PREDICTIVE CONTRACTING

Spencer Williams*

This Article examines how contract drafters can use data on contract outcomes to inform contract design. Building on recent developments in contract data collection and analysis, the Article proposes "predictive contracting," a new method of contracting in which contract drafters can design contracts using a technology system that helps predict the connections between contract terms and outcomes. Predictive contracting will be powered by machine learning and draw on contract data obtained from integrated contract management systems, natural language processing, and computable contracts. The Article makes both theoretical and practical contributions to the contracts literature. On a theoretical level, predictive contracting can lead to greater customization, increased innovation, more complete contract design, more effective balancing of front-end and back-end costs, better risk assessment and allocation, and more accurate term pricing for negotiation. On a practical level, predictive contracting has the potential to significantly alter the role of transactional lawyers by providing them with access to previously unavailable information on the statistical connections between contract terms and outcomes. In addition to these theoretical and practical contributions, the Article also anticipates and addresses limitations and risks of predictive contracting, including technical constraints, concerns regarding data privacy and confidentiality, the regulation of the unauthorized practice of law and the potential for exacerbating information inequality.

I.   INTRODUCTION

This Article examines how contract drafters can use data on contract outcomes[1] to inform contract design.[2] Building on recent developments in contract data collection and analysis, the Article proposes "predictive contracting," a new method of contracting in which contract drafters can design contracts using a technology system that helps predict the connections between contract terms and outcomes.

On July 25, 2018, the major technology company Qualcomm announced that it was walking away from its $44 billion acquisition of NXP Semiconductors.[3] Qualcomm and NXP had been working on closing the acquisition for almost two years.[4] Qualcomm finally decided to terminate the deal after failing to receive regulatory approval from China.[5] Yet despite Qualcomm's best efforts to close the deal,[6] it did not

---

[1] This Article uses the term "contract outcomes" to refer to a broad set of outcomes that can be used to assess a contract's performance. This includes outcomes such as whether the contract resulted in litigation, the quality and timing of counterparty performance, how much the contract cost to draft and administer, etc. For a discussion of contract outcomes, see *infra* Section III.B.2.

[2] This Article uses the term "contract design" to collectively refer to the set of terms that make up a contract. *See* Albert Choi & George Triantis, *The Effect of Bargaining Power on Contract Design*, 98 VA. L. REV. 1665, 1665–69 (2012); Spencer Williams, *Venture Capital Contract Design: An Empirical Analysis of the Connection Between Bargaining Power and Venture Financing Contract Terms*, 23 FORDHAM J. CORP. & FIN. L. 105, 106–07 (2017).

[3] *See* Don Clark, *Qualcomm Scraps $44 Billion NXP Deal After China Inaction*, N.Y. TIMES (July 25, 2018), https://www.nytimes.com/2018/07/25/technology/qualcomm-nxp-china-deadline.html [https://perma.cc/8D52-7JLT].

[4] *See* Chad Bray & Quentin Hardy, *Qualcomm to Acquire NXP Semiconductors for $38.5 Billion*, N.Y. TIMES (Oct. 27, 2016), https://www.nytimes.com/2016/10/28/business/dealbook/qualcomm-acquire-nxp-semiconductors.html [https://perma.cc/7RDX-T85D].

[5] *See* Clark, *supra* note 3.

[6] The CEOs of Qualcomm and NXP even exchanged text messages thanking each other for their work on the deal despite the end result. *See* Stu Woo, *"I'm Sorry": Qualcomm and NXP Chiefs Lament Failed Deal Via Texts*, WALL ST. J. (July 26, 2018), https://www.wsj.com/articles/qualcomm-

get to walk away for free.[7] Instead, it was forced to pay NXP a termination fee of $2 billion under a breakup provision in the acquisition agreement.[8]

As the breakup provision in the Qualcomm-NXP example highlights, contract terms can have significant effects on outcomes for the parties involved. This is the case in business-to-business contracts and business-to-consumer contracts.[9] For example, experimental evidence suggests that transfer provisions in mortgage contracts can increase the likelihood of homeowners engaging in strategic default[10] and that parties are more likely to exploit efficient-breach opportunities if the contract contains a liquidated damages clause.[11] In addition, empirical results show that anti-dilution mechanisms in venture capital contracts can alter the division of control between entrepreneurs and investors[12] and that earnout clauses in complex acquisition agreements can have

---

deal-collapse-forces-nxp-to-forge-a-new-path-1532600411 (on file with the *Columbia Business Law Review*).

[7] *See* Clark, *supra* note 3.

[8] *Id.*

[9] While this Article focuses primarily on business-to-business contracts, it also highlights some important issues unique to business-to-consumer contracts. For a categorization of contracts along these lines, see Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 YALE L.J. 541, 544 (2003) (breaking contracts down into four categories based on whether the parties are businesses and/or individuals).

[10] *See* Tess Wilkinson-Ryan, *Breaching the Mortgage Contract: The Behavioral Economics of Strategic Default*, 64 VAND. L. REV. 1547, 1573–74 (2011).

[11] *See* Tess Wilkinson-Ryan, *Do Liquidated Damages Encourage Breach? A Psychological Experiment*, 108 MICH. L. REV. 633, 633–38 (2010).

[12] Robert E. Hall & Susan E. Woodward, *The Incentives to Start New Companies: Evidence from Venture Capital* 11 (Nat'l Bureau of Econ. Research, Working Paper No. 13056, 2007) (finding that "[d]own-round anti-dilution provisions shift venture ownership upward and non-venture (entrepreneurs, angels, and employees) downward by an average of 4.8 percentage points"). For a discussion of anti-dilution provisions, see Michael A. Woronoff & Jonathan A. Rosen, *Understanding Anti-Dilution Provisions in Convertible Securities*, 74 FORDHAM L. REV. 129, 140–55 (2005).

large effects on acquirer returns.[13] Despite the importance of contract design to outcomes, little attention has been given to how contract drafters use historical outcomes to inform subsequent contract drafting. This is in sharp contrast to the focus on outcomes in other fields, such as medicine, engineering, philanthropy, and education.[14]

The contracts literature has long been divided over how contract terms evolve over time. Efficient contracting theory takes the view that contract terms evolve via a market-based natural selection process.[15] According to this view, the goal of contracting parties is to maximize the joint value created by

---

[13] *See* Leonidas Barbopoulos & Sudi Sudarsanam, *Determinants of Earnout as Acquisition Payment Currency and Bidder's Value Gains*, 36 J. BANKING & FIN. 678, 678 (2012); Reena Kohli & Bikram Jit Singh Mann, *Analyzing the Likelihood and Impact of Earnout Offers on Acquiring Company Wealth Gains in India*, 16 EMERGING MKTS. REV. 203, 203 (2013).

[14] *See generally* PAUL BREST & HAL HARVEY, MONEY WELL SPENT: A STRATEGIC PLAN FOR SMART PHILANTHROPY 135–66 (2008) (highlighting the importance of measuring the return on philanthropic investments and using this information when making subsequent investment decisions); SIGURD SKOGESTAD & IAN POSTLETHWAITE, MULTIVARIABLE FEEDBACK CONTROL: ANALYSIS AND DESIGN (2d ed. 2001) (discussing the use of feedback mechanisms in systems engineering); Raj Chetty, John N. Friedman & Jonah E. Rockoff, *Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood*, 104 AM. ECON. REV. 2633 (2014) (demonstrating that students who are taught by high value-added teachers have better outcomes in adulthood including higher rates of college attendance and higher salaries); Scott L. Pomeroy et al., *Prediction of Central Nervous System Embryonal Tumour Outcome Based on Gene Expression*, 415 NATURE 436 (2002) (showing that the clinical outcomes of children with embryonal tumors of the central nervous system known as medulloblastomas are highly predictable based on gene expression profiles).

