NLSIR

THE PROMISE OF MACHINE LEARNING FOR THE COURTS OF INDIA

—Sandeep Bhupatiraju, Daniel L. Chen & Shareen Joshi*

Abstract—Artificial Intelligence and machine learning offer new opportunities for improving the decision-making capacity and productivity of the Indian judiciary. First, the algorithmic analysis of legal data can provide human decision-makers with timely alerts of biases at critical decision-making moments, and also propose real-time corrections for these behaviors. Analysis of texts for patterns of bias and discrimination, for example, can augment the capabilities of judges and lawyers and systematise processes of review. Second, machine learning tools can also be deployed to clean, systematize, and standardize legal data. Though the judiciary has made significant investments in data systems, the variations in quality across states and administrative boundaries prevent a deeper analysis of the data. Third, the deployment of machine learning methods creates new opportunities to ensure procedural fairness and also enables legal scholars to better study the courts themselves. When cases are randomly assigned to judges researchers can evaluate the impact of judicial decisions — since judges in this scenario do not choose their cases and end up with them randomly, observed rulings reflect their deliberations in the case rather than the process of justice that led them to be assigned the case. We emphasize however, that technology must be viewed as a complement to human decision-makers and not a substitute. Only technologies that aid humans, rather than replace them, are suitable in this setting.

* **Sandeep Bhupatiraju**, World Bank, 1818 H Street Washington, DC 20433. Email: sbhupatiraju@worldbank.org. **Daniel L. Chen**, Toulouse School of Economics, Université de Toulouse Capitole, Toulouse, France and World Bank, 1818 H Street Washington, DC 20433. Email: dlchen@nber.org.**Shareen Joshi,** Walsh School of Foreign Service, Georgetown University, 3700 O Street, Washington DC, 20057. Email: shareen.joshi@georgetown.edu.

I. INTRODUCTION

Artificial Intelligence ('AI') and machine learning ('ML')— adaptive computer programs that attempt to perform functions typically associated with the human mind — offer new opportunities for improving the productivity of large-scale organisations.[1] Recent advances in data collection, systems of aggregation, algorithms, and processing power have transformed computers from machines that perform rigidly defined tasks to ones that can learn without supervision and adapt to new inputs without the need to be reprogrammed. These advances have already brought tangible benefits in the world of business, medicine and large-scale systems more generally.[2]

AI platforms offer a variety of opportunities to improve justice systems.[3] The large volume of data from surveillance systems, digital payments platforms, newly computerised bureaucratic systems and even social media platforms can be analysed to detect anomalous activity, investigate potential criminal activity, and improve systems of justice. AI systems can also reduce the barriers to accessing courts by providing users with timely information directly, rather than through lawyers. In the United States, AI is already used in the processing of bail applications, DNA analysis of crimes, gunshot detection, and crime forecasting.[4]

For the purpose of this article, we will focus on the potential of AI systems to improve data quality and the processes of justice themselves. This set of applications does not feature AI systems as replacements for human decision-makers. Rather, they help *improve* human decision-making capacity and productivity. A growing body of work demonstrates how small external factors, most of which the participants are unaware of, can leave their mark on the outcomes of legal cases . Analysis of the US, French, Israeli, UK, and Chilean courts for example, find in various settings, that the tone of words used in the first three minutes of a hearing, the incidence of birthdays, the outcomes of sporting events, and even the time of day of a hearing or defendant's name, affect the outcome of cases.[5] These methods can also identify both con-

---

1 S Russel and P Norvig, *Artificial intelligence: A Modern Approach* (3rd edn, Pearson Education 2013).

2 E Brynjolfsson and A McAfee, 'The Business of Artificial Intelligence' (2017) Harvard Business Review, 1-20; JH Chen and SM Asch, 'Machine Learning and Prediction in Medicine—Beyond the Peak of Inflated Expectations' (2017)376 (26) The New England Journal of Medicine 2507.

3 DL Chen, 'Judicial analytics and the great transformation of American Law' (2019) 27(1) Artificial Intelligence and Law15-42; C Rigano, 'Using Artificial Intelligence to Address Criminal Justice Needs' (2019) 280 National Institute of Justice 1; DL Chen, 'Machine Learning and the Rule of Law' in M Livermore and D Rockmore (eds), *Computational Analysis of Law* (Santa Fe Institute Press) (forthcoming).

