

Michigan Law Review

Volume 119 | Issue 6

2021

Natural Language Processing for Lawyers and Judges

Frank Fagan

EDHEC Business School (France)

Follow this and additional works at: <https://repository.law.umich.edu/mlr>



Part of the [Computer Law Commons](#), [Legal Profession Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

Frank Fagan, *Natural Language Processing for Lawyers and Judges*, 119 MICH. L. REV. 1399 (2021).
Available at: <https://repository.law.umich.edu/mlr/vol119/iss6/17>

<https://doi.org/10.36644/mlr.119.6.natural>

This Review is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

NATURAL LANGUAGE PROCESSING FOR LAWYERS AND JUDGES

Frank Fagan*

LAW AS DATA: COMPUTATION, TEXT, & THE FUTURE OF LEGAL ANALYSIS. Edited by *Michael A. Livermore* and *Daniel N. Rockmore*. Santa Fe: Santa Fe Institute Press. 2019. Pp. xxx, 490. Cloth, \$32.99; paper, \$13.99.

INTRODUCTION

If written forty years ago, *Law as Data: Computation, Text, & the Future of Legal Analysis* would have surely contained several chapters on how to code a computer to evaluate facts and generate a legal decision. The earlier fascinations of legal academics with computers focused heavily on replacing lawyers and judges, and tracked much of the early enthusiasm of computer scientists for developing a general artificial intelligence (AI) that could replicate the human mind.¹ Work on that project continues,² much like legal scholars continue to develop automated forms of rulemaking and adjudication.³ Nonetheless, there is today widespread recognition that less ambitious

* Associate Professor of Law, EDHEC Business School (France). Email: frank.fagan@edhec.edu.

1. See ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 18–20 (Stuart Russell & Peter Norvig eds., 3d ed. 2010) (describing the early enthusiasm and great expectations between 1952 and 1969 for “general problem solver[s]” that were designed to implement fundamental human problem solving “protocols”). That era was followed by a dose of reality between 1966 and 1973, when computer scientists broadly recognized that the earlier generation had failed to come to grips with the “combinatorial explosion,” especially of hypotheses, that plagues AI to this day. *Id.* at 20–22. On this last point, see PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 73 (2015).

2. See, e.g., Steve Lohr, *A.I. Is Doing Legal Work. But It Won’t Replace Lawyers, Yet.*, N.Y. TIMES (Mar. 19, 2017), <https://www.nytimes.com/2017/03/19/technology/lawyers-artificial-intelligence.html> [<https://perma.cc/G2N3-EBS7>].

3. Consider that smart contracts feature contingent or “automated” enforcement much like covenants in loan agreements. Max Raskin, *The Law and Legality of Smart Contracts*, 1 GEO. L. TECH. REV. 305, 309, 323 (2017). Legislation can be contingent, too, as when one set of rules overrides another when something probabilistic happens. See Eric A. Posner, *Introduction* to THE TIMING OF LAWMAKING 3 (Frank Fagan & Saul Levmore eds., 2017). The automated-law project has more in common with the phrase “Code is Law,” as opposed to “Law as Data.” In the former, the underlying architecture of the legal system regulates citizens. The tagged phrase comes from LAWRENCE LESSIG, CODE VERSION 2.0, at 1 (2006), which describes how the hardware and software of the Internet regulate the actions that take place in cyberspace. The general idea is discernible in French postmoderns such as Michel Foucault, who

applications of artificial intelligence can improve life and legal practice. General AI is not needed. With bigger stores of data, concrete advances in life and law can be gained, even with “weak” AI.⁴

Against this historical background, Michael Livermore⁵ and Daniel Rockmore⁶ construct an important collection of chapters that, taken together, ask the reader to consider the relevance of data for law. Common lawyers, of course, are deeply familiar with the practice of sifting through case law and distilling legal rules, and the act of matching and distinguishing cases can be readily understood as a form of textual classification. In some way, law has always been data, and lawyers have always done the work of today’s machines. This is obvious. After all, artificial intelligence mostly does the things that humans have done for years. It is when the machine does something more, such as discover a connection between a fact and a legal outcome previously hidden from humans for centuries, that makes us stand up and take notice.⁷ Livermore and Rockmore clearly wish the reader to consider what these advances mean for law.