[15] *See* Michelle E. Boardman, *Contra Proferentem: The Allure of Ambiguous Boilerplate*, 104 MICH. L. REV. 1105, 1116 (2006); Charles J. Goetz & Robert E. Scott, *The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms*, 73 CALIF. L. REV. 261, 278 (1985) (describing a "quasi-Darwinian evolutionary process" for contract terms); Marcel Kahan & Michael Klausner, *Standardization and Innovation in Corporate Contracting*, 83 VA. L. REV. 713, 760–61 (1997); Michael Klausner, *Corporations, Corporate Law, and Networks of Contracts*, 81 VA. L. REV. 757, 767, 787 (1995).

the contract.[16] Over time, sub-optimal terms that do not maximize joint value are weeded out by value-maximizing contract drafters. This process leads to an optimal steady-state contract design in which the terms of the contract efficiently maximize the joint value for the parties.[17] Yet there is substantial scholarship that raises theoretical and empirical challenges to this natural selection view, citing examples of steady-state contracts that contain sub-optimal terms and/or are inefficiently incomplete.[18]

[16] *See* Schwartz & Scott, *supra* note 9, at 544, 552 (arguing that contracting parties aim to maximize joint value and that contract law should facilitate this value-maximization); Robert E. Scott & George G. Triantis, *Incomplete Contracts and the Theory of Contract Design*, 56 CASE W. RES. L. REV. 187, 188 (2005).

[17] *See* FRANK H. EASTERBROOK & DANIEL R. FISCHEL, THE ECONOMIC STRUCTURE OF CORPORATE LAW 1–39 (1991); Marcel Kahan & Michael Klausner, *Path Dependence in Corporate Contracting: Increasing Returns, Herd Behavior and Cognitive Biases*, 74 WASH. U. L.Q. 347 (1996). This view of contract design efficiency is a contractual application of the classic Coase Theorem that argues that with perfect information and no transaction costs, resources will be allocated efficiently regardless of initial allocation. *See* R.H. Coase*, The Problem of Social Cost*, 3 J.L. & ECON. 1, 15–16 (1960). The term "steady state" refers to the state of a system that has reached equilibrium following an external shock. For a discussion of how contract terms respond to external shocks, see Stephen J. Choi & G. Mitu Gulati, *Innovation in Boilerplate Contracts: An Empirical Examination of Sovereign Bonds*, 53 EMORY L.J. 929, 933–36 (2004) (discussing how interpretive shocks can lead to changes in steady state contract terms and providing empirical evidence from sovereign bonds); Stephen J. Choi, Mitu Gulati & Eric A. Posner, *The Dynamics of Contract Evolution*, 88 N.Y.U. L. REV. 1, 7–10, 27, 35–36 (2013) (proposing and testing a three-stage model of contract evolution that includes pre-shock standardization, post-shock innovation, and post-shock standardization).

[18] Frequently cited challenges to the efficient contracting theory include network and learning externalities, agency costs, cognitive biases and bounded rationality. *See* Luca Anderlini & Leonardo Felli, *Bounded Rationality and Incomplete Contracts*, 58 RES. IN ECON. 3, 5 (2004) (describing how bounded rationality forces parties to write incomplete contracts); Claire A. Hill, *Why Contracts Are Written in "Legalese"*, 77 CHI.-KENT L. REV. 59, 60 (2001) (noting that iterations of contract forms do not always improve forms and sometimes make the form worse); Kahan & Klausner, *supra* note 17, at 350–64; Avery W. Katz, *Contractual Incompleteness: A Transactional Perspective*, 56 CASE W. RES. L. REV. 169,

As the above disagreement in the contracts literature highlights, contract drafters are impeded in their ability to iterate on contract design based on outcomes. This is partly due to two technical barriers long faced by contract drafters. First, contract drafters have traditionally had limited data on contract terms and outcomes.[19] Most companies take an ad

172–73 (2005); Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203, 1206 (2003) (arguing that due to bounded rationality, term-takers in the context of contracts of adhesion only consider a limited set of contract terms and therefore term-givers have an incentive to choose inefficient, allocatively favorable forms of the terms that are not considered); Russell Korobkin, *Inertia and Preference in Contract Negotiation: The Psychological Power of Default Rules and Form Terms*, 51 VAND. L. REV. 1583, 1586–87 (1998) (discussing an inertia theory of contract negotiation in which parties prefer previously used terms because of status quo and endowment bias); Russell Korobkin, *The Status Quo Bias and Contract Default Rules*, 83 CORNELL L. REV. 608 (1998); Barak Richman, *Contracts Meet Henry Ford*, 40 HOFSTRA L. REV. 77, 78 (2014); Oliver E. Williamson, *The Economics of Organization: The Transaction Cost Approach*, 87 AM. J. SOC. 548, 553–54 (1981); Kenneth A. Adams, *Copyright and the Contract Drafter*, N.Y. L.J., Aug. 23, 2006, at 1–5 (discussing the difficulty of copyrighting novel contract terms). For an excellent example of empirical evidence that runs counter to the natural selection view of contract evolution, see MITU GULATI & ROBERT SCOTT, THE 3 1/2 MINUTE TRANSACTION 2–3 (2012) (describing the continued widespread use of a "pari passu" clause in cross-border sovereign bond contracts following an adverse judicial ruling that upset the standing interpretation of the clause). As Gulati and Scott note, the pari passu clause continued to be used in ninety percent of sovereign bond contracts despite the adverse judicial interpretation. *Id.* In fact, use of the clause increased even as understanding of its meaning decreased. *Id.* at 141.

[19] *See* ANUJ SAXENA, ENTERPRISE CONTRACT MANAGEMENT: A PRACTICAL GUIDE TO SUCCESSFULLY IMPLEMENTING AN ECM SOLUTION 11–12, 16–17 (2008) (noting that most organizations manage contracts in an ad hoc manner, which results in numerous problems including fragmented contract data, poor visibility into contracts, ineffective contract monitoring, and inadequate analysis of contract performance); GULATI & SCOTT, *supra* note 18, at 4, 150 (identifying that the traditional structural division between litigation and transactional law practice prevents transactional lawyers from systematically modifying contract drafting based on litigation outcomes); Hill, *supra* note 18, at 75–76 (describing the tension between a contract as a document that meets the needs of the parties and a contract as a method of capturing data for future use); Matthew Roach, *Toward a New Language of Legal Drafting*, 17 J. HIGH TECH. L. 43, 46–48 (2016);

hoc approach to managing their contracts.[20] In many cases, little or no contract data are collected in a systematic manner. When contract data are collected, these data rarely include contract outcomes such as whether a contract resulted in litigation.[21] As a result, even if companies have data on the terms contained in their contracts, they cannot identify the effects of those terms on key outcomes without outcome data. Second, contract drafters have typically not had the analytical tools necessary to conduct robust analysis of contract data.[22] Many companies engage in low levels of systematic data analysis, or even forego the process entirely.[23] Even if companies do engage in analysis of contract data, it is

---

Robert E. Scott & George G. Triantis, *Anticipating Litigation in Contract Design*, 115 YALE L.J. 814, 822–23 (2006).

[20] *See* SAXENA, *supra* note 19, at 11–12.

[21] *See* Hill, *supra* note 18, at 69 (discussing how contract drafters lack the incentive to follow-up with a contract once it has been drafted and signed); George G. Triantis, *Improving Contract Quality: Modularity, Technology, and Innovation in Contract Design*, 18 STAN. J.L., BUS. & FIN. 177, 183–84 (2013) (noting that contract drafters pay little attention to the consequences of contract drafting).