4 Rigano (n 4); [7] WJ Epps Jr and JM Warren, 'Now Being Deployed in the Field of Law' 59(1) The Judges' Journal 16-39.

5 Chen (n 4).

scious and unconscious bias. The analysis of 18,686 judicial rulings, collected over seventy-seven years, by the twelve US circuit courts (also known as courts of appeals or federal appellate court) illustrated that judges demonstrate considerable bias before national elections.[6] Similarly, there is new evidence on sequencing matters in high-stakes decisions: decisions made on previous cases affect the outcomes of current cases, even though the cases have nothing to do with each other. Refugee asylum judges are two percentage points more likely to deny asylum to refugees if their previous decision granted asylum.[7]

AI systems have enormous potential to improve the processes of justice in India. Human capacity has already been identified as a significant constraint in the system. India has just nineteen judges per million people, and twenty-seven million (2.7 crore) pending cases.[8] The justice system has already made considerable advances in adoption of information technology, released large volumes of data to court users and encouraged them to use electronic systems. Yet, legislative, institutional, and resource constraints have limited their full impact.[9]

We argue that the integration of machine learning tools with newly available legal data offers a mechanism to identify biases in judicial behavior and propose real-time corrections for these behaviors. This can result in a more streamlined system and reduction in backlog . Such tools can identify discrimination and bias even when these are not evident to the participants in the courts themselves, thereby strengthening the credibility of the judiciary.[10] The availability of data enables new varieties of academic research on the efficiency and effectiveness of the legal system itself: macro-level studies conducted as 'distant reading' of the system, rather than micro-level 'close reading' as commonly done in legal studies.

The adoption of AI systems, however, is not a panacea for a court system. The technological sophistication of the tools creates what is often called

---

6 C Berdejo and DL Chen, 'Electoral Cycles Among us Courts of Appeals Judges' (2017) 60(3) The Journal of Law and Economics 479-496.

7 DL Chen, TJ Moskowitz and K Shue, 'Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires' (2016)131(3) The Quarterly Journal of Economics 1181-1242.

8 VA Kumar, 'Judicial Delays in India: Causes & Remedies' (2012) 4 Journal of Law Policy & Globalization 16; M Chemin, 'Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India' (2012) 28(3) The Journal of Law, Economics, & Organization 460-485; A Amirapu, 'Justice Delayed is Growth Denied: The Effect of Slow Courts on Relationship-Specific Industries in India' (2020) Economic Development and Cultural Change <https://doi.org/10.1086/711171> accessed on 16 August 2021.

9 Amirapu (n 9); D Damle and T Anand, 'Problems with the e-Courts Data' (2020) National Institute of Public Finance and Policy Working Paper 314 <https://www.nipfp.org.in/media/medialibrary/2020/07/WP_314__2020.pdf> accessed 16 August 2021.

10 K Kannabiran, 'Judicial meanderings in Patriarchal thickets: Litigating sex discrimination in India' (2009) 44(44) Economic and Political Weekly 88-98; M Galanter, *Competing Equalities: Law and the Backward Classes in India* (OUP 1984); P Bhushan, 'Misplaced Priorities and Class Bias of the Judiciary' (2009) 44(14) Economic and Political Weekly 32-37.

the "black box" problem:[11] their technological sophistication makes them less interpretable to many. The challenge of interpretability also raises concerns about the accountability and oversight for these systems. Furthermore, the gap between those who can and cannot access and understand these technologies exacerbates existing social divisions and intensifies polarisation. For all these reasons, we refrain from advocating that the tools of AI and ML be used to replace human decision-making through, for example, automation of bail applications. Rather, we believe that the systems be leveraged to *aid* and *improve* human decision-making within the system. We believe that the adoption of AI systems stimulates reflection on questions that are truly fundamental to the systems of justice and that the benefits of the system are considerable when leveraged in a careful and ethical way.

II. THE PROMISE OF NEW DATA

In the past fifteen years, considerable efforts have been made to adopt and deploy information technology systems in the courts of India. One of the most significant projects, the e-courts project, was first launched in 2005 by the Supreme Court of India through the "National Policy and Action Plan for Implementation of Information and Communication Technology (ICT) in the Indian Judiciary".