For lawyers and judges, and perhaps non-empirically minded legal academics, there is an uncertainty or skepticism that pervades this consideration because, among other things, data science has not (yet) transformed legal practice broadly. Humans still try cases and render judgments. Lawyers can be observed, relatively unaided by machines, making strategic decisions in litigation or transactions. Judges rule on disputes or questions of law without additional assistance than that provided by their law clerks and secretaries.⁸ This is true despite significant advances in computer-assisted

describe “normalizing” or “regulatory” forces emanating from many things, including the layout of cities, the architecture of buildings, the classification of psychiatric disorders and disease, and of course, the classification of right and wrong. See, e.g., MICHEL FOUCAULT, 2 THE HISTORY OF SEXUALITY: THE USE OF PLEASURE 12 (Robert Hurley trans., Vintage Books 1990) (1985); see also Gerald Turkel, *Michel Foucault: Law, Power, and Knowledge*, 17 J.L. & SOC’Y 170, 185 (1990).

4. This Review uses artificial intelligence as a general term to encompass human tasks carried out by machines. Weak AI sometimes refers to a machine with the capacity to carry out a specific or narrow set of tasks. General AI, sometimes referred to as strong AI, refers to a machine with the capacity to carry out an open-ended set of tasks while harnessing a full range of cognitive abilities, much like a human. See ARTIFICIAL INTELLIGENCE, *supra* note 1, at 22–27. Machine learning, an important subset of AI, refers to the ability of a machine to learn on its own with some amount of guidance from a human. *Id.* at 2.

5. Edward F. Howrey Professor of Law, University of Virginia School of Law.

6. William H. Neukom 1964 Distinguished Professor of Computational Science, Dartmouth College.

7. For example, Macey and Mitts examine 9,380 veil-piercing cases and find evidence that courts are more likely to veil-pierce when doing so furthers a federal statutory purpose. Jonathan Macey & Joshua Mitts, *Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil*, 100 CORNELL L. REV. 99, 115 (2014). Previous theories focused elsewhere. *Id.* at 104–10.

8. Although the use of algorithmic risk assessment in sentencing and bail decisions is growing. See Brandon Garrett & John Monahan, *Assessing Risk: The Use of Risk Assessment in Sentencing*, JUDICATURE, Summer 2019, at 42, 43; John Logan Koepke & David G. Robinson,

search for evidence (e-discovery) and legal rules (Lexis and Westlaw).⁹ On the other hand, one cannot deny machine learning's growing influence in law as evidenced by, among other things, increased entrepreneurial activity and investment in automated legal technology.¹⁰ Nonetheless, things still feel new, and the scope of future transformation remains unclear for legal practitioners.

I. NLP RESEARCH FOR LAW

A. *Prediction vs. Causality*

The chapters contained within *Law as Data* successfully reduce some of that uncertainty by spending significant time on method (about one-fifth of the book) as well as providing a number of applications (roughly four-fifths). This structure, and the presentation of the book in general, clearly demonstrates that method drives the selection of research questions. This may seem unremarkable, but it is perhaps the book's most important insight to be gained by newcomers to empirical legal studies. Popular conceptions of artificial intelligence and machine learning paint a picture that machines can do it all. This is not true. While AlphaZero may be able to defeat any human opponent, the algorithm is helpless if the rules of chess change frequently enough.¹¹ Many legal domains are characterized by rapid change, but algorithms require sufficiently stable environments for making accurate predictions.¹² Besides, law itself is situated within social and political life, and even general artificial intelligence may prove useless, and certainly undesirable, for political competitors. A general and strong form of AI, for instance, may be able to instruct lawmakers where to place a bridge so as to cause minimal traffic, pollution, building costs, and accidents, but it is easy to imagine that

Danger Ahead: Risk Assessment and the Future of Bail Reform, 93 WASH. L. REV. 1725, 1729 (2018).

9. See Livermore & Rockmore, pp. xiv–xv; Copus, Hiibert & Laqueur, p. 22.

10. For example, the Stanford Legal Tech List, hosted by CodeX at tech.law.stanford.edu, lists 1,744 (as of March 15, 2021) companies that are “changing the way legal is done.” About Page, CODEX TECHINDEX, <https://techindex.law.stanford.edu/about> [<https://perma.cc/2CWH-F28R>].