[22] *See* LAWRENCE S. MAISEL & GARY COKINS, PREDICTIVE BUSINESS ANALYTICS: FORWARD LOOKING CAPABILITIES TO IMPROVE BUSINESS PERFORMANCE 62 (2014) (citing studies regarding how companies frequently fail to engage in adequate analysis); GRETTA RUSANOW, KNOWLEDGE MANAGEMENT AND THE SMARTER LAWYER 346 (2003) (describing how contract drafters easily get overwhelmed when trying to process contract data); Peter J. Gardner, *A Role for the Business Attorney in the Twenty-First Century: Adding Value to the Client's Enterprise in the Knowledge Economy*, 7 MARQ. INTELL. PROP. L. REV. 17, 44–45 (2003) (discussing the difficulty of processing large amounts of contract information in a meaningful way in a short amount of time); Hill, *supra* note 18, at 76 (discussing how "noise" in contract outcomes makes it difficult to determine the effects of terms on outcomes); Daniel Martin Katz, *Quantitative Legal Prediction—Or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L.J. 909, 928–29 (2013) (noting that even if lawyers had all the data they could ask for, it would be impossible for them to process these data using traditional mental heuristic models).

[23] *See* MAISEL & COKINS, *supra* note 22, at 62.

typically summarization of historical terms rather than prediction of future outcomes.[24]

Recent technological developments in contract data collection and analysis are lowering the above barriers.[25] Building on these developments, this Article proposes "predictive contracting," a new method of contracting in which contract drafters can design contracts using a technology system that helps predict the connections between contract terms and outcomes. For example, a predictive contracting system with data on the terms and outcomes of thousands of prior procurement contracts could inform a contract drafter that version A of a delivery term is ten percent more likely to result in late performance by a particular type of counterparty than version B. Predictive contracting will be powered by machine learning[26] and draw on contract data obtained from integrated contract management systems,[27] natural language processing,[28] and computable contracts.[29] Initially, predictive contracting will be applied to relatively simple, high volume contracts, such as sales and nondisclosure agreements. As predictive contracting systems improve over time, they can begin to be applied to more complex contracts, such as financing and acquisition agreements. Unlike previous

---

[24] *Id.* at 5.

[25] In the words of Oliver Williamson, "[b]ut for the limited ability of human agents to receive, store, retrieve and process data, interesting economic problems vanish." Oliver E. Williamson, *Transaction-Cost Economics: The Governance of Contractual Relations*, 22 J.L. & Econ. 233, 234 n.5 (1979).

[26] Machine learning is a category of artificial intelligence research that focuses on building mathematical computer models that learn from data to improve over time. *See infra* Section II.A.

[27] Contract management refers to a broad category of workflow processes and technology systems that allow companies to track and manage their contracts from beginning to end. *See infra* Section II.C.1.

[28] Natural language processing is a category of machine learning research focused on enabling computers to understand natural language communication, such as documents written in English. *See infra* Section II.C.2.

[29] A contract is computable if it is both machine-readable and machine-executable. *See infra* Section II.C.3.

contract automation mechanisms, such as LegalZoom, that have primarily focused on making contracts cheaper and faster to draft,[30] predictive contracting aims to substantively improve contract design by statistically connecting terms to outcomes.

This Article makes both theoretical and practical contributions to the contracts literature. On a theoretical level, predictive contracting can lead to greater customization, increased innovation, more complete contract design, more effective balancing of front-end and back-end costs, better risk assessment and allocation, and more accurate term pricing for negotiation. On a practical level, predictive contracting has the potential to significantly alter the role of transactional lawyers by providing them with access to previously unavailable information on the statistical connections between contract terms and outcomes. In addition to these theoretical and practical contributions, the Article also anticipates and addresses limitations and risks of predictive contracting, including technical constraints, concerns regarding data privacy and confidentiality, the regulation of the unauthorized practice of law, and the potential for exacerbating information inequality.

The remainder of this Article proceeds as follows. Part II introduces predictive contracting, discusses and provides examples of the underlying technologies, and distinguishes predictive contracting from prior versions of contract automation. Part III discusses the theoretical and practical implications of predictive contracting and addresses limitations and risks. The Article ends with a short conclusion that discusses opportunities for further research.

---

[30] *See* GULATI & SCOTT, *supra* note 18, at 6 (citing an interview with a transactional lawyer in which the lawyer describes how contract automation technology allows an associate to draft a sovereign bond contract in only three and a half minutes); Triantis, *supra* note 21, at 179 (discussing how contract automation has focused on commoditizing transactional legal work with the goal of cutting costs rather than improving contract quality).

II. PREDICTIVE CONTRACTING

This Part introduces predictive contracting, discusses and provides examples of the underlying technologies, and distinguishes predictive contracting from previous versions of contract automation. Predictive contracting is a new method of contracting in which contract drafters can design contracts using a technology system that helps predict the connections between contract terms and outcomes given a set of exogenous conditions. Figure 1 depicts predictive contracting.

Figure 1: Predictive Contracting[31]

Terms		Conditions		Outcomes
T_1		C_1		O_1
T_2	$+$	C_2		O_2
T_3		C_3		O_3
$\vdots$		$\vdots$		$\vdots$
T_x		C_y		O_z

Insights from Predictive
Contracting System

---
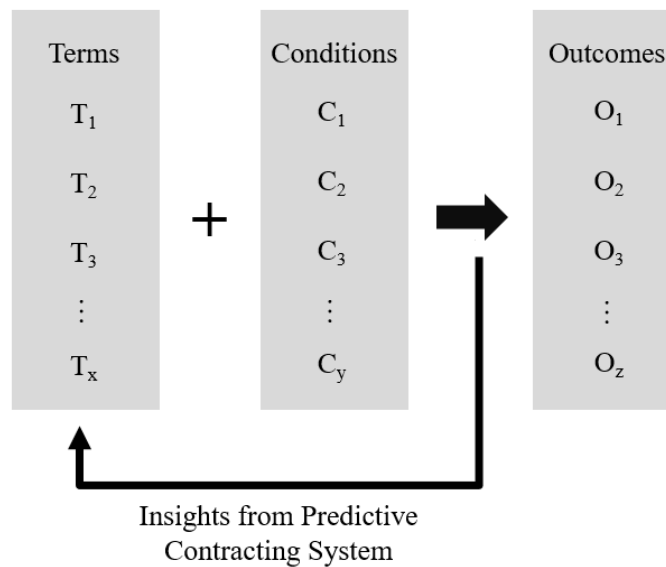
[31] The total number of variables for terms, conditions and outcomes (represented by the subscripts x, y, and z) do not need to be equal. For example, a predictive contracting scenario could contain three terms, ten exogenous conditions, and two outcomes.

As can be seen in Figure 1, contract terms result in contract outcomes given a set of exogenous conditions.[32] The goal of predictive contracting is to help contract drafters predict how terms are connected to outcomes given specified conditions. Contract drafters can then use this information to iteratively improve contract design over time.

A hypothetical example of predictive contracting can be illustrative. In this example, assume a company enters into simple sales contracts with its customers. One of the terms in the template sales contract is a choice of law provision between State A and State B. Looking to gain insights into the effects of this choice of law provision, the company compiles a data set of numerous past sales contracts. This data set contains data on (1) whether a contract used State A or State B for the choice of law provision, (2) whether the contract resulted in arbitration and if so the costs associated with the arbitration, and (3) a variety of exogenous conditions including demographic data on the counterparty. The company then uses this data to build a predictive contracting model to see if the choice of law provision has an impact on the likelihood and costliness of arbitration. The model identifies that for a specific category of counterparty, State A reduces overall arbitration costs, but for all other counterparties, State B reduces arbitration costs. The company uses this insight to set State B as the default for the choice of law provision except when dealing with this particular type of counterparty, thereby reducing overall arbitration costs.

As the example shows, a predictive contracting system is comprised of two primary technical components: (1) an analytical model and (2) a data set of contract terms, outcomes and exogenous conditions. These components are addressed in the Sections below. As these Sections will demonstrate, the individual technical components of predictive contracting

---

[32] In most cases, outcomes will result probabilistically from terms and exogenous conditions, though in some instances the association may be determinative. For a discussion of how contract outcomes result probabilistically from external contingencies and conditional instructions contained within a contract, see GULATI & SCOTT, *supra* note 18, at 143.

already exist and are currently being used in a variety of other real-world contexts. The existence of these technologies supports the technical feasibility of predictive contracting. The Sections below provide examples of these technologies and discuss how they will facilitate predictive contracting.