The e-courts initiative introduced technology in Indian courts in a variety of ways. The most obvious and observable feature of the system was the deployment of technology within the court rooms itself. Judges were provided with LCD touch screen machines, screens and projectors were connected via a local network to disseminate information to lawyers, electronic boards at the courts display the queue of case numbers for hearing scheduled on a particular day, etc. Outside of the courtroom, e-filing procedures were established, and an architecture of data management was created that ranged from scanning of old cases into the electronic system, the creation of digital archives and the establishment of direct electronic communication with litigants and an online case management system. These investments have eventually paved the way for the creation of the National Judicial Data Grid (NJDG), a database of twenty-seven million cases available to court users to view the status of pending cases and access information on past hearings.

For our purposes, the most significant resource available through this has been the digital archives of cases. We were able to scrape these publicly available digital archives to construct an e-Courts district court dataset of eighty-three million cases from 3289 court establishments.[12] We were able to curate details like the act under which the case is filed, the case type (criminal or

[11] F Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (HUP 2015).

[12] The eCourts data is public and can be accessed via the district court websites, the eCourts Android/iOS app, or the district court services webpage.

civil), district where it originates, the parties to the case, and the history of case hearings, in a manner that makes the data amenable to large-scale analysis.

We supplement this database with a variety of other data sources. Three examples of these are below.

Data on Judges: In order to better understand the impact of specific judges — their identity, training and experience — we have constructed a database of judges for the courts of India. We have begun this task by extracting data from the Judges Handbooks, released by the Supreme Court of India, and appending to it information from various High Court websites. Thus far, we have assembled details of 2,239 judges from the handbooks for the years between 2014 and 2020. Most notably, 93.5% of these judges are males and 6.5% are females and their range of experience covers a period spanning approximately seventy years.

Database of Central Acts: This auxiliary dataset is intended to give a definitive list of standardised act names. This could then be used to standardise the act names appearing in the various cases. This allows us to analyse all cases filed under a given act. We have, for example, examined all cases related to the Water Act of 1974 and found a total of 978 such cases at the Supreme Court and High Courts of India. The list of central (federal) acts can be viewed on the Legislative Department website of the Ministry of Law and Justice. There is currently no centralised source for all state legislation — this needs to be obtained from the state websites separately.

Other Administrative Data: Data on other institutions can be linked to the judicial data at the district as well as the state level. For example, data on Indian banks and their branches is available through the Reserve Bank of India. This database contains information on their name, location, license number, license date, address, and other unique identifiers. We have scraped, cleaned and organised this data for further analysis. It contains about 160,000 entries. The unique identifiers and location information allow us to merge this data with banks appearing in litigation in courts that are present in the e-Courts databases. The merging of this data with the legal data allows us to examine a variety of interesting questions about the development of financial markets in an area, participation in the justice system, and the impacts of legal rulings.

III. POTENTIAL APPLICATIONS OF MACHINE LEARNING IN THE COURTS OF INDIA

Legal data released by the Indian judiciary is voluminous, messy, and complex.[13] The typical case has clear tags for some key dates (filing date, order

[13] Damle and Anand (n 10).

date, etc.), key actors (petitioner, respondent, and judges) and court name, but information about the type of case, outcome of the deliberations, and pertinent acts cited are not clearly identifiable in the body of the orders or judgements. Cleaning and pre-processing the data is critical for any form of analysis, especially so for supervised algorithms trained on this data. Traditional empirical legal studies have typically addressed this issue by relying on small-scale data sets, where legal variables are manually coded, and the scope of inference is related to a small body of legal cases that is pertinent to a single issue.[14]

One of the biggest challenges in pre-processing the data is the variability of reporting across states and districts. The quality of the data varies significantly — there is no nationally standardised system for defining variables or reporting on them. For instance, in some states the act name and section numbers are well delineated and a larger proportion of cases have orders uploaded and in others this is not the case. This makes it difficult to compare individual case-types across courts and across states.[15] There are no standardised identifiers within the data to follow a case throughout its potential sequence of appeals in higher courts. In a similar vein, there is no easy way to track a criminal case from its entry into the system as a FIR to its exit as a judgment. There are inconsistencies in identifying information about participants, their attributes and the types of laws or acts that the case relates to. There are also issues of incorrect reporting and spelling mistakes. Entries from one field can sometimes show up in another, requiring careful cleaning and systematic recoding of variables.

Various ML tools can be harnessed to improve data quality and address the issues discussed above. We have constructed a robust pipeline to scrape, clean and prepare this data for analysis. We briefly describe some of these methods below.