11. AlphaZero is an AI program developed to play chess and other games. David Silver et al., *A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go Through Self-Play*, 362 SCIENCE 1140–44 (2018). A major assumption underlying AI systems, however, is that the context in which the system is operating remains relatively unchanging. See LESLIE VALIANT, *PROBABLY APPROXIMATELY CORRECT* 61–62 (2013).

12. In addition, they require “big data.” See IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, *DEEP LEARNING* 20 (2016), which notes that “[a]s of 2016, a rough rule of thumb is that a supervised deep learning algorithm will generally achieve acceptable performance with around 5,000 labeled examples *per category* and will match or exceed human performance when trained with a dataset containing at least 10 million labeled examples” (emphasis added). It is plausible that many questions in law can only draw on small amounts of data and cannot be answered with current tools.

politicians will desire to build the bridge elsewhere to provide jobs and construction contracts to their favored constituents.

For the most part, *Law as Data* provides a narrow but important focus on questions related to the feasibility of machine learning in law by analyzing the distinction between prediction and causality. Three chapters, placed early within the first section of the book, provide a substantial discussion, and the reader might benefit from approaching them together as a whole. Algorithms predict; they are not equipped with the tools of causal reasoning.¹³ Lawmakers may wonder, for example, if the introduction of universal basic income will *cause* a 10 percent reduction in crime, but predictive algorithms cannot answer that question. Policy decisions, of course, can be made on the basis of noncausal forecasts, but one must be comfortable implementing new rules without tested reasons—other than “it seems to work often”—and trust that this prediction is sufficiently robust and stable over time.

Prediction examines the past in order to make sense of the future. The accuracy of a predictive inference is therefore beholden to its understanding of the past. Causal inference shares this approach. It relies on observations made earlier in time in order to deduce a relationship that will come about tomorrow. Of course causal inference can be used to identify relationships that obtain in the past and present, but lawmakers are generally interested in the future. For instance, an empirical legal scholar engaged in causal inference may produce a study that demonstrates a relationship between incarceration in maximum-security prison and reduced inmate misconduct.¹⁴ The study may conclude that placing an inmate in a maximum-security prison reduces the likelihood of misconduct by 10 percent. The causal claim asserts that this relationship holds in the past, present, or future; but a policymaker evaluating plans for prison expansion is obviously interested in whether the security level of a prison will impact inmate misconduct tomorrow. Lawmaking is a forward-looking endeavor.

13. On the other hand, researchers are actively working on supplying machine-learning algorithms with the tools of causal inference. See generally Judea Pearl, *Theoretical Impediments to Machine Learning with Seven Sparks from the Causal Revolution*, ARXIV (Jan. 15, 2018), <https://arxiv.org/pdf/1801.04016.pdf> [<https://perma.cc/5ZZ5-TUUZ>]. The task is challenging. Simply adding a new variable can generate exponential growth in the number of candidate hypotheses of a causal relationship. See DOMINGOS, *supra* note 1, at 73. In principle, more data could reduce the number of hypotheses, but adding more data typically introduces more variables, which in turn leads to new, additional hypotheses of the causal relationship. Thus, it is common for the number of hypotheses to exceed the number of observations. See VALIANT, *supra* note 11, at 74. For further discussion and an example in bankruptcy law, see Frank Fagan & Saul Levmore, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, 93 S. CAL. L. REV. 1, 23 (2019).

14. See, e.g., Richard A. Berk & Jan de Leeuw, *An Evaluation of California's Inmate Classification System Using a Generalized Regression Discontinuity Design*, 94 J. AM. STAT. ASS'N 1045, 1052 (1999). For further discussion of the Berk and de Leeuw study, as well as a helpful discussion of regression discontinuity methods for law, see Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. & SOC. SCI. 17 (2011).

For this reason, both predictive and causal inferences of legal outcomes are susceptible to changing circumstances and the changing behavior of parties.¹⁵ Inmates may learn of a new deadly virus and begin to practice social distancing on their own, which may, in turn, reduce levels of misconduct independent of a prison's security level. A causal study carried out in 2018, even if perfectly designed to mimic an ideal random experiment, would generate incorrect inferences in 2020. Changing circumstances generate errors no matter how perfectly a study simulates law's random application. Indeed, even if equal protection and other constitutional rules were ignored and a law were applied randomly,¹⁶ any conclusion regarding that law's causal effect is accurate only to the extent that the future resembles the past.

