Artificial Intelligence and the Courts: MATERIALS FOR JUDGES

Artificial Intelligence, Legal Research, and Judicial Analytics

SEPTEMBER 2022

AAAS
AMERICAN ASSOCIATION FOR
THE ADVANCEMENT OF SCIENCE

Artificial Intelligence and the Courts: Materials for Judges

The American Association for the Advancement of Science (AAAS) is honored to have been entrusted, by the National Institute of Standards and Technology (NIST), with the task of developing educational materials on artificial intelligence (AI) for judges and courts.

AAAS therefore offers this compilation of educational materials for judges, covering a wide, yet appropriate, set of issues. (Please see the list below). AAAS' goal is to provide a set of user-friendly and accurate, yet readily comprehended, definitions, analyses and perspectives, on a variety of terms and topics with which the judiciary ought to become familiar.

The materials contained herein were developed by teams of scientific and legal experts who focused on a particular topic. The topics considered worthy of inclusion were selected based both on the mandate provided by NIST and guidance received by AAAS from an Advisory Committee composed of a large and diverse group of legal and AI experts. Drafts of the materials were subsequently submitted to Advisory Committee members, and outside expert "Reviewers," to obtain any suggestions for adjustments before each team of authors finalized their contribution (paper, podcast, annex, etc.).

It is not expected that courts will become experts regarding these sometimes complex or technical matters. Rather, this collection presents facts and overviews in a manner intended to make judges aware of key issues and to enable courts to find useful information contained herein, easily.

Finally, it is hoped that courts will appreciate certain innovative elements of this product, notably the inclusion of podcasts. These will provide courts with facts and analysis of important questions in a format that courts may find agreeable and, given the accompanying transcripts included, useful. AAAS thanks NIST for allowing a team of experts to undertake this forward-leaning approach to providing courts with needed information and insights as part of this project.

Materials in this series include:

1. ***Artificial Intelligence – Foundational Issues and Glossary***
2. ***Artificial Intelligence and the Justice System*** **(Podcast Series and Transcripts)**
   - Episode 1: *AI and Risk Scores* (49 minutes)
   - Episode 2: *AI in the Legal Field – Commercial and Unexpected Uses* (70 minutes)
   - Episode 3: *AI, Decision-Making, and the Role of Judges* (58 minutes)
3. ***Artificial Intelligence, Trustworthiness, and Litigation***
4. ***Artificial Intelligence, Legal Research, and Judicial Analytics***
5. ***Artificial Intelligence and Bias – An Evaluation***

Acknowledgements

The American Association for the Advancement of Science (AAAS) gratefully acknowledges the authorship of John Facciola, Adjunct Professor of Law at Georgetown University School of Law and Judge (ret.), US District Court for the District of Columbia, and Nicholas Mignanelli, Research and Instructional Services Librarian at the Lillian Goldman Law Library and Lecturer in Legal Research, Yale University School of Law.

We also extend our gratitude to the members of the Advisory Committee for their input and guidance throughout the design and production of this product: Rediet Abebe, Assistant Professor, Computer Sciences, University of California, Berkeley; Theodore F. Claypoole, Partner, Womble, Bond, Dickenson, LLP; John Cooke, Director, Federal Judicial Center; Cynthia Cwik, Senior Legal Advisor, Former Stanford Fellow, and Former Partner, Jones Day and Latham & Watkins; Herbert B. Dixon, Jr., Judge (ret.), Superior Court of the District of Columbia, and Senior Legal Advisor to the Center for Legal and Court Technology, William & Mary School of Law; Finale Doshi-Velez, Associate Professor, Computer Science, School of Engineering and Applied Sciences, Harvard University; Paul W. Grimm, Judge, US District Court for the District of Maryland; James Harris, Technology Consultant, National Center for State Courts; Joseph Sawyer, Director, Faculty Development and Online Learning, National Judicial College; Elham Tabassi, Chief of Staff, Information Technology Laboratory, National Institute of Standards and Technology; and Zachary Zarnow, Principal Court Management Consultant, National Center for State Courts.