Predictive contracting systems will likely come in two forms: (1) systems built by third-party technology companies that are sold as software solutions to customers, and (2) systems built in-house by large companies. Third-party predictive contracting systems will primarily be marketed to small and midsized companies. While most of these companies will not have an incentive to develop predictive contracting on their own due to relatively low contracting volume, third-party technology providers can overcome this incentive problem by selling predictive contracting systems to multiple small and midsized customers. In addition, technology providers will be able to supply the complex technical expertise needed to develop a predictive contracting system that small-to-midsized companies may lack. This pattern is already observable in the contract management industry.[33] Large companies, on the other hand, will likely have sufficient contracting volume to be incentivized to develop predictive contracting in-house and the resources to do so. In both cases, predictive contracting systems will be designed such that contract drafters can use them without a background in statistics or computer science.

Predictive contracting will initially be applied to relatively simple, high-volume contracts, such as sales and nondisclosure agreements. In addition, predictive contracting systems will begin by examining relatively narrow problem specifications with a limited number of terms, outcomes, and exogenous conditions. This is because the ability of a model to predict outcomes of a system decreases as the complexity of the system increases.[34] Furthermore, the predictive capability of a model generally increases as the amount of data the model has access to increases. Therefore, examining narrow

---

[33] *See infra* Section II.C.1.

[34] *See* Katz, *supra* note 22, at 959–63.

problems related to simple contracts with large amounts of data is an ideal starting point for early versions of predictive contracting. As predictive contracting models improve and contract data becomes more robust, contract drafters can begin to use predictive contracting to analyze broader problems and more complex contracts such as financing and acquisition agreements.

The remainder of this Part proceeds as follows. Section II.A discusses the predictive contracting model, Section II.B examines the necessary types of contract data, Section II.C discusses potential sources of these data, and Section II.D distinguishes predictive contracting from prior versions of contract automation.

A. Model

The predictive contracting model is the analytical mechanism that uses contract data to provide contract drafters with insights into the statistical connections between contract terms and outcomes given exogenous conditions. To build a predictive contracting model, developers are likely to rely heavily on a rapidly growing area of analytical innovation: machine learning. Machine learning is a category of artificial intelligence research that focuses on building mathematical computer models that learn from data to improve over time.[35] Unlike earlier versions of artificial intelligence that attempted to replicate the way the human mind learns, machine learning instead seeks to achieve analytical results by using data-driven statistical models powered by computer processors.[36] Machine learning models

---

[35] *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89 (2014).

[36] *See id.* at 95–100; *see also* Katz, *supra* note 22, at 913–18 (discussing how the increase in data-driven predictive analysis is made possible in part due to the continually increasing power of computer processors described by Moore's Law and the continually decreasing cost of data storage described by Kryder's Law). For a discussion of Moore's Law, see Gordon E. Moore, *Cramming More Components onto Integrated Circuits*, 38 ELECTRONICS 114 (1965). For a discussion of Kryder's Law, see Chip Walter, *Kryder's Law*, SCI. AM., Aug. 2005, at 32.

have two main advantages over alternative analytical methods such as causal inference. First, machine learning models improve over time as more data are added to the data set from which the model learns, which is known as the "training set."[37] Second, machine learning models do not necessarily require a pre-specified relationship between independent and dependent variables.[38] This "black box" approach allows machine learning models to provide valuable predictive insights without the user needing to specify (or even understand) the potential relationships between variables in the model.[39]

Thus far, machine learning in the legal industry has primarily been applied to litigation issues such as discovery,[40] legal search,[41] the setting of bail,[42] and even jury selection.[43] Perhaps the most interesting application of machine learning to the law (and the most relevant for predictive contracting) has been the prediction of case outcomes and judicial decisions.[44] Previously an academic endeavor, multiple

---

[37] *See* Surden, *supra* note 35, at 92–93.

[38] *See* Katz, *supra* note 22, at 949–53.

[39] *Id.* While machine learning models do not necessarily require the user to prespecify a relationship between the variables in the model, the user must still select the data upon which the model is trained. For a discussion of the risks associated with the "black box" nature of machine learning models, see *infra* Section III.C.1.

[40] *See* Katz, *supra* note 22, at 945; John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services*, 82 FORDHAM L. REV. 3041, 3047–48 (2014); Mark McKamey, *Legal Technology: Artificial Intelligence and the Future of Law Practice*, 22 APPEAL 45, 49 (2017).

[41] *See* McGinnis & Pearce, *supra* note 40, at 3048–50.

[42] *See* Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401, 1428 (2017).

[43] *See Social-Enriched Voir Dire*, VIJILENT, https://www.vijilent.com/jury-selection [https://perma.cc/QDP5-WRNJ]; VOLTAIRE, https://voltaire app.com [https://perma.cc/NG3R-A5PM]. Voltaire is a technology system for jury selection that is powered by machine learning. *See id.*

[44] *See* Katz, *supra* note 22, at 936–39; Andrew D. Martin, Kevin M. Quinn, Theodore W. Ruger & Pauline T. Kim, *Competing Approaches to Predicting Supreme Court Decision Making*, 2 PERSP. ON POL. 761, 761 (2004); McGinnis & Pearce, *supra* note 40, at 3052–53; Theodore W. Ruger,

companies[45] are now using machine learning to engage in case prediction.[46] Notable among these companies is Judicata.[47] Judicata has built a machine learning prediction model trained on publicly available case law and opinions.[48] Users upload court documents for analysis such as motions and briefs as well as contextual information, such as the cause of action, the identity of the judge, and the location of the court.[49] Judicata analyzes the document and generates a report that assess the document in three categories: drafting, arguments, and context.[50] With respect to drafting, Judicata analyzes the document's citations and quotes for errors and potentially

Pauline T. Kim, Andrew D. Martin & Kevin M. Quinn, *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150, 1151–59 (2004); Surden, *supra* note 35, at 108–10; Dana Remus & Frank Levy, *Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law* 30 (Nov. 27, 2016) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2701092 [https://perma.cc/W5TY-EC28]; Daniel Martin Katz, Michael J. Bommarito II & Josh Blackman, *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, PLOS ONE (Apr. 12, 2017), http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174698 [https://perma.cc/JMQ4-EG48].

[45] Much of this section is based on interviews conducted by the author, which are cited throughout. These individuals spoke on conditions of anonymity and therefore, out of respect for their privacy, names are omitted from the citations. Notes from each interview are on file with the *Columbia Business Law Review*.

[46] *See, e.g.*, LEXMACHINA, https://lexmachina.com [https://perma.cc/F37M-B6G2]; PREMONITION, https://premonition.ai [https://perma.cc/WFS3-8837]; RAVEL, http://ravellaw.com [https://perma.cc/4AHT-VHVB].

[47] *See* JUDICATA, https://www.judicata.com [https://perma.cc/2PZP-3T6A]; *see also* Itai Gurari, *From Judging Lawyers to Predicting Outcomes*, OFFICIAL JUDICATA BLOG (Feb. 6, 2018), https://blog.judicata.com/from-judging-lawyers-to-predicting-outcomes-f46aedeb8684 (on file with the *Columbia Business Law Review*); Beth Hoover, *Judge Insights: Understanding the Forest and the Trees*, OFFICIAL JUDICATA BLOG (June 28, 2017), https://blog.judicata.com/judge-insights-understanding-the-forest-and-the-trees-c3164b767a4c (on file with the *Columbia Business Law Review*).

[48] Telephone Interview with Judicata Representatives (Apr. 2, 2018).