Because so much of empirical legal scholarship attempts to make causal inferences,¹⁷ there is a real need to define what, exactly, predictive empirical legal scholarship should (and can) be doing. In a lucid passage written by Marion Dumas and Jens Frankenreiter, this need is clearly stated:

[O]ne important challenge for researchers seeking to exploit the full potential of machine learning techniques is to identify questions that can be answered meaningfully by means of prediction and classification. This constitutes a rather new epistemological approach for social scientists, and research agendas based on predictive inference are just starting to emerge. (Dumas & Frankenreiter, pp. 63–64; citation omitted)

15. Causal studies often side-step the second challenge of “changing behavior of parties” by assuming a fixed distribution of party characteristics. For instance, when evaluating whether an additional hour of study for an exam enhances performance, various attributes of the members of a student body, such as stamina, can be considered fixed or “exogenous” to the treatment of an additional hour of study. Of course, an additional hour of study might increase stamina. There are generally two groups of techniques used to sidestep this problem. One group adjusts for differences in party characteristics through *observation*. Popular techniques include regression analysis, propensity scoring, and matching. See GUIDO W. IMBENS & DONALD B. RUBIN, *CAUSAL INFERENCE FOR STATISTICS, SOCIAL, AND BIOMEDICAL SCIENCES: AN INTRODUCTION* (2015). Another adjusts for differences through *modeling*, particularly “structural modeling,” which basically identifies the characteristics (or “preferences”) of people, as well as a model of how those characteristics come about. *Id.*; Susan Athey, *Machine Learning and Causal Inference for Policy Evaluation*, 21 KDD '15, at 5, 5 (2015), <https://doi.org/10.1145/2783258.2785466> [<https://perma.cc/P3QD-GSHA>]. Model parameters are then estimated from data. Both categories of techniques for identifying party characteristics, whether through observation or modeling, are susceptible to changing circumstances that remain uncaptured. Advanced techniques that leverage machine learning for identifying characteristics remain susceptible to changing and uncaptured circumstances as well.

16. For a discussion, see Michael Abramowicz, Ian Ayres & Yair Listokin, *Randomizing Law*, 159 U. PA. L. REV. 929, 967–74 (2011).

17. Empirical legal studies exhibit a wide range of approaches. Daniel E. Ho & Larry Kramer, *Introduction: The Empirical Revolution in Law*, 65 STAN. L. REV. 1195, 1197 n.3 (2013). However, studies that “prioritize research design (often approximating a natural experiment) [can] credibly estimate the causal effect of some intervention of interest.” *Id.* at 1200. The “credibility revolution” in social sciences has led to an increase in causal inference studies in law.

So much more of this statement is to be made for law. What can the data available to lawmakers accomplish? One way of approaching this question is to ask in which situations predictive inference is superior to its causal counterpart. As should be clear, the accuracy of both inference types benefits from environmental stability. Causal inference works well in medicine, for instance, because biochemical reactions are analyzed within a relatively closed environment (i.e., a human body), and differences across humans can be accounted for by means of randomization.¹⁸ Across time, fundamental changes to the typical human body tend to occur gradually, and at a pace slow enough for making valid causal inferences in, say, 2020 or 2025 on the basis of tests performed throughout a prior decade. In contrast, causal inference tends to work poorly when engaged in tasks such as inferring future stock prices because financial events must be analyzed within relatively open environments subject to change. Markets adapt quickly, and better predictions are quickly arbitrated away. Moreover, changes to underlying market structure in which financial events occur are common but relatively unexpected, as when central banks make dramatic and systemic interventions.

What does this mean for law? In situations where the legal environment is comprehensively understood and stability is likely, well-designed causal studies of policy will tend to generate fewer errors.¹⁹ Where comprehension is low and the policy domain undertheorized, or where the stability of the domain is less likely, causal inference should be trusted less. Further, legal domains that reflexively adapt to intervention (like markets) will likely be better suited to prediction in most instances. In these settings predictive inference offers at least two advantages: (1) its conclusions are less ambitious and can therefore be trusted more relative to an inference of the same relationship with causal methods; and (2) prediction involves far-reaching collection of seemingly unrelated variables that can soak up poorly understood aspects of the policy environment. Generous inclusion of variables can lead to robust inferences that can later be refined with theory and further testing. One of the achievements of *Law as Data* is that it provides a guide for exploring and developing predictive inference research agendas in law.