AAAS appreciates the contributions of Solomon Assefa, Vice President, IBM Research; Nicole Clark, Chief Executive Officer, Trellis Research; Rebecca Crootof, Associate Professor of Law, University of Richmond School of Law; Patrick Huston, Brigadier General and Assistant Judge Advocate General for Military Law and Operations, US Department of Defense; Maia Levy-Daniel, Attorney and Public Policy Consultant; Matthew Stepka, Managing Partner, Machina Ventures, and Lecturer, Haas School of Business, University of California, Berkeley; Peter Stone, Professor, Department of Computer Sciences, and the Founder and Director, Learning Agents Research Group, Artificial Intelligence Laboratory, University of Texas, Austin; Rebecca Wexler, Assistant Professor of Law and the Faculty Co-Director, Berkeley Center for Law & Technology, University of California, Berkeley; and others for their review and insightful comments on previous drafts of the *Artificial Intelligence and the Courts: Materials for Judges* series.

This paper was edited by Joel Ericsen, Program Associate, and Alain Norman, Senior Program Associate, Science and the Law, AAAS. Special thanks to Theresa Harris, Program Director, Scientific Responsibility, Human Rights and Law Program, and Nathaniel Weisenberg, Program Associate, AAAS, and to Barbara "Bebe" Holloway, University of Virginia, for their contributions.

AAAS Creative and Brand Manager, Paula Fry, created the cover design for this publication.

The production of this publication and other materials in the *Artificial Intelligence and the Courts: Materials for Judges* series was funded by the United States National Institute of Standards and Technology (NIST) through Award 60NANB21D031.

Contact: AAAS welcomes comments and questions regarding its work. Please send information, suggestions and any comments to the AAAS Scientific Responsibility, Human Rights and Law Program at srhrl@aaas.org.

Cite as: Facciola, J. and Mignanelli, N. "Artificial Intelligence, Legal Research, and Judicial Analytics." *Artificial Intelligence and the Courts: Materials for Judges*. American Association for the Advancement of Science (2022). Available at https://doi.org/10.1126/aaas.adf0787.

Abstract

This paper touches on two matters—often overlooked—but of potentially very direct significance to judges: How artificially intelligent systems (AI) may affect the process and results of legal research, and whether or how assessments (i.e., "judicial analytics") of judges' prior rulings, decisions, or even style, might be leveraged by parties' counsel to gain some advantage. That is, key-word searches, or other forms of "technology assisted review," may be more or less effective, or be affected by, the way search-engines are designed. Separately, but increasingly, vendors offer products—based on analyses of judges' behaviors and rulings—that purport to provide insights that will, in turn, reduce risks for litigants or parties. Finally, it may also be that AI will be able to usefully shed light on whether or when reforms to rules or procedures have proven to be effective in improving the administration of justice.

Table of Contents

Artificial Intelligence, Legal Research, and Judicial Analytics

Authors:

John Facciola
Adjunct Professor of Law, Georgetown University School of Law
Judge (ret.), US District Court for the District of Columbia

Nicholas Mignanelli
Research and Instructional Services Librarian, Lillian Goldman Law Library
Lecturer in Legal Research, Yale University School of Law

1. Introduction

In recent years, **artificial intelligence**[1] (AI) has become a buzzword in the legal industry, with countless articles and blogposts heralding its arrival and touting the manifold ways in which it promises to radically transform law practice. But what is AI? AI is simply a stage of technological development in which computers are able to perform tasks that have traditionally required human intelligence. AI can be divided into two categories: rule-based AI and **machine learning** (ML). In the rule-based approach, processes are automated through the use of rules coded by humans. Machine learning, on the other hand, involves computers identifying patterns in data and creating rules accordingly.

Where legal research is concerned, machine learning allows for natural language searching and the retrieval of relevant documents as determined by the behavior of prior users, among other factors, in accordance with the **algorithms** (instructions) created by programmers. Yet, the details of these instructions and how they operate often exist in a metaphorical "**black box**," meaning that they are inaccessible to users and, in some cases, even to programmers. Accordingly, while a user can see the input and the output, he or she cannot see why a particular input leads to a particular output. For instance, when an attorney runs a keyword search in a legal database, he or she cannot know why certain results appear while others do not, nor why the results appear in the order that they do.