[49] *Id.*

[50] *Id.*

recommends better sources.[51] With respect to arguments, the system breaks the document down into its constituent arguments and displays the statistical favorability of those arguments based on past cases.[52] The system can also recommend missing arguments that it believes should be in the document.[53] Lastly, with respect to context, Judicata will provide the user with information on the outcomes of past cases with similar contextual characteristics.[54] In the future, the company plans to enable the system to generate first drafts of litigation documents.[55]

Machine learning is also beginning to be applied in the contracting context. Numerous contract technology companies are leveraging machine learning to provide contract drafters with insightful analysis. Contract Standards[56] and Legal Robot[57] are using machine learning to enhance compliance efforts by creating a map of a contract's constituent parts that can be connected with a map of an area of regulation to determine if there are any regulatory conflicts.[58] Legal technology companies are also using machine learning for predictive analysis. For example, Kira Systems[59] is working on a machine learning model for risk prediction in corporate acquisitions.[60] Contract drafters would be able to use this model to identify the likelihood of litigation risk associated

---

[51] *Id.*

[52] *Id.*

[53] *Id.*

[54] *Id.*

[55] *Id.*

[56] CONTRACT STANDARDS, https://www.contractstandards.com/ [https://perma.cc/8LAG-CEH5].

[57] LEGAL ROBOT, https://www.legalrobot.com/ [https://perma.cc/VP2F-HYMG].

[58] Telephone Interview with Contract Standards Representative (Mar. 6, 2018); Telephone Interview with Legal Robot Representative (Mar. 14, 2018). Contract Standards has mapped HIPAA and Legal Robot has mapped regulatory changes pertaining to Brexit, the United Kingdom's exit from the European Union.

[59] KIRA, https://kirasystems.com/ [https://perma.cc/P4UD-S4G3].

[60] Telephone Interview with Kira Systems Representative (Mar. 12, 2018).

with specific terms in acquisition agreements based on data of past agreements and litigation.[61]

A predictive contracting system using a machine learning model would function as follows. First, a predictive contracting technology provider (or a group within a large company building a predictive contracting system) would compile a data set of contract terms, outcomes, and exogenous conditions.[62] This data set would serve as the training set for the machine learning model. The predictive contracting company would then train a model based on this data set that would identify connections between terms and outcomes of interest given a set of exogenous conditions. Contract drafters would then use this information to inform contract design when drafting subsequent contracts. Data on the terms, outcomes, and conditions associated with these subsequent contracts would be collected and periodically added to the training set to retrain the model.[63] As the data set expands over time, the model would become more powerful and therefore able to take on more complex prediction problems.

Building a machine learning model for a predictive contracting system will require real-world contract data and is beyond the scope of this Article. While there are numerous standard machine learning models available for prediction analysis,[64] the design of a predictive contracting model will

---

[61] *Id.*

[62] For a discussion of these different types of data, see *infra* Section II.B. For a discussion of potential sources of these data, see *infra* Section II.C.

[63] This is similar to how insurance companies use actuarial data to update the terms and conditions of insurance contracts. *See* Boardman, *supra* note 15, at 1114–16 ("Not only does past language become clearer over time in the insurer's eyes, but the cost of each clause becomes increasingly clear as actuarial data is collected and pooled.").

[64] For example, IBM offers off-the-shelf machine learning software through its Watson initiative. *See Watson Machine Learning*, IBM, https://www.ibm.com/cloud/machine-learning [https://perma.cc/6J2S-59E4]. Common machine learning models include support vector machines and random forest decision trees. Support vector machines are commonly used in classification problems to sort data into defined categories. *See generally* COLIN CAMPBELL & YIMING YING, LEARNING WITH SUPPORT VECTOR

depend in large part on the characteristics of the contracting problem being analyzed. Two key dimensions of any predictive contracting model will be: (1) whether the model is single-task or multi-task, and (2) whether the model is supervised or unsupervised. These dimensions are discussed in the Sections below.

1. Single-Task vs. Multi-Task

Single-task machine learning models are ideal for scenarios in which the objective of the model is to solve a single learning task.[65] In the context of predictive contracting, single-task learning will be effective when the outcomes being predicted are largely independent from one another.[66] For example, assume that a contract drafter wants to understand how a variety of contract terms and exogenous conditions affect two outcomes: the amount of drafting time spent obtaining internal approvals and the quality of service provided by the counterparty under the contract. These outcomes are unlikely to be related, so the contract drafter could use two separate single-task learning models to predict each outcome. The type of single-task model that is most effective will depend on factors such as the amount of data in the training set and whether the outcome variable is binary, categorical, or numerical.[67]

Multi-task machine learning models, on the other hand, are better suited for scenarios in which multiple learning

---

MACHINES (2011); NELLO CRISTIANINI & JOHN SHAWE-TAYLOR, AN INTRODUCTION TO SUPPORT VECTOR MACHINES AND OTHER KERNEL-BASED LEARNING METHODS (2000); INGO STEINWART & ANDREAS CHRISTMANN, SUPPORT VECTOR MACHINES (2008). Random forests use a series of decision trees to analyze classification and regression problems. *See generally* Tin Kam Ho, *Random Decision Forests*, 1 PROC. 3RD INT'L CONF. ON DOCUMENT ANALYSIS & RECOGNITION 278, 278–82 (1995); Leo Breiman, *Random Forests*, 45 MACHINE LEARNING 5 (2001); Tao Shi & Steve Horvath, *Unsupervised Learning with Random Forest Predictors*, 15 J. COMPUTATIONAL & GRAPHICAL STATS. 118, 118 (2006).

[65] Interview with Private Technology Company Representative, in S.F., Cal. (May 2, 2018).

[66] *Id.*

[67] *Id.*

tasks need to be solved at the same time.[68] Multi-task learning draws on similarities and differences between the different learning tasks to improve efficiency and accuracy.[69] In the context of predictive contracting, multi-task learning will be useful when the outcomes being predicted are related to one another.[70] In the previous example, assume instead that the contract drafter wants to predict both the likelihood that the counterparty renews the contract and the likelihood that the contract results in litigation. These outcomes are likely to be related (a counterparty is less likely to renew a contract that resulted in litigation), so the contract drafter could use a multi-task learning model to predict both outcomes. Multi-task learning will frequently utilize an artificial neural network (often called a "neural net"), a machine learning mechanism commonly used in image recognition.[71]

2. Supervised vs. Unsupervised

Supervised machine learning models learn from a training set of labeled data.[72] For example, a supervised model

---

[68] *See generally* Andreas Argyriou, Theodoros Evgeniou & Massimiliano Pontil, *Convex Multi-Task Feature Learning*, 73 MACHINE LEARNING 243 (2008); Rich Caruana, *Multitask Learning*, 28 MACHINE LEARNING 41 (1997); Carlo Ciliberto, Youssef Mroueh, Tomaso Poggio & Lorenzo Rosasco, Convex Learning of Multiple Tasks and Their Structure 1 (Apr. 17, 2015) (unpublished manuscript), https://arxiv.org/abs/1504.03101 [https://perma.cc/NW2K-JJJ3].

[69] *See supra* note 68.

[70] Interview with Private Technology Company Representative, *supra* note 65.

[71] *Id. See generally* SIMON HAYKIN, NEURAL NETWORKS AND LEARNING MACHINES (3d ed. 2009); Jürgen Schmidhuber, *Deep Learning in Neural Networks: An Overview*, 61 NEURAL NETWORKS 85 (2015).

[72] *See* MEHRYAR MOHRI, AFSHIN ROSTAMIZADEH & AMEET TALWALKAR, FOUNDATIONS OF MACHINE LEARNING (2012); STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (3d ed. 2010); Surden, *supra* note 35, at 93; Irene Ng (Huang Ying), *The Art of Contract Drafting in the Age of Artificial Intelligence: A Comparative Study Based on US, UK and Austrian Law* 25 (Transatlantic Tech. L.F. Working Paper No. 26, 2017); Remus & Levy, *supra* note 44, at 9–10.

designed to identify whether an image contains an apple would be trained on a data set of images that a human had gone through ahead of time and tagged which ones contained apples. Based on this labeled data, the model would learn to identify the image characteristics of an apple. In the predictive contracting context, a supervised model would be trained on labeled data of contract terms and outcomes. For example, a supervised predictive contracting model designed to identify connections between anti-dilution provisions in venture financing agreements and entrepreneur ownership at liquidation would be trained on a data set of contracts labeled with their type of anti-dilution provision and the percent of entrepreneur ownership at liquidation. The necessity of data labeling for supervised learning emphasizes the importance of collecting contract data that are properly formatted and structured for use in machine learning. Part II.C below discusses potential means of collecting contract data in this manner.