B. *Basic NLP Research for Law*

Most of the remaining fourteen chapters supply the reader with applications. Within this grouping, one can loosely track the familiar distinction found in the hard sciences between early-stage basic research and later-stage applications for industrial use and consumption. “Basic” research chapters include:

18. This includes differences that are unobserved. Ho & Rubin, *supra* note 14, at 22 (“Randomization over a large number of units ensures that treatment and control units are comparable in all respects other than the treatment.”).

19. These studies can identify party characteristics with a number of methods, including observational study, structural modeling, and machine learning. See *supra* note 15 and accompanying text; see also Athey, *supra* note 15, at 5–6.

- a textual examination of law that maps its broad domains, features, and contours (Livermore & Rockmore, Chapter One);
- analysis of the Supreme Court’s writing style to uncover broad trends in substance and intelligibility (Carlson, Rockmore, Riddell, Ashley & Livermore, Chapter Five);
- analysis of the same for the European Court of Justice (Frankenreiter, Chapter Seven);
- a demonstration of how the text of judicial opinions can be represented by machines to enable downstream prediction tasks (Ash & Chen, Chapter Eleven);
- a demonstration of how measurement of cross-references within civil codes can be used to support qualitative assertions about differences among legal systems (Badawi & Dari-Mattiacci, Chapter Twelve);
- detection of conservative versus liberal phrases in environmental case law and classification of judicial ideology (Dumas, Chapter Fourteen); and
- an analysis of semantic intelligibility of state and federal trial court opinions (Feldman, Chapter Fifteen).

While some of these chapters include potential administrative applications, their use for lawyers and judges is less apparent and immediate. As a result, this grouping of chapters makes a strong case that natural language processing (NLP) techniques clearly work well for broad research questions that examine the workings of law from afar. Description and investigation of the general contours of law and legal systems, the writing style of judges in terms of substance and intelligibility, and the detection of partisanship in judicial opinions are, all, clearly of scientific value and merit sustained research. Nonetheless, there remains the important question of “What next?”

It is true that several of the chapters listed above present obvious normative questions. Should the use of cross-references within codes and other laws be restricted or expanded, perhaps with legislation?²⁰ What, if anything, should be done about judicial partisanship?²¹ Should judges and their clerks adopt simplified writing styles?²² But it is hard to imagine concrete applications, such as Congress passing a law mandating threshold Flesch-Kincaid scores²³ for judicial opinions, though at least one state requires a minimum score for insurance policies.²⁴

20. See Badawi & Dari-Mattiacci, chapter 12.

21. See Dumas, chapter 14.

22. See Feldman, chapter 15.

23. Flesch-Kincaid scores measure readability based upon word, sentence, and syllable counts. Feldman, pp. 415–16.

24. See FLA. STAT. § 627.4145(1)(a) (2019), which notes that certain insurance policies must “achieve[] a minimum score of 45 on the Flesch reading ease test as computed in subsec-

C. *Applied NLP Research for Law*

The remaining chapters contained within *Law as Data* tilt more toward practical and immediate applications for law practice. “Applied” research chapters include:

- prediction of the likelihood that state legislation will reach the floor on the basis of characteristics of legislatures, legislators, and importantly for drafters, the legislation text (Eidelman, Kornilova & Argyle, Chapter Six);
- an examination of whether particular phrases uttered by inmates during parole hearings, nonverbal inmate characteristics, and decisionmaker biases increase the likelihood of parole (Laqueur & Venancio, Chapter Eight);
- the development of a natural language processing approach to analyze and summarize mass commenting on administrative agency rules (Eidelman, Grom & Livermore, Chapter Nine);
- an analysis of the case files for all employment law cases that were closed within a U.S. district court over a seven-year period to determine whether various features of a lawsuit can predict its case-ending event such as settlement, dismissal, and summary judgment; and a description of (1) the frequency and distribution of case-ending events, (2) the legal doctrines used in summary judgment, and (3) the strategic patterns used by defense lawyers and judges (Alexander, al Jadda, Feizollabi & Tucker, Chapter Ten);
- an examination of whether an attorney’s vocal style of uttering “Mr. Chief Justice” and “May it please the Court” can predict Supreme Court outcomes (Chen, Halberstam, Kumar & Yu, Chapter Thirteen);
- a discussion of how machine learning can be used to detect extra-legal bias in judicial decisionmaking (Chen, Chapter Sixteen); and
- early development of a machine-learning task in which an algorithm predicts which case citations will be used by judges based solely on the words contained within their opinions (Livermore & Rockmore, Chapter Seventeen).