Furthermore, "every database has a point of view." In a 2017 article, Professor Susan Nevelow Mart compared the top 10 results of 50 searches performed in the same dataset across six legal databases.[2] She found that the results differed dramatically from database to database. This demonstrated that what a researcher finds in the process of searching depends heavily on who builds the search algorithm and what choices the builder (programmer) makes in the process.[3] Thus, the assumptions and biases of the programmers are built into the search algorithms humans write, and the biases of past researchers are incorporated into the user data that

[1] Bolded red terms appear in the Glossary.

[2] See "Further Readings" for reference, above.

[3] Algorithms can encompass training algorithms as well as traditional rule-based algorithms. Note, too, that this example (of Mart) involved searching the same dataset in all six databases; this shows that even when a dataset is the same, results can be different!

algorithms rely upon. This is especially troubling because the programmers typically hail from homogenous groups of people with particular incentives, and past researchers consist of those who could afford access to the expensive legal research platform in question.[4] This phenomenon is called "**algorithmic bias**."

A related but distinct development in legal research technology is found in the emergence of so-called "**judicial analytics**." Vendors create judicial analytics products by culling data from publicly available dockets and using that data to identify patterns in the way a particular judge has ruled in the past. These patterns can then be used by an attorney to attempt to predict the future behavior of the judge he or she is appearing before. Using judicial analytics products, attorneys can learn how likely a judge will be to grant their motion, the average amount of time a judge takes to decide cases like theirs and what sources a judge typically relies upon in deciding similar cases.

Although attorneys, especially those employed by large law firms, are drawn to these expensive products by the promise of risk mitigation, the accuracy of judicial analytics depends on the availability and comprehensiveness of the dockets from which the data are being culled. Additionally, these products have raised serious ethical questions about equity and access to justice, with critics claiming that only wealthy individuals and entities will be able to retain attorneys who can afford access to these products. In fact, France has criminalized the use of certain judicial analytics.[5]

AI and judicial analytics are said to be revolutionizing the practice of law. Such assertions should, however, be met with a healthy dose of skepticism. As it remains unclear how these new technologies will change law practice, it is important to understand the basics of how these new technologies work and what their shortcomings are.

---

[4] The question may arise: Are the biases of programmers and past researchers likely to result in certain judicial precedents being ranked higher in search results than others of similar authority, thereby possibly burying good law? The authors hold that technology is still developing in this area and, so, it would be inappropriate to engage in conjecture about what this might mean for the results that "AI-powered" legal research tools provide researchers. Nevertheless, it is important to raise awareness of the "algorithmic bias" phenomenon.

[5] More specifically, France prohibits the use of certain "identifying" data in efforts to "predict the actual or supposed professional actions/practices [of judges or court personnel]." The relevant lines of French law may be translated as stating, "Data identifying magistrates (judges) and personnel of the court cannot be (re)utilized in order to effectuate evaluation, analysis, comparison of, or to predict, their actual or supposed professional actions/practices." (In French, the pertinent lines read: "*Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées[.]"* See: Section, III, Article 33 (as amended, 2019) of French law (pertaining to the judiciary), entitled "*Concilier la publicité des décisions de justice et le droit au respect de la vie privée"* ["Conciliating the Publication of Judicial Decisions and the Right to Respect for Private Life"], at: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000038261761.

2. The Pragmatic Use of Predictive Software

2.1. AI and Research, Generally

For as long as there have been lawyers and more than one court, lawyers have always tried to find a court that will favor their client. Whether it is the judicial district on the federal side, the judge, or the potential jury pool, lawyers pride themselves on their ability to "forum shop." Although lawyers' efforts are usually no more scientific than calling a friend and asking about the proclivities of a certain judge, lawyers trust shared information implicitly, even though it may be nothing more than warmed-over courthouse gossip.