Unsupervised machine learning models, on the other hand, learn from unlabeled data.[73] In the apple image recognition example above, an unsupervised model would be provided with a data set of unlabeled images, some containing apples, some not. In the predictive contracting context, an example of unlabeled data would be a contract in Microsoft Word or PDF format with none of its terms labeled ahead of time by a human. Because unsupervised models do not require labeled data, they are well-suited for situations in which labeling data is difficult or impossible. Yet unsupervised models can prove less effective than supervised models in situations in which the user is interested in identifying connections between specific inputs and outputs because these inputs and outputs are not pre-defined by the user. For predictive contracting, if a contract drafter is interested in understanding the connections between a specific term or set of terms and a

---

[73] *See* UNSUPERVISED LEARNING: FOUNDATIONS OF NEURAL COMPUTATION (Geoffrey Hinton & Terrence J. Sejnowski eds., 1999); RICHARD O. DUDA, PETER HART & DAVID G. STORK, PATTERN CLASSIFICATION 517–600 (2d ed. 2001); Ng, *supra* note 72, at 25–26; Remus & Levy, *supra* note 44, at 10–11.

defined set of outcomes, a supervised model would likely prove more effective. As a result, early versions of predictive contracting will likely rely more heavily on supervised learning due to initial applications of predictive contracting being relatively specific and narrow in scope. As predictive contracting expands in scope to address more complex relationships between terms and outcomes, developers may turn to unsupervised learning.

B. Data

A predictive contracting system running on machine learning technology as discussed in Part II.A will require the following three categories of contract data: (1) terms, (2) outcomes, and (3) exogenous conditions. These data categories are discussed in the Sections below.

1. Terms

The first category of data needed for predictive contracting is data on contract terms. In the relationship depicted in Figure 1, terms are the endogenous inputs that, when combined with exogenous conditions, result in contract outcomes. Terms are endogenous to the relationship between the contracting parties because the parties chose to include the terms in the contract. This is true even of non-negotiated, boilerplate terms because the parties ultimately chose to enter into a contract containing these terms. As a result, contract terms are considered endogenous for the purposes of predictive contracting even in the case of consumer contracts of adhesion in which consumers are faced with "take-it-or-leave-it" contracting scenarios.[74] While contract design also

---

[74] *See generally* Oren Bar-Gill, *The Behavioral Economics of Consumer Contracts*, 92 MINN. L. REV. 749 (2008); Shmuel I. Becher, *Asymmetric Information in Consumer Contracts: The Challenge That Is Yet to Be Met*, 45 AM. BUS. L.J. 723 (2008); Friedrich Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629 (1943); Michael I. Meyerson, *The Efficient Consumer Form Contract: Law and Economics Meets the Real World*, 24 GA. L. REV. 583 (1990); Andrew A. Schwartz, *Consumer Contract Exchanges and the Problem of Adhesion*, 28 YALE J. ON

includes implied terms that are not expressly included in the contract but are inferred by courts in the event of litigation, these terms are difficult to incorporate into predictive contracting because data on them cannot be collected from contract documents like express terms.[75]

Contract terms can be analyzed as individual terms or as sets of multiple terms. The modular nature of contracts lends itself to breaking contracts down into their constituent terms.[76] For example, a contract could be broken down into Terms A, B, C, D, and E. A contract drafter could then use a predictive contracting system to analyze the effects of each of these terms on outcomes of interest given a set of exogenous conditions. Recent research, however, has highlighted the importance of the interconnectedness of contract terms to overall contract design.[77] In the previous example, the contract drafter could also use predictive contracting to analyze the joint effects of subsets of Terms A–E on the outcomes of interest.

Contract term data can broadly be classified as binary, categorical, or numerical.[78] Binary term data can be used to represent the presence or absence of a term in a contract. For example, a contract data set could contain binary data on whether or not an acquisition agreement contained an earnout provision. Binary term data is most useful for

---

REG. 313 (2011); W. David Slawson, *Standard Form Contracts and Democratic Control of Lawmaking Power*, 84 HARV. L. REV. 529 (1971); Anne Brafford, Note, *Arbitration Clauses in Consumer Contracts of Adhesion: Fair Play or Trap for the Weak and Unwary?*, 21 J. CORP. L. 331 (1996). For a discussion of the implications of predictive contracting for consumer contracts, see *infra* Section III.C.4.

[75] *See* Goetz & Scott, *supra* note 15, at 262. For a discussion of issues raised by terms and conditions that are difficult to capture, see *infra* Section III.C.1.

[76] *See* Cathy Hwang, *Unbundled Bargains: Multi-Agreement Dealmaking in Complex Mergers and Acquisitions*, 164 U. PA. L. REV. 1403, 1417–27 (2016) (discussing the modularity of merger and acquisition contracts).

[77] *See* Cathy Hwang & Matthew Jennejohn, *Deal Structure*, NW. U. L. REV. 279, 279–85 (2018).

[78] This is also true of outcome and condition data.

examining the effects of terms for which there is little or no variation in the form of the term across contracts other than whether the term is present or absent. Categorical term data can be used to represent different versions of a term.[79] For example, a contract data set could contain categorical term data on the type of anti-dilution provision contained in a venture financing agreement. Categories of terms can be ordered or unordered. Ordered categories represent a logical progression between the categories, such as different versions of a penalty term with increasing levels of severity. Categorical term data is most useful for examining the effects of terms for which there is meaningful variation in the form of the term across contracts. Numerical term data can be used to represent terms with magnitude. For example, a contract data set could contain numerical data on the interest rate in a debt contract or the price in a procurement contract.

2. Outcomes

The second category of data needed for predictive contracting is data on contract outcomes. In the relationship depicted in Figure 1, outcomes are the outputs that result from endogenous contract terms and exogenous conditions. Contract outcomes are a means of assessing and measuring a contract. Predictive contracting aims to help contract drafters understand and predict the effects of terms on outcomes given conditions so that drafters can design subsequent contracts that result in better outcomes. The improvement of contract outcomes is therefore the ultimate goal of predictive contracting. In most contracting scenarios, parties and their drafters will have to balance certain expected outcomes against others when designing a contract. This aspect of contract design will become a key role for transactional lawyers in the presence of predictive contracting.[80]

---

[79] Note that for categorical term data, one of the versions of a term could be the absence of the term.

[80] *See infra* Section III.B.

Contract outcomes can assess both the front-end and back-end of a contract's life.[81] The front-end is when the parties and their agents negotiate and design the contract. Examples of front-end outcomes include how long the contract took to negotiate and how much the contract cost to draft. The back-end is when the parties perform the obligations under the contract and/or potentially dispute the contract. Examples of back-end outcomes include the timing and quality of counterparty performance and whether the contract resulted in litigation. Contract outcomes can also be objective or subjective. Examples of objective outcomes include whether the contract was amended and the amount of any payments made under the contract. Examples of subjective outcomes include whether the parties believed the contract adequately met their needs and whether the contract resulted in any reputational effects for the parties.

Predictive contracting users will likely use a wide variety of outcomes to assess their contracts, including front-end, back-end, objective, and subjective outcomes.[82] While many outcomes will be specific to the contracting scenario being analyzed, some general outcomes of interest include:

- The amount of time to negotiate and draft the contract

---

[81] *See* Albert Choi & George Triantis, *Strategic Vagueness in Contract Design: The Case of Corporate Acquisitions*, 119 YALE L.J. 848, 851 (2010); Richard A. Posner, *The Law and Economics of Contract Interpretation*, 83 TEX. L. REV. 1581, 1583–84 (2005); Scott & Triantis, *supra* note 19, at 814; Scott & Triantis, *supra* note 16, at 190. For a proposed model that includes a third "midstream" stage, see Triantis, *supra* note 21, at 183–84.