A. Inference about the identity of participants

Some databases of judgements provide no identifying information on the participants in the cases themselves. To better understand who participates in the courts, we first extract litigant names from the raw text of the judgements, and then using some matching algorithms, we identify the type of litigant

---

[14] V Gauri, 'Public interest litigation in India: Overreaching or Underachieving?' (2009) World Bank Policy Research Working Paper 5109 <https://poseidon01.ssrn.com/delivery.php?ID=709021124002094091083101097022125125019054034003088001030009059006043060390390971260910161050640670260310500571030050231240260300040261130670290270970071051250220650690830940820970170130 24&EXT=pdf&INDEX=TRUE> accessed 18 August 2021; S Krishnaswamy, S K Sivakumar and S Bail, 'Legal and Judicial Reform in India: A Call for Systemic and Empirical Approaches' (2014) 2(1) Journal of National Law University Delhi1-25; U Baxi, *Towards a Sociology of Indian Law* (1st edn, New Delhi: Satvahan 1986).

[15] Damle and Anand (n 10).

(individuals, companies, or state institutions). Classifying participants can be challenging. If we are interested in cases that involve the government for example, we must be able to identify all the different agencies of the state government, national government, and the additional agencies that fall within the definition of "state" under Article 12 of the Indian constitution. Manually tagging these entities is prohibitively time consuming and the existence of latent patterns in the names makes this fertile ground for ML applications.

Once we extract names, we can draw inferences about the characteristics of individual participants. Some obvious attributes of interest are a participant's gender, caste, and religion. These attributes, however, are not officially recorded in court proceedings. Again, ML methods provide a possible solution. We have focused here on people's first and last names, for illustration.

We first format individual names to ensure that each individual could be identified by an honorific title, a first name, and a last name. Honorifics such as Shri, Sri, Smt., Mr., Mrs., and Ms. enabled us to directly identify gender. To extend this classification to names without an honorific, we train an algorithm on a publicly available corpus of labeled common Indian first names. Training this algorithm, often referred to as training a classifier, is the process of learning patterns within the data related to the classification. Here, these patterns are the statistics of co-occurrence of alphabets in names, length of the name, and other features which have some predictive information. In order to reduce the generalisation error, we use the majority vote from multiple trained classifiers, including a logistic regression model and a random forest classifier to make predictions on gender.[16] A logistic regression models the probability of a binary outcome or event. A random forest classifier will use decision trees (nested if-then statements) on features of the data to make the prediction.

We have also made predictions of religion and caste using similar approaches. Muslims can be recognised in the data through the distinctiveness of Muslim names: common names such as *Khan* and *Ahmed* can easily be assigned and coded, but for others we utilise the occurrence of specific alphabets (such as Q and Z) through appropriate classifiers, to identify additional names. These algorithms formalise our intuitive notions of why a name belongs to a given group by identifying frequently occurring patterns within names associated to that group. Caste assignment is more complicated because the same last name can be associated with multiple caste groups. The name

[16] The features (x-values) in all the models were hand-engineered co-occurrence statistics of blocks of alphabets in various locations within the names; The voting procedure is a way to ensemble models so as to reduce the generalisation error. For instance, if we had three prediction algorithms for gender and all of them make a prediction of M or F, then we use the majority vote as the final prediction of the ensemble. In this case, at least two of the algorithms would have predicted the same class and we use this as our final prediction. We used the cross-entropy function as the loss function to quantify how well a given classifier did and obtained an accuracy of 0.92 for the ensemble model.

Kumar, for example, could be the name of a person belonging to SC, ST or 'Other' category. In the case of such names, we generate the distributions of the last name across the different caste categories . We use this distribution to generate a prediction and then combine this with predictions of other models to ensure a robust prediction. We assign a caste to each household based on a simple majority vote between these models.

B. Identification of Laws and Acts

Legal texts do not currently employ any standardised citation style for referring to acts or laws. For example, the Hindu Marriage act may be referred to in a variety of ways: "u/s 13 clause 2 of Hindu Marriage Act," "u/s 13(b) Hindu Marriage Act," or "u/s 13 of Hindu Marriage Act 1995". Again, ML tools can also be used to address this issue.