Something new has been added. Software engineers are creating programs that will predict the outcome of a case based on the data that is analyzed by the program. There is nothing revolutionary about computer programs that seek to predict what humans may like, and/or do. It is the same technology that the music streaming service, Pandora, uses to tell us that if we like Bach or Drake, there is a high probability that we will also like Mozart and Bruno Mars. Indeed, fortunes are made because the software can predict with supposed accuracy whether the quarterback's next pass will result in a touchdown.

Persons of a certain age may remember when the television camera panned the dugout and caught the pitcher stuffing a wad of tobacco in his mouth. Now, it shows us the next batter scanning his iPad to calculate the likelihood of getting a hit in his next at-bat if he does not swing at a curveball.

All of this is nothing more than the now nearly universal use of the algorithm.

In a recent opinion, Judge Grimm said of algorithms:

> Algorithms are not omniscient, omnipotent, or infallible. They are nothing more than a systematic method of performing some particular process from a beginning to an end. Paul W. Grimm, et al, *Artificial Intelligence as Evidence*, 19 NW. J. Tech & Intell. Prop. 9. 11 (2021) (defining algorithm). If improperly programmed, if the analytic steps incorporated within them are erroneous or incomplete, or if they are not tested to confirm their output is the product of a system or process capable of producing results… then the results they generate cannot be shown to be relevant, reliable, helpful to the fact-finder, or to fit the circumstances of the particular case in which they are used.[6]

[6] In re: Marriott Int'l Inc., Customer Data Security Search Litigation, CV No. 19-MD-1879 (D. Md. May 3, 2022) slip op. at 31.

Algorithms are as present in our lives as rain and sunshine. Judges and lawyers are becoming more comfortable and familiar with them in one significant aspect: their use in civil discovery.

As technology transformed the creation of data by American businesses, lawyers found that their clients had mountains of data. Unfortunately, few of these corporations had information governance policies. Instead, most of the data mountains consisted of high hills of junk. Buried with the crucial data were invitations to the holiday party, an announcement of softball practice and the entries for the office-wide NCAA March Madness bracket challenge.

Lawyers were confronted with discovery demands that would force them to review what could be as much as a terabyte or even a petabyte of this co-mingling of the crucial and the inane using nothing more than their eyes and a pencil. Unfortunately, not too many lawyers have the life expectancy to read a petabyte of data.[7] In what may have been desperation, they turned to information science.

There was already in existence a scientific discipline devoted to creating mechanical systems for finding pertinent data in large datasets. That discipline yielded what the lawyers called "**technology assisted review**" (TAR). That term meant the use of algorithms to define and then narrow searches. The lawyer, aided by the algorithm, searched the data using different methodologies.

One was search terms, words that lawyers thought should appear in the documents they were searching for. A more refined method earned the title of "machine learning." This is not to suggest that a machine "learns" the way a human being learns. A machine is not conscious of the increase in its capacity to identify a pertinent document. Instead, we should say that using a well-defined algorithm and an iterative process will persistently refine a search to the point where a lawyer can defend the process used as a reasonable means of finding the documents demanded by their opponent's discovery demand.

Two terms are helpful here, recall and precision *which are used universally no matter the means of sampling a large accumulation of data.* **Recall** is "the fraction of Relevant Documents that are identified as Relevant by a search or review effort." **Precision** is the fraction of Documents identified as Relevant by a search or review effort, that are, in fact Relevant.[8] As the authors of that Glossary point out in their definition of another term, precision and recall work inversely:

[7] A "terabyte" has been defined as follows: "A terabyte (TB) is a unit of digital data that is equal to about 1 trillion bytes. In decimal notation (base 10), a terabyte is exactly 1 trillion bytes. In binary notation, a terabyte is equal to 2^{40} bytes, or 1,099,511,627,776 bytes. The terabyte is typically used as a measure for storage capacity or the amount of stored data." (At: https://www.techtarget.com/searchstorage/definition/terabyte). A petabyte has been defined as: "One petabyte (abbreviated "PB") is equal to 1,000 terabytes and precedes the exabyte unit of measurement. A petabyte is slightly less in size than a pebibyte, which contains 1,125,899,906,842,624 (2^{50}) bytes." (At: https://techterms.com/definition/petabyte).