[82] *See* Stephen J. Choi & G. Mitu Gulati, *Contract as Statute*, 104 MICH. L. REV. 1129, 1158 (2006) (discussing how contracting parties balance several goals when designing a contract); D. Gordon Smith & Brayden G. King, *Contracts as Organizations*, 51 ARIZ. L. REV. 1, 25–26 (2009) (applying organization theory to contracts to argue that contracts are strategically important mechanisms that enable an organization to advance its strategic goals); Ron Dolin, Measuring Legal Quality 2 (June 18, 2017) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2988647 [https://perma.cc/WX5D-AHLD] (citing an example of DuPont developing a diverse set of assessment outcomes for analyzing its litigation portfolio).

- The cost to negotiate and draft the contract
- The extent of deviation between the first draft and the final draft
- The timing of counterparty performance
- The quality of counterparty performance
- Whether and how the contract was amended
- Whether and why the contract resulted in a dispute
- If the contract resulted in a dispute, how the dispute was resolved (negotiation, arbitration, litigation, etc.)
- If the contract resulted in a dispute, the cost of the dispute
- The total cost of negotiating, drafting, administering, and resolving the contract
- Whether the parties were satisfied with the contract
- Whether the contract resulted in any reputational effects for the parties

3. Conditions

The third category of data needed for predictive contracting is data on exogenous conditions. In the relationship depicted in Figure 1, conditions are the exogenous inputs that, when combined with endogenous terms, result in contract outcomes. Contracting does not exist in a vacuum, but rather against a backdrop of external factors that can influence contract design and outcomes.[83] For example, whether the parties to an acquisition agreement are public or private can have substantial effects on the design of the agreement and numerous outcomes of interest. While some insights could be gained from a predictive contracting model trained only on terms and outcomes, the quality of the

---

[83] *See* Williams, *supra* note 2, at 149–54 (demonstrating that the total supply of venture capital financing had statistically significant connections with a variety of contract terms based on a set of over 5000 venture capital financing contracts from 2004–2015). The author controlled for a number of other external factors such as industry, location, and the risk-free treasury rate. *See id.* at 151–53.

results generated by the model can be improved by including data on relevant exogenous conditions.[84]

For the purposes of predictive contracting, a condition is considered exogenous if one or more of the parties cannot feasibly modify it as part of the contract design process. This includes conditions over which the parties have no control, such as general economic conditions and geopolitical factors. This also includes conditions over which the parties do have control but cannot feasibly modify as part of the contracting scenario being analyzed. For example, while the location of a party's headquarters may be a relevant condition for a simple procurement contract, and while the party does have control over this condition, it cannot feasibly move its headquarters for the purposes of designing the contract. As a result, the location of the party's headquarters would be considered an exogenous condition when analyzing the procurement contract using a predictive contracting system.

While many conditions will be specific to the contracting scenario being analyzed, some general conditions include:

- Party characteristics (identity, location, size, industry, etc.)
- Drafter characteristics (identity, location, law firm, etc.)
- Whether the parties have a preexisting relationship and the nature of that relationship
- Whether the contract is based on a prior contract between the parties
- General industry conditions
- General economic conditions
- Geopolitical conditions

---

84 Determining which exogenous conditions are "relevant" will require iterative trial and error. Contract drafters will initially include exogenous conditions for which they have data and believe could potentially influence design and/or outcomes. The models they develop will then provide insight into whether those conditions are in fact relevant.

C. Data Sources

Data on contract terms, outcomes, and conditions for use in predictive contracting will primarily come from three sources: (1) contract management systems, (2) natural language processing, and (3) computable contracts. These data sources are discussed in the Sections below.

1. Contract Management

Contract management refers to a broad category of workflow processes and technology systems that allow companies to track and manage their contracts from beginning to end.[85] While many companies still manage their contracts through a combination of email and Excel spreadsheets, a growing percentage of companies are turning to dedicated contract management systems.[86] Contract management systems include both systems developed in-house for use by a single company,[87] as well as third-party contract management providers that sell contract management software to a wide range of customers.[88] While many companies still use contract management systems primarily as repositories for contract documents, a growing

---

[85] *See* SAXENA, *supra* note 19, at 12 ("Enterprise Contract Management (ECM) encompasses a wide spectrum of applications, protocols, and systems for managing an enterprise's contracts from A to Z.").

[86] In a 2017 survey by SpringCM, thirty-two percent of respondents reported that they manage their contracts with a contract management tool. *See* MATT STERN, SPRINGCM, 2017 STATE OF CONTRACT MANAGEMENT REPORT 9 (2017) (on file with the *Columbia Business Law Review*).

[87] Interview with Oracle Representative, in S.F., Cal. (Apr. 10, 2018); Telephone Interview with Airbnb Representatives (Apr. 26, 2018); Telephone Interview with Microsoft Representatives (Apr. 12, 2018); Telephone Interview with Public Technology Company Representative (Apr. 30, 2018).

[88] For a list of some of the primary contract management providers, see *Contract Management Software*, CAPTERRA, https://www.capterra.com/contract-management-software [https://perma.cc/9E2Z-G2FA].

number of companies are starting to use these systems as data sources for a variety of applications.[89]

Contract management systems increase the availability of data on contract terms. Companies use contract management systems to track data on key terms as well as conditions such as party identities, locations, and dates.[90] Companies will often begin the contracting process with an internal template.[91] Contract drafters will negotiate the template-based contract with the counterparty based on a set of pre-approved negotiating ranges for various terms.[92] Any modifications that fall outside these pre-approved ranges must typically go through an internal approval process.[93] Once the contract has been finalized, contract drafters can track any deviations from the template terms in the contract

---

[89] *See* Andrew Bartels & Charlotte Wang, *The Forrester Wave™: Contract Life-Cycle Management, Q3 2016*, FORRESTER, July 25, 2016, at 2–3 (on file with the *Columbia Business Law Review*).

[90] Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Dell Representative (Apr. 11, 2018); Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Private Technology Company Representative (Apr. 13, 2018); Telephone Interview with Public Technology Company Representative, *supra* note 87.

[91] Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Dell Representative, *supra* note 90; Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Private Technology Company Representative, *supra* note 90; Telephone Interview with Public Technology Company Representative, *supra* note 87.

[92] Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Dell Representative, *supra* note 90; Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Private Technology Company Representative, *supra* note 90; Telephone Interview with Public Technology Company Representative, *supra* note 87.

[93] Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Dell Representative, *supra* note 90; Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Public Technology Company Representative, *supra* note 87.

management system.[94] In some cases, these systems can be used to collect data on hundreds of contract terms. For example, Dell works with Axiom, an alternative legal services company that provides contract management solutions, to collect over three hundred data points from each of its contracts.[95] This process of tracking contract terms is even more streamlined if a company uses an end-to-end contract management system that also supports drafting and negotiation. These systems allow contract drafters to draft and negotiate contracts entirely within the system, thereby increasing the ability to collect data on contract terms.[96] For example, Icertis is a leading contract management company that provides customers with an integrated, end-to-end system within which they can draft, negotiate, and track all of their contracts.[97]

In addition to collecting data on contract terms, contract management systems can also collect valuable data on contract outcomes. This includes front-end outcomes such as the time required to draft and negotiate the contract, overall drafting costs, and term-by-term negotiating outcomes.[98] For example, Contract Room enables contract drafters to collect front-end outcomes, such as, how long a contract takes to

---

[94] Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Dell Representative, *supra* note 90; Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Public Technology Company Representative, *supra* note 87.

[95] Telephone Interview with Axiom Representative (Mar. 9, 2018); Telephone Interview with Dell Representative, *supra* note 90.

[96] *See* Interview with Contract Room Representative, in Palo Alto, Cal. (Apr. 9, 2018); Telephone Interview with Icertis Representative (Mar. 15, 2018); Telephone Interview with Sirion Labs Representative (Apr. 3, 2018); Telephone Interview with Sirion Labs Representative (Apr. 27, 2018).