In our work, we are using a set of tools that create mathematical representations of the text in the form of vectors. "Term Frequency - Inverse Document Frequency" (TF-IDF) is one such popular method to represent a string of words as a vector of scores that reflects how frequently a word is used within a given text and how infrequently it appears in the corpus.[17] With this representation of the act names as vectors, we use several different unsupervised clustering algorithms for further analysis.[18] These algorithms allow us to convert sentences into mathematical objects and transform the problem of distinguishing sentences into one of measuring distances between these mathematical objects. Put differently, this uses an inductive process to group the underlying data in a manner that best preserves the coherence within groups and the distance across groups in order to make the classification.

Identification of specific laws and acts opens up new opportunities of legal analysis. We can, for example, compare the types of cases across courts, and the time it takes to process them. This is critical in being able to understand where the biggest delays in the system come from, and how they may be resolved.

---

[17] "Term frequency" measures the number of times a term appears in a document. Inverse Document Frequency refers to the log(N/D), where N is the total number of documents, and D is the number of documents that contain that specific term. The TF-IDF is the product of these two terms.

[18] Examples of this are the "agglomerative hierarchical clustering" and "k-means" algorithms. The latter first assigns "K" random data points (TF-IDF vectors generated from text of acts in our case) as cluster centroids (or means) and then categorises rest of the data points to the means closest to them. After this, a new position of the centroid (mean) is calculated, taking the average of the data points categorised in that mean. This process goes on iteratively till the centroid stops moving, or say is at the centre of the final cluster.

C. New interpretations of text

The ideology of judges, and how this affects their rulings on specific cases, has fascinated legal scholars for a long time.[19] Empirical estimates of ideological imprints on cases and rulings, however, are difficult to determine. Recent work on natural language processing and computational linguistics makes this a new possibility.[20] Instead of looking at clusters of specific words as discussed above for acts and laws, algorithms within a class called 'neural networks' can examine the contexts in which words are used and produce representation of these words as vectors such that words used in similar contexts are close together in this vector space. Interestingly, mathematical relations between these vectors encode semantic relations between the encoded words. For example, Word2Vec is one such algorithm that "learns" conceptual relations between words. A trained model can produce synonyms, antonyms, and analogies for a given word. These vectors, often referred to as "word embeddings" can be used to identify structure in the data and make predictions. More recently, "document embeddings" have built upon the success of word embeddings to represent words and documents in a joint geometric space.[21] Like word embeddings, these document embeddings can be used to interpret and classify large volumes of text.

The application of these algorithms on a corpus of cases allows for the identification of important patterns. In the United States for example, they uncover the distinctive ideology of judges, and the variations in this ideology based on birth cohort, partisan affiliation, and/or legal training.[22] For the embeddings approach to the citation network, we can identify similar cases based on how often they are cited together. A more intriguing example is identifying legal analogies through document embeddings. Word embeddings have been documented to know that man is to woman as king is to queen, from the way the four words are used in the English language. A document embedding might identify when the latter case is a related application of a legal principle articulated in the former case.

D. Identification of Discrimination and Bias

Concerns about stereotyping and discrimination in the courts of India typically cite a single case or small set of cases to highlight the issue. Bias shown

---

[19] GH Gadbois Jr, *Judges of the Supreme Court of India: 1950–1989* (OUP 2011); GH Gadbois, 'Indian Supreme Court Judges: A Portrait' (1969) 3 Law & Society Review 317-336; T Mikolov and others, 'Distributed Representations of Words and Phrases and their Compositionality' (2013) 26 Advances in Neural Information Processing Systems 3111-3119.

[20] Q Le and T Mikolov, 'Distributed Representations of Sentences and Documents' (2014) 32(2) Proceedings of the 31st International Conference on Machine Learning 1188-1196.

[21] E Ash and DL Chen, 'Case Vectors: Spatial Representations of the Law Using Document Embeddings' in M Livermore and D Rockmore (eds), *Law as Data: Computation, Text, & the Future of Legal* (Santa Fe Institute Press 2019).

[22] D Kahneman, *Thinking, fast and slow* (Farrar, Straus and Giroux 2011).

by a court, or even a single judge, is difficult to identify and analyse rigorously. This challenge is of course not unique to judiciaries. Many papers in the academic literature have demonstrated that bias by a human decision-maker can have conscious as well as unconscious drivers and may manifest in complex ways than can be difficult to prove in a variety of contexts.[23] In other settings, such as labor markets and educational institutions, algorithms — rules that take "inputs" (like the characteristics of a job applicant) and predict some outcome (like a person's salary) —have been shown to create new forms of transparency and serve as methods to detect discrimination.[24]In the courts of India, algorithms could help judges make some critical decisions about cases (for example, dismissals or bail applications). They could also help courts evaluate a judge's performance.