[8] "The Grossman-Cormack Glossary of Technology-Assisted Review," with a Foreword by John M. Facciola, 7 Fed. Cts. L. Rev. 25-326.

- **Precision-Recall Tradeoff**: The notion that most search strategies can be adjusted to increase Precision at the expense of Recall, or vice versa. At one extreme, 100% Recall could be achieved by a search that returned the entire Document Population, but Precision would be low (equal to Prevalence). At the other extreme, 100% Precision could be achieved by a search that returned a single Relevant Document, but Recall would be low (equal to 1/N, where N is the number of Relevant Documents in the Document Population). More generally, a broader search returning many Documents will have higher Recall and lower Precision, while a narrower search returning fewer Documents will have lower Recall and higher Precision.[9]

Given this tradeoff, information scientists and lawyers who use their discipline, attempt to achieve what they call **F_1**, defined by the authors of the Glossary as follows:

- **F_1**: The Harmonic Mean of Recall and Precision, often used in Information Retrieval studies to measure the effectiveness of a search or review effort, which accounts for the tradeoff between Recall and Precision. In order to achieve a high F_1 score, a search or review effort must achieve *both* high Recall and High Precision.[10]

Lawyers might use this harmonic mean of F_1 to defend their search process against their opponent's attack that they did not find everything they were supposed to in response to the opponent's discovery demand.

The difference in the method used by the human being and the machine is crucial. The human being looks at the data and decides its relevance to a discovery demand. The machine is using what Grossman and Cormack, the authors of the Glossary, call **continuous active learning**. The human being puts into the computer, let us say, 100 documents. The human being then examines the 100 documents now captured in the computer's memory. The human being finds 20 that appear relevant and now programs the computer to find more like the 20. The computer then finds more like the 20 and produces, let us say, another 40. The human being eliminates the irrelevant ones and programs the computer to find more like the relevant ones. This process continues until what we can call F_1 is achieved.

Scientific analysis of this search process has yielded significant information about this difference. First, science establishes that human beings search large databases poorly. In one study called "Blair Marrion," the lawyers searched a large database based on a train derailment. They estimated that they had found about 75% of the documents they were looking for. In fact, they found 20%; this meant that, on average, the "STAIRS" system[11]—being assessed in this

[9] *Id*. at 26.

[10] Id. at 16.

[11] STAIRS is "IBM's full-text retrieval system, STAIRS. STAIRS, an acronym for "STorage And Information Retrieval System," is a very fast, large-capacity, full-text document-retrieval system[.]" See: David C. Blair & M.E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, 28 COMMC'NS ACM 289 (1985), at pg. 289.

study – could be used to retrieve only 20% of the relevant documents, whereas the lawyers using the system believed they were retrieving a much higher percentage (i.e., over 75%).[12] [13]

On the other hand, thanks to several studies sponsored by the National Institute of Standards and Technology (NIST), Grossman and Cormack, the authors of the Glossary quoted above, established indubitably that machine learning or technology assisted review can achieve results at least as good as what human beings may, yet at a fraction of the time and expense.[14] Indeed, Grossman has specifically advised this paper's authors that the Grossman/Cormack team, and other researchers, have used technology assisted review (TAR) in hundreds of matters, and that courts have taken judicial notice, frequently, of the accuracy of the results, and those of similar studies, dealing with many different matters.[15]

To lawyers handling **big data** cases, this technology-assisted review process is becoming second nature. Judges are now spending their time assessing the validity of the searches conducted by that process.

---

[12] David C. Blair & M.E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, 28 COMMC'NS ACM 289 (1985).