These chapters demonstrate that natural language processing methods have potential for changing legal practice.

One may wonder, however, about how extended knowledge of language could introduce new opportunities to “game the system” that lead to social loss. Will, for example, knowledge and use of the “magic” words that should be uttered by inmates during parole hearings degrade the quality of parole decisions? Should lawyers fret about voice intonation when they approach

tion (5) or an equivalent score on any other test comparable in result and approved by the office.”

the bench? Perhaps these questions simply reflect the general problem of separation between words and meaning as patterns of speech become habituated.²⁵ Surely, judges and other decisionmakers will search for new guides as signal quality degrades. In any case, it is apparent that sustained development of these types of studies will raise important normative questions regarding the mechanization of law.

Related is Daniel Chen's chapter on the detection of extralegal bias in judicial decisions.²⁶ Chen cites research that shows that decisionmakers tend to reduce bias when alerted.²⁷ Machines, equipped with a capacity to observe many variables at once, can detect and uncover various biases unbeknownst to the judge who perpetuates them.²⁸ Once notified, that judge can engage in introspection and self-correction.²⁹ Like the chapters on inmate utterances and lawyer voice patterns, Chen's chapter raises the broader question of how lawyers and judges should react to the hidden connections uncovered by machines. Gaming the system often carries a negative connotation, but machines can teach us something socially worthwhile, too, especially when they provoke self-reflection.

The remaining chapters contained within this grouping are of obvious use to legal practitioners as well. The chapter written by Charlotte Alexander and others develops a tool that predicts case outcomes on the basis of filings.³⁰ While it achieves moderate accuracy on the basis of filings from the early stages of litigation (Alexander, al Jadda, Feizollabi & Tucker, p. 307), further work in this area is clearly of social value inasmuch as predicting case outcomes can lower the costs of litigating and judging. When parties can accurately estimate case outcomes, judicial caseloads tend toward plaintiff victories 50 percent of the time.³¹ This means that parties will tend to only file close cases and will either settle or choose not to litigate unlikely victories. Setting aside the noneconomic compulsions to sue, further development of

25. Contemporary formulations of this idea can be found in FERDINAND DE SAUSSURE, *COURSE IN GENERAL LINGUISTICS* (Perry Meisel & Haun Saussy eds., Wade Baskin trans., Columbia Univ. Press 2011) (1959). Saussure emphasized that the meaning of language is socially constructed, which implies that the meaning of signifiers (words or voice intonation) is dependent upon human relationships. *Id.* at 15. Put differently, magic words may lose their magic once everyone is using them.

26. See Chen, chapter 16.

27. He gives the example of NBA referees, who, once alerted of racial bias, proceed to reduce it. Chen, p. 440 (citing Devin G. Pope, Joseph Price & Justin Wolfers, *Awareness Reduces Racial Bias* (Nat'l Bureau of Econ. Rsch., Working Paper No. 19765, 2013), https://www.nber.org/system/files/working_papers/w19765/w19765.pdf [<https://perma.cc/39N2-4FFC>]).

28. *E.g.*, Chen, pp. 437–38.

29. *Cf.* MAHZARIN R. BANAJI & ANTHONY G. GREENWALD, *BLINDSPOT: HIDDEN BIASES OF GOOD PEOPLE* 1–5 (2013) (describing optical illusions, produced by automatic processes in the mind, that easily can be reversed as soon as the illusion is brought to a viewer's attention).