[97] Telephone Interview with Icertis Representative, *supra* note 96; *see also* ICERTIS, https://www.icertis.com/contract-management-software/ [https://perma.cc/P6Z8-T8BA].

[98] Interview with Contract Room Representative, *supra* note 96; Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Public Technology Company Representative, *supra* note 87.

negotiate, how frequently certain terms are negotiated, which terms create the most negotiating roadblocks, and the total cost to draft and negotiate the final contract.[99] Contract management systems can also collect data on back-end outcomes such as payments and deliveries made under the contract, whether deal risks flagged during negotiation actually occur, and whether the contract results in a dispute (and if so, the outcome of the dispute).[100] For example, Sirion Labs enables contract drafters to collect data on numerous back-end contract outcomes.[101] For a particular contract, a Sirion user can see if the contract is in dispute, and if so, the current stage of the dispute.[102] The system links disputes to specific terms within the contract so the user can see which terms cause disputes.[103] Once a dispute is resolved, the system displays the outcome of the dispute and any associated costs.[104] Contract managers can also use Sirion to track counterparty performance under a contract.[105] For a contract with a server provider, for example, the system can track the percentage of time, within a defined period, during which the servers were online and running properly.[106] The system also

---

[99] Interview with Contract Room Representative, *supra* note 96.

[100] Interview with Contract Room Representative, *supra* note 96; Telephone Interview with Dell Representative, *supra* note 90; Telephone Interview with Microsoft Representatives, *supra* note 87; Telephone Interview with Sirion Labs representatives, *supra* note 96; Telephone Interview with Sirion Labs Representative, *supra* note 96.

[101] Telephone Interview with Sirion Labs Representative (Apr. 27, 2018), *supra* note 96; Telephone Interview with Sirion Labs Representative (Apr. 3, 2018), *supra* note 96.

[102] Telephone Interview with Sirion Labs Representative (Apr. 27, 2018), *supra* note 96; Telephone Interview with Sirion Labs Representative (Apr. 3, 2018), *supra* note 96.

[103] Telephone Interview with Sirion Labs Representative (Apr. 27, 2018), *supra* note 96; Telephone Interview with Sirion Labs Representative (Apr. 3, 2018), *supra* note 96.

[104] Telephone Interview with Sirion Labs Representative (Apr. 27, 2018), *supra* note 96; Telephone Interview with Sirion Labs Representative (Apr. 3, 2018), *supra* note 96.

[105] Telephone Interview with Sirion Labs representative (Apr. 27, 2018), *supra* note 96.

[106] *Id.*

uses customizable formulas to convert counterparty performance data into payment obligations under a contract.[107] As the use of contract management systems such as Sirion continue to grow, contract drafters will have increasingly better access to data on contract terms and outcomes.

2. Natural Language Processing

One of the primary hurdles to collecting data on contract terms for machine learning is that these data are stored in an unstructured format within natural language contract documents such as English-language Microsoft Word files and PDFs.[108] Despite contract documents containing significant amounts of data,[109] these data are not in a form that is easily useable for machine learning analysis due to their lack of structure and labeling.[110] To systematically analyze contract terms, companies have traditionally had to manually extract, structure, and label data from natural language documents, which is an incredibly time and labor-intensive process.[111] For example, some large law firms will have junior associates

---

[107] Telephone Interview with Sirion Labs Representative (Apr. 27, 2018), *supra* note 96; Telephone Interview with Sirion Labs Representative (Apr. 3, 2018), *supra* note 96.

[108] *See* Roach, *supra* note 19, at 46; Harry Surden, *Computable Contracts*, 46 U.C. DAVIS L. REV. 629, 642–44 (2012) (distinguishing "natural languages" such as English from "formal languages" such as computer programming languages).

[109] *See* Roach, *supra* note 19, at 50–51 (describing contracts as a mineable source of data).

[110] *See* Surden, *supra* note 108, at 642–44; Roach, *supra* note 19, at 46.

[111] Interview with Oracle Representative, *supra* note 87; Telephone Interview with Airbnb Representative, *supra* note 87; Telephone Interview with Contract Assistant Representative (Mar. 8, 2018); Telephone Interview with Private Technology Company Representative, *supra* note 90; Telephone Interview with Public Technology Company Representative (Apr. 4, 2018); Telephone Interview with Public Technology Company Representative, *supra* note 87.

review contracts after signing for the purpose of entering contract data into an internal database.[112]

Natural language processing (NLP) is a category of machine learning research focused on enabling computers to understand natural language communication.[113] Most NLP techniques are statistical in nature.[114] Drawing on a training set of existing natural language documents, NLP models can be trained to understand natural language text based on statistical relationships between components of the text such as individual words, groups of words, word sequencing, and physical layout features like paragraph breaks and page positioning.[115] An NLP model is often adjusted and retrained until it is sufficiently accurate at understanding natural language text.[116] The model can then be used to process new natural language documents outside of the training set. In the legal context, NLP has primarily been applied to litigation discovery to help human document reviewers sort through massive amounts of discovery documents.[117]

Numerous legal technology companies have begun to use NLP to extract structured contract term data from natural

---

[112] *See* Elisabeth de Fontenay, *Law Firm Selection and the Value of Transactional Lawyering*, 41 J. CORP. L. 393, 397 (2015). For example, an associate might note in the database whether a venture financing contract contains an anti-dilution provision, and if so, what type.

[113] *See* Surden, *supra* note 108, at 643. For an overview of NLP, see generally CHRISTOPHER D. MANNING & HINRICH SCHÜTZE, FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING 3–5 (1999); RUSSELL & NORVIG, *supra* note 72, at 860–67; Robert Dale, *Classical Approaches to Natural Language Processing, in* HANDBOOK OF NATURAL LANGUAGE PROCESSING 1–7 (Nitin Indurkhya & Frederick J. Damerau eds., 2d ed. 2010); Prakash M. Nadkarni, Lucila Ohno-Machado & Wendy W. Chapman, *Natural Language Processing: An Introduction*, 18 J. AM. MED. INFORM. ASS'N 544 (2011).

[114] *See* Surden, *supra* note 108, at 644.

[115] Telephone Interview with Kira Systems Representative (Mar. 5, 2018); Telephone Interview with Kira Systems Representative, *supra* note 60.

[116] Telephone Interview with LawGeex Representative (Mar. 8, 2018); Telephone Interview with LegalSifter Representative (Mar. 14, 2018).

[117] *See* Surden, *supra* note 108, at 644.

language contracts.[118] Using NLP to generate structured contract term data is far more efficient, cost-effective and scalable than the manual alternative. Many legal NLP companies also create application programming interfaces ("APIs") that allow their products to integrate with contract management systems.[119] This enables a company to track and use the contract data obtained via NLP within its contract management system. While non-legal NLP companies often use off-the-shelf NLP software,[120] legal NLP companies must typically create their own models due to the highly technical and unnatural nature of legalese.[121] For example, LawGeex developed their own NLP model specifically for understanding contractual legalese called Legalese Language Processing ("LLP").[122] LawGeex's proprietary LLP model was trained for over three years on over 400,000 contracts to understand the unique phrasing, sentence structure, and terminology of contractual legalese.[123]

The main differentiating factor among legal NLP products is whether the NLP model is pretrained. Pretrained (also known as "out-of-the-box") models are typically trained on large data sets (thousands, tens of thousands, or even hundreds of thousands) of relatively simple contracts such as

---

[118] *See Our Services*, CONTRACTSTANDARDS, https://www.contract standards.com/Services [https://perma.cc/48YC-3R3V]; EBREVIA, https://ebrevia.com [https://perma.cc/6K6K-H8T8]; *How Kira Works*, KIRA SYSTEMS, https://www.kirasystems.com/how-it-works [https://perma.cc/N9N9-CXUU]; LAWGEEX, https://www.lawgeex.com [https://perma.cc/ZM36-GG3U]; LEGAL ROBOT