[13] More specifically, as summarized in a subsequent paper: "The STAIRS study described the design, execution and analysis of a large scale, search and retrieval experiment aimed at evaluating the effectiveness of a simple full-test retrieval system. The study examined and evaluated IBMs full-text retrieval system STAIRS as used in a litigation support situation. The STAIRS database contained roughly 350,000 pages of documents which included engineering reports, internal management memos, progress reports, minutes of meetings, etc. The results of this test showed that Recall was, on average, no better than 20% with a 79% mean Precision level. Thus, no more than one in every five relevant documents in the database was retrieved-even though the lawyers using the system were convinced that, after multiple search iterations, they had in fact retrieved over 75% of the relevant documents. These conclusions about the poor Recall of the STAIRS system cannot be contested - they are the facts that the study produced. However, the study went beyond these conclusions and offered two theoretical arguments to support the view that these poor results should have surprised no one. These arguments showed why it would be difficult indeed to obtain higher Recall using a simple full-text retrieval model with a large document database.

In order for a simple full-text system to retrieve effectively, the user/searcher must be able to predict (and use as his query terms) those words, phrases and word combinations that occur in most of the relevant documents, and which do not occur in most of the non-relevant documents. (See also Maron, 1988.) If a searcher can construct such a query, we shall call that an "effective query." We see that there are two interrelated parts to an effective query; predicting A, the words, word combinations, etc., that occur in the relevant documents and then B, reducing that set of terms by excluding those word or word combinations which are likely also to occur in nonrelevant documents."

David C. Blair & M.E. Maron, *Full Text Information Retrieval: Further Analysis and Clarification*, Information Management & Processing, Vol. 26, No. 3, pg. 438 (1990); online at: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/28883/0000719.pdf?sequence=1.

[14] Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual, Review*, XVII RICH. J.L. & TECH. 11 (2011) http://jolt.richmond.edu/v17i3/article11.pdf.

[15] See: Gordon V. Cormack, *Navigating Imprecision in Relevance Assessments on the Road to Total Recall: Roger and Me*, *SIGIR 17, August 7-11*, available at: *https://dl.acm.org/doi/10.1145/3077136.3080812*.

2.2. AI and Judicial Analytics

To return to where we started, information specialists will use the same methodology to make predictions about the results, in a given court, or about one of its judges' rulings.

Such analysts will collect all the data they can from the court records and use machine learning to find the pertinent documents. Having found them, they will use algorithms to analyze what they have found to support their predictions.

Take a case where the lawyer can establish venue in a patent case in either the District of Delaware or the Southern District of Texas. The "prediction" software will capture all the public records about patent cases in both Districts.

Once collected, the lawyer can first use the software to assemble data about the two Districts. How many cases were dismissed by judges on a motion under FR Civ. P. 12(b)(6) for failure to state a claim or grant summary judgment? Which judges seem to have had the largest number of patent cases? How many of the cases involved chemical patents, and how many mechanical? Were there any law firms that were particularly successful?

While gathering that information is important, the real goal of the prediction software is to arm counsel with predictions as to the ultimate result so that counsel can decide where to file the lawsuit.

Therefore, the analysis must shift to the science of statistics and the mathematical derivation of probabilities. Although scientists may use various means to demonstrate the validity of a thesis, Federal courts assess that purported validity by using factors now captured in the Federal Rules of Evidence, notably in Fed. R. Evid. 702. It should be recalled the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579 (1993) was motivated by a desire to eliminate what was called "junk science." Instead, the Supreme Court insisted that "the subject of an expert's testimony must be "scientific… knowledge." The adjective "scientific" implies a grounding in the methods and procedures of science." Id. at 590. It therefore follows that; "In short, the requirement that an expert's testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability." Id.

The authors of this paper are legal academics, including a federal judge who applied F. R. Evid. 702 to the cases before him. As such, they do not pretend to be able to specify how the methodology of that Rule differs from the scientific method that may be applied in a true, laboratory-used, scientific process. They can say, however, that lawyers would default to Rule 702 in assessing the validity of a scientific technique. They are therefore comfortable in using it to assess the likelihood of lawyers' relying upon it.

Accordingly, we then must ask the questions the Rule demands we ask: Is the prediction based on sufficient data? As indicated above, we know that the specter of **bias** haunts artificial intelligence. Was the probability based on an objectively derived dataset? Was that data biased in its inception even if it was objectively collected?[16]

Second, how was the algorithm created? What scientific principles (if any) were relied upon by its creator, and did the creator apply those principles to create the algorithm correctly?

