Mohr, J. W. (1998). Measuring meaning structures. *Annual Review of Sociology, 24*(1), 345–370.

Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics, 41*(6), 545–569.

Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. federal courts. *The Journal of Law and Economics, 44*(1), 285–314.

Osnabrügge, M., Ash, E., & Morelli, M. (2020). Measuring topics using crossdomain supervised learning: Methods and application to New Zealand Parliament. Technical Report No. 04/2020, ETH Zurich.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543, Doha, Qatar. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162. URL http://aclweb.org/anthology/D14-1162.

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review, 84*(3), 1195–1246.

Rachlinski, J. J., Wistrich, A. J., & Guthrie, C. (2015). Can judges make reliable numeric judgments: Distorted damages and skewed sentences. *Indiana Law Journal, 90*(2), 695–739.

Ramos-Maqueda, M. & Chen, D. L. (2024). The role of justice in development i: How rule of law spurs economic growth. Policy Research Working Paper, 9720.

Rehavi, M. M., & Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy, 122*(6), 1320–1354.

Reno, J., Dwyer, J. C., Robinson, L., & Gist, N. E. (1997). Trial court performance standards with commentary. Technical report, U.S. Department of Justice, Washington, D.C., July 1997. URL: https://www.ojp.gov/pdffiles1/161570.pdf.

Ricoeur, P. (1981). Hermeneutics and the human sciences: Essays on language, action and interpretation. Cambridge University Press. ISBN 978-0-521-28002-0. Google-Books-ID: 8L1CwJ0U DsC.

Rodriguez, P. L., & Spirling, A. (2022). Word embeddings: What works, what doesn't, and how to tell the difference for applied research. *The Journal of Politics, 84*(1), 101–115.

Rudin, C., & Ustun, B. (2018). Optimized scoring systems: Toward trust in machine learning for healthcare and criminal justice. *Interfaces, 48*(5), 449–466.

Sadka, J., Seira, E., & Woodruff, C. (2017). Overconfidence and settlement: Evidence from Mexican labor courts. Working Paper.

Schanzenbach, M. (2005). Racial and sex disparities in prison sentences: The effect of District-level judicial demographics. *The Journal of Legal Studies, 34*(1), 57–92.

Shayo, M., & Zussman, A. (2017). Conflict and the persistence of ethnic bias. *American Economic Journal: Applied Economics, 9*(4), 137–165.

Steffensmeier, D., & Demuth, S. (2000). Ethnicity and sentencing outcomes in U.S. federal courts: Who is punished more harshly? *American Sociological Review, 65*(5), 705–729.

Steven, D., De Langen, M., Muller, S., & Weston, M. (2020). Justice for all and the economic crisis. Technical report, Pathfinders for Peaceful, Just, and Inclusive Societies. URL: https://cic.nyu.edu/sites/default/files/justice-for-all-and-the-economic-crisis-web-final.

United Nations (2018). The sustainable development goals report: 2018. Technical report, New York. URL: https://unstats.un.org/sdgs/files/report/2018/TheSustainableDevelopmentGoalsReport2018-EN.

Yeung, L. L., & Azevedo, P. F. (2011). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics, 22*(4), 343–356.

Zingraff, M., & Thomson, R. (1984). Differential sentencing of women and men in the U.S.A. *International Journal of the Sociology of Law, 12*(4), 401–413.