Building such algorithms requires a feature-rich dataset typically consisting of variables that include litigant characteristics (caste, gender, location, type of crime committed), lawyer characteristics, court characteristics, case details (filing details and evidence provided), additional variables (day, month, year, weather, etc.) and case outcomes (such as granting of bail or dismissal of a case). An algorithm builder would write a "learning procedure" that would aim to provide a predicted outcome from a broad range of inputs, choosing from a variety of models like support vector machines, decision trees, Bayesian networks and neural networks. The architectures of these models feature multiple sequential layers of intermediate variables that connect input features and output classes, so that the outputs of one layer serve as inputs of the next layer.[25] These models stand in sharp contrast to traditional statistical methods such as linear regression, which is more deductive (presuming a linear fit between a few sets of variables) than inductive (allowing the data to report the best fit between a large set of variables).

These insights could be invaluable not only within the court room itself, but also in judicial education. Experiments are currently underway, in the Judicial Academy of Peru for example, to assess methods to improve case-based teaching by using the history of a judge's past decisions — which can

---

[23] A Banerjee and others, 'Labor Market Discrimination in Delhi: Evidence from a Field Experiment' (2009) 37(1) Journal of Comparative Economics 14-27; M Bertrand and S Mullainathan, 'Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination' (2004) 94(4) American Economic Review 991-1013; M Ewens, B Tomlin and LC Wang, 'Statistical Discrimination or Prejudice? A Large Sample Field Experiment' (2014) 96(1) The Review of Economics and Statistics 119-134; J Kleinberg and others, 'Discrimination in the Age of Algorithms' (2018) 10 Journal of Legal Analysis 113.

[24] J Kleinberg and others, 'Algorithms as Discrimination Detectors' (2020) 117(48) Proceedings of the National Academy of Sciences 30096; Y LeCun, Y Bengio and G Hinton, 'Deep Learning' (2015) 521(7553) Nature 436-444.

[25] DL Chen, M Ramos and S Bhupatiraju, 'Data and Evidence for Justice Reform (DE JURE)' (Development Impact Evaluation (DIME) Group, World Bank, November 2019).

reveal potential bias or error.[26] The data is also suitable for creating personalised dashboards and interfaces that provide judges, mediators and other decision-makers with real-time information about their own performance relative to their own previous decisions and others who are comparable.[27] This information can be used to augment the capabilities of judges and lawyers, increase their productivity and devote greater attention to complex cases.

E. Identification of Causal Effects of Legal Rulings

Legal scholars and judges have long made arguments for or against the implementation of various laws and regulations and justified their arguments with theories about the effects of these legal rules. This situation resembles the field of medicine a century ago — prior to the advent of clinical trials, there were only theories without rigorous causal evidence.

A growing body of empirical research now demonstrates that causal inference is possible in judicial studies. For example, in situations where cases are randomly assigned to judges, the random assignment itself can be used as an exogenous source of variation to evaluate the impact of judicial decisions — since judges in this scenario do not choose their cases and end up with them randomly, observed rulings reflect their deliberations in the case rather than the process of justice that led them to be assigned the case.

Randomising cases to judges predicted to be harsh or lenient generates the inference on the long-run causal impacts of the length of sentences .[28] To get at the causal effect of a sentence length of eight months or eight years, a randomised control trial would need to randomise the sentence. However, assigning a defendant to a judge predicted to assign eight months or another judge predicted to assign eight years sentence length generates the causal impact of sentence length on subsequent life outcomes. The same framework can examine the causal effects of debt relief on individuals' earnings, employment, and mortality.[29] Discretion in decision-making sheds light on a myriad of topics where judicial habits exist. This includes the decision to protect patent rights [38]. Machine learning can be used to create predictions of judges, which can then be used to estimate the causal effects of their predicted decisions on long-term outcomes.

---

[26] JR Kling, 'Incarceration Length, Employment, and Earnings' (2006) 96(3) American Economic Review 863-876.

[27] ibid.