Finally, while it is not mentioned in Fed. R. Evid. 702, the question: Can the algorithm be tested? The answer seems to be no. It is hard to imagine how a prediction of how a judge will rule in a given case can be tested until the judge rules.

Therefore, it will be left to researchers to examine the *bona fides* of prediction software, if they see fit. If it is a commercial product, the market may tell us whether it will survive because lawyers buy it, or not.

Meanwhile, the authors of this paper hope to have shown judges that prediction software should be de-constructed and de-mystified. All a prediction software product does is collect data, using what has become familiar technology. It takes that data and a computer, programmed with an algorithm, and uses math and statistics to predict an outcome.

Judicial analytics is nothing more than that. It is just like predicting whether the quarterback will complete that pass or that the batter will finally hit that curveball.

3. Conclusions

Part A: Putting Analytics to Good Use – Assessing Reforms, Not Judges?

Time will tell whether the supposed ability to predict how a judge will rule will have any practical significance. While we talk about "forum shopping" we must remember that lawyers are not in Walmart. They cannot simply shop for the court they like. Considerations of what lawyers call "venue,"—the statutory constraints on where a lawsuit may be brought, and of establishing a sufficient presence in a state to justify the exercise of personal jurisdiction over the putative defendant—constrain significantly where a lawyer can bring a lawsuit.

Moreover, certain courts, like the federal claims court, only entertain certain types of cases and cases of that type must be filed in that court. Indeed, state and city courts in all but the smallest towns have subdivided their courts into parts that only handle a certain type of a case.

There will therefore be a landlord tenant court, a probate court, a criminal court, a small claims court, a family court and a traffic court. A case that falls within the jurisdiction of one of those courts must be filed in that court. The Chief Judge of the entire court will assign the judges of

[16] It might be asked what bias has to do with the application of FRE 702: The authors would respond that the rule requires an expert's testimony be based on sufficient facts and data, If data is "cherry picked," so that only data supporting the conclusion is used, they are certain that a court would find that to be a violation of the rule and thus the court would rule such expert's evidence inadmissible.

the court to one of those courts for a term. Thus, for instance, Judge Smith will sit in probate court for a year and all probate cases go that judge.

The vast majority of all court cases in America are tried in state court.[17] If those courts are structured in the manner just described, forum shopping is impossible. Predicting how a judge may rule would, then, be an expensive waste of time.

There are multi-national corporations with offices in many American cities. When they are plaintiffs or defendants, the concerns about venue and personal jurisdiction lessen. That means predicting judicial behavior may be more useful to those who believe in it. But that also means that we may be creating another expensive toy for the rich to which the poor do not have access. That is the last thing we need: more inequity in the administration of justice.

The authors do not mean to suggest that the collection of data and its analysis by artificial intelligence cannot be justified. To the contrary: Efforts to assess the impacts of rule or procedural reforms might benefit from the application of the technology and analysis discussed, above. For, courts—unlike businesses—do not usually create data about their operations and analyze such data to improve the courts' practices.

For example, it may take years to amend the Federal Rules of Civil Procedure. Once an amendment is enacted, however, there is no one who collects data indicating whether the aspirations that animated its enactment have been realized. Instead, the best that can be had is anecdotal evidence from lawyers about how the amendment(s) may work.[18]

The practice of law in inherently adversarial. The bar often breaks down into, for example, a plaintiff's bar and a defendant's bar. Comments on a rule amendment will inevitably be colored in favor of the kind of clients they have and their original position of whether the amendment should have been enacted. Naturally, then, those lawyers who wanted the amendment think that it is working well while the ones who did not want it think that it is working poorly.[19]

Surely, there has to be a better way of assessing the impact of new procedures on the court's operation. Artificial intelligence may point the way.

Indeed, there are some fascinating developments in the collection of court data. For instance, scholars at Georgetown are collecting all the data from all the dockets of every court in America

[17] One study reported that in 2020, there were 11, 691, 816 state courts cases. https://www.courtstatistics.org/court-statistics/interactive-caseload-data-displays/csp-stat-nav-cards-first-row/csp-stat-civil. In the same year there were 470, 581 federal court cases. https://www.supremecourt.gov/publicinfo/year-end/2020year-endreport.pdf.