30. See Alexander, al Jadda, Feizollabi & Tucker, chapter 10.

31. Yoon-Ho Alex Lee & Daniel Klerman, *The Priest-Klein Hypotheses: Proofs and Generality*, 48 *INT'L REV. L. & ECON.* 59, 60 (2016).

predictive litigation tools promises to reduce aggregate litigation costs and unburden dockets. But note that predictive accuracy can become muddled over time, much the same way that the connection between magic words and their signal might become severed if overused. As litigants rely more on predictive tools to make litigation choices, fewer cases will be decided by judges and precedent stocks will be reduced.³² In addition, case features will likely diverge rather than converge over time as environments change. All of this points to cyclical patterns of certainty then uncertainty, which means that predictive accuracy will sharpen, then dull, and then sharpen again as precedent will constantly need to play catch up with environmental change.³³ This implies that there exists a socially optimal level of effort that should be directed toward building legal prediction systems.

This logic applies broadly. Michael Livermore and Daniel Rockmore's interesting task of predicting case citations on the basis of the words contained within a judicial opinion can be made easier by delimiting a time frame such as "any opinion written between 2020 and 2025."³⁴ Other approaches correlated with time might be fruitful, such as, "any opinion written before the introduction of the self-driving car," or "all asylum decisions prior to the XYZ civil war," even if these approaches sacrifice the generality inherent in Livermore and Rockmore's task. An artificial intelligence that can handle completely new environments cannot rely on the tools of predictive inference because it has no data from which to build conclusions. To accurately predict outcomes, the AI would need to be equipped with the tools of causal inference, which are currently unavailable to machines as Chapters Two, Three, and Four make clear. Even then, its domain must be adequately understood and remain sufficiently stable as emphasized above.³⁵

II. DOCTRINAL NLP STUDIES WITHOUT PREDICTION

Law as Data is interdisciplinary. The authors of its chapters include lawyers, computer scientists, political scientists, policy studies scholars, and economists. They are brought together by a shared interest in applying methods for evaluating textual data to answer questions about law and legal institutions broadly considered. Because so much of legal practice is expressed textually—and takes the form of written pleadings, opinions, statutes, and regulations (let alone the written workings of administrative processes and legislative routines)—it would seem that natural language pro-

32. This is true given a uniform or normal distribution of likelihoods of plaintiff victory.

33. See William M. Landes & Richard A. Posner, *Legal Precedent: A Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249, 250–51 (1976) (noting that the value of precedent depreciates due to obsolescence). Accident law developed for collisions of automobiles, for instance, will likely provide guidance, but not hard rules, for collision of self-driving automobiles.

34. See *supra* notes 15–16 and accompanying text (noting how predictive accuracy benefits from environmental stability).

35. See *supra* notes 15–16 and accompanying text.

cessing has a bright future in law. In addition, legal data is “high-dimensional”; that is, a single unit of study (an observation) consists of many variables. For instance, a judicial opinion may consist of 4,000 words, and each word might be counted as a variable. Natural language processing tools are crafted specifically for analyzing numerous observations that contain thousands of variables.

One area full of potential is the conversion of doctrinal standards into rules.³⁶ For example, in their study of corporate veil piercing, Jonathan Macey and Joshua Mitts analyze the full text of 9,380 decisions and provide evidence that courts rarely use undercapitalization as a veil-piercing rationale.³⁷ If this research were internalized by courts, then undercapitalization would be one less rationale to rule on, or at least evaluate, when determining whether to veil pierce. In addition to providing evidence that undercapitalization is rarely dispositive, Macey and Mitts use NLP to develop a coherent and streamlined taxonomy for what is widely considered a messy and sprawling doctrine.³⁸ I take the same approach in a study of 2,100 successor-liability decisions.³⁹ The NLP analysis uncovers evidence that courts incoherently apply de facto merger doctrine and offers an updated taxonomy.⁴⁰

Of course, natural language processing can reveal relative imprecision and a lack of rules. For instance, in a study of approximately 20,000 cases related to contractual good faith, the texts uncovered consistent application of a standard.⁴¹ While additional fine-grained analysis of the opinion texts might reveal highly contextual rule application,⁴² revelation is possible only to the extent that a legal domain remains sufficiently stable over time.⁴³

36. On the conversion of standards to rules generally, beyond doctrine, see Fagan & Levmore, *supra* note 13, at 1.

37. Macey & Mitts, *supra* note 7, at 99.

38. They note that courts pierce for three reasons: “(1) achieving the goals of a particular regulatory or statutory scheme; (2) avoiding fraud or misrepresentation by shareholders trying to obtain credit; and (3) promoting the bankruptcy value of eliminating favoritism among claimants to the cash flows of a firm.” *Id.* at 113.