[28] W Dobbie and J Song, 'Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection' (2015) 105(3) American Economic Review 1272-1311.

[29] B Sampat and HL Williams, 'How do Patents Affect Follow-on Innovation? Evidence from the Human Genome' (2019) 109 (1) American Economic Review 203-236.

IV. A NEW RELATIONSHIP BETWEEN HUMAN AND MACHINE

Thus far, we have argued that ML presents a powerful tool for improving the systems of organising and interpreting the voluminous, unstructured and complex data that has been released from the Indian judiciary over the past fifteen years. Algorithms can be written to draw inferences about the identity of participants and study the deliberative processes they employ within court rooms. ML tools can also convert a high volume of textual data to numerical estimates that can be used for understanding the processes and outcomes of the systems of justice themselves. Analysis of texts for patterns of bias and discrimination, for example, can augment the capabilities of judges and lawyers and systematise processes of review.

These tools, however, have several limitations and requirements that need to be addressed before they can be effectively deployed in the courts. At the very outset, there are significant issues related to the privacy related to personally identifiable information, security, and control of legal data. Next, the algorithms require data pre-processing, training on large and high-frequency datasets, and iterative refinement with respect to the actual cases where they are deployed. This requires strong pilot programs that are studied as part of randomised control trials (RCTs). Insights on data privacy, costs as well as outcomes require these pilots to be constructed on a reasonable scale.

Only technologies that aid humans, rather than replace them, should be adopted in the courts. This is for a variety of reasons. As noted earlier, algorithms have the "black-box" problem of interpretability, i.e., it is not easy to trace the output of complex algorithms to the data inputs themselves.[30] The word-embeddings algorithms discussed earlier, for example, learn biases existing in the corpora. Using these in downstream tasks and decisions without critical oversight raises the risk of replicating these biases elsewhere in the system. The inherent choices of model architecture can also reinforce existing biases by decision-makers within the system. Addressing these issues requires a participatory and deliberative approach towards the design, implementation, and evaluation of the adoption of these technologies.

How may AI and ML methods aid human decision-making? An AI based recommender system might start by offering a judge the best prediction of themselves, based on the specific judge's previous decision-making, from a model using only legally permitted features. At the very least, it can help judges be consistent across similar cases by offering the most relevant reference points—and to cabin the influence of extraneous factors. Deviating from defaults can facilitate more conscious, slow deliberation . Moreover, by

[30] Pasquale (n 12).

showing the judge statistics about oneself, it explicitly leverages motives for judges to want to self-improve.

Showing how other judges might make the decision—based on the machine's model of the other judges—offers yet another way for AI to assist rather than dictate decision-making, this time by creating a customised community of instantly available experts (trained based on data from the behavior of other experts, possibly over time and across geographic and subject matter contexts). Showing statistics of other judges may also leverage self-image concerns of being a better judge that can motivate moral behavior. To be sure, this can lead to group think, though research in the US finds that judges have a strong sense of identity and want to be different from the others. Depending on a judge's tendency to conform or the opposite in a manner deleterious to effective justice, this aspect of the AI system can be toggled on or off.

A predictive algorithm to detect judicial error might nudge judges into more appropriate decisions. The predicted errors allow for analysis of the extent to which systematic factors affect the predictability of judges' errors. By predicting error, we might highlight areas in which decision-makers potentially need more decision support. Of course, it is always possible that the AI system's suggestion would fail to take into account some reliable private information that the judge might have access to. Where this happens, the AI system would be steering the judge off course rather than correcting for their inconsistencies. Therefore, a conversation between the judge and the AI can be encouraged so that the AI can learn this private information from the judge as well.

A reasonable demand to guarantee trust and fairness is that algorithms be interpretable. A judge may request a reason for why the deviation may lead to mistakes. Such an incremental integration leverages judges' normative commitments and self-perceptions of being a good judge to facilitate the adoption of these systems.

Given the complexities of working with AI and ML algorithms, it is essential that any roll out be preceded by a phase of comprehensive study and rigorous testing of the systems themselves. Randomised controlled trials that carefully estimate the causal impacts of the adoption of these algorithms to properly evaluate their costs and benefits are essential. A carefully constructed trial can provide important benchmarks on cost, efficiency, user satisfaction and outcomes, all essential for a justice system to credibly serve citizens. Overall, we believe that when leveraged in a careful and ethical way, AI and ML offer significant potential for the courts of India.