[18] Lexis and West collect opinions, but cannot tell you whether the amendments accomplished what the draftspersons hoped to achieve via some reform. Although it is possible to assess whether a court has been faithful to the purpose of an amendment in an individual case, it is not possible to assess whether all, or some, or none, of all the judges have been.

[19] Again, Lexis and Westlaw tell you what the opinions are. That a judge reached a certain opinion does not tell you anything about what th lawyers think about a Rule amendment.

that have digital dockets.[20] That data—analyzed by artificial intelligence—may yield useful insights about public access to the courts.

For example, states created small claims courts so that litigants who lack the money could resolve their disputes cheaply. But the data seems to show that small claims courts have become default judgment mills where few of the defendants ever show up.[21] Also, the COVID-19 pandemic forced courts to use tele-conferencing. Does the data show that the traffic court—where everyone appears virtually—is a cheaper and wiser alternative to court appearances? Analysis of this data by AI or ML may prove to be useful in finding answers.

Thus, there is good work to be done and it may provide insights into how courts can be organized to permit greater access to them, at a reduced cost. These authors submit that this is a much better use of artificial intelligence than a lawyer using it to determine whether old Judge Jones, who hates motions for summary judgment, will deny this one.

Part B: Maintaining Perspective or Healthy Skepticism?

Finally, although talk of judicial analytics and "AI-powered" legal research products can be intimidating for members of the judiciary who first learned legal research in print or on a terminal, it is hoped that this paper sheds some light on what these newer tools are, and how they work. Also, the authors hope that, as such AI/ML based tools continue to develop, and their use becomes widespread, judges and lawyers will approach them with a healthy dose of skepticism. While researchers should, of course, use these tools to their full advantage, it is important always to remember that they are tools created by humans and so will have many of the limitations that traditional legal research tools do.

---

[20] *See* Georgetown Law, *A Civil Justice Data Commons*, https://www.law.georgetown.edu/news/georgetowns-civil-justice-data-commons-seeks-to-unlock-court-data/ .

[21] PEW, How Debt Collectors Are Transforming the Business of State Courts, https://www.pewtrusts.org/en/research-and-analysis/reports/2020/05/how-debt-collectors-are-transforming-the-business-of-state-courts.

Annex A: Further Readings

Paul D. Callister, *Law, Artificial Intelligence, and Natural Language Processing: A Funny Thing Happened on the Way to My Search Results*, LAW LIBRARY JOURNAL Vol. 112:2 [2020-6], https://www.aallnet.org/wp-content/uploads/2021/02/01_LLJ_112n2_callister.pdf.

Center for the Legal Profession, Harvard Law School, *Data in Courts: Judicial Analytics in Practice*, 8 Prac., no. 3, Mar./Apr. 2022, https://thepractice.law.harvard.edu/article/data-in-the-court/.

Lauri Donahue, *A Primer on Using Artificial Intelligence in the Legal Profession*, JOLT Digest (Jan. 3, 2018), https://jolt.law.harvard.edu/digest/a-primer-on-using-artificial-intelligence-in-the-legal-profession.

Carolyn Elefant, Part I: The Basics-What is AI?, myshingle.com (July 19, 2019), https://myshingle.com/2019/07/articles/future-trends/part-i:-the-basics---what-is-ai/.

Rick Merrill, *Top 5 most important judicial analytics metrics*, A.B.A. J. (Oct. 11, 2019), https://www.americanbar.org/groups/journal/articles/2021/top-5-most-important-judicial-analytics-metrics/.

Susan Nevelow Mart, *Results May Vary in Legal Research Databases*, A.B.A. J. (Mar. 1, 2018), https://www.abajournal.com/magazine/article/results_vary_legal_research_databases .

Jason Tashea, *France Bans Publishing of Judicial Analytics and Prompts Criminal Penalty*, A.B.A. J. (June 7, 2019), https://www.abajournal.com/news/article/france-bans-and-creates-criminal-penalty-for-judicial-analytics.