39. Frank Fagan, *From Policy Confusion to Doctrinal Clarity: Successor Liability from the Perspective of Big Data*, 9 VA. L. & BUS. REV. 391 (2015).

40. In particular, the article notes that courts incoherently use de facto merger to spread risk in jurisdictions where successor liability is supposed to be used for suppressing bad behavior and evasion of predecessor liabilities. *Id.* at 394–95. NLP is used to develop a new taxonomy of successor liability that eliminates the confusion. *Id.* at 394.

41. Frank Fagan, *Waiving Good Faith: A Natural Language Processing Approach*, 16 N.Y.U. J.L. & BUS. 633 (2020). The majority of the opinions demonstrated that judges applied “community standards of decency, fairness, and reasonableness” between contractual partners when assessing good faith violations. *Id.*

42. *Cf.* Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401, 1401 (2017) (describing the development of microdirectives on the basis of contingent and specific factual contexts).

43. See Fagan & Levmore, *supra* note 13, at 16, 19 (noting that sufficient stability of a legal domain is a necessary condition for good machine-generated rules).

There is, of course, a possibility that extratextual analysis could reveal hidden rules inasmuch as judges simply pay lip service when uttering standards, but in either scenario, NLP can help. It will either distill doctrine to closed-ended rules, or provide further rationale for maintaining open-ended standards.⁴⁴

These examples demonstrate that NLP on its own—without relying on machine-learning algorithms that make predictive (or even causal) inferences—can be of considerable use to lawyers and judges. Older descriptive studies of legal doctrine that populate law reviews and treatises rely on hand-coded cases and small datasets.⁴⁵ They suffer from selection bias. Old-fashioned positive studies of black-letter law carried out with sophisticated NLP tools can clarify standards and help make law more precise.⁴⁶ The social benefits of this work include reduced litigation costs and fewer judicial errors.⁴⁷ *Law as Data* describes this work as a quantitative, internal study of law (Livermore & Rockmore, p. 10). It is certainly descriptive work of great potential value.

CONCLUSION

Law as Data provides a comprehensive overview of what natural language processing can bring to law. Through its wide-ranging chapters, it demonstrates that predictive inferences drawn from large stocks of data can transform legal practice in areas as diverse as word choice and intonation, selection of litigation strategy, legislative drafting, and judging. Several chapters, when read together, supply the reader with a good understanding of the technical differences between predictive and causal inference and what those differences mean for policymaking. Specialists will undoubtedly consult individual chapters, but *Law as Data* can be profitably read as a whole. What emerge from its entirety are the potential outcomes of a slow, but persistent, evolution of legal research and practice. To lawyers and judges, the possibilities will be both novel and familiar.

44. For more on how NLP can streamline doctrine, see Frank Fagan, *Big Data Legal Scholarship: Toward a Research Program and Practitioner's Guide*, 20 VA. J.L. & TECH. 1, 25–29 (2016).

45. See, e.g., Robert S. Summers, “Good Faith” in *General Contract Law and the Sales Provisions of the Uniform Commercial Code*, 54 VA. L. REV. 195, 203 (1968). Summers’s seminal study continues to provide the basis for the Restatement approach to contractual good faith despite relying on examples of bad-faith conduct derived from fewer than 100 cases. RESTATEMENT (SECOND) OF CONTRACTS § 205 reporter’s note (AM. L. INST. 1981) (citing Summers’s survey alongside two other authorities as the basis for a new section on good faith).

46. This is more likely when legal domains are sufficiently settled and unchanging. In sufficiently unstable environments, law will do better with standards. See Fagan & Levmore, *supra* note 13, at 33.

47. Of course other important applications, in addition to streamlining doctrine, are being developed. For instance, researchers are actively working on the development of legal-reasoning tools on the basis of predictive inference. If successful, these tools will generate legal arguments—both for and against particular outcomes—that can then be evaluated by legal practitioners. See KEVIN D. ASHLEY, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS: NEW TOOLS FOR LAW PRACTICE IN THE DIGITAL AGE 3–4 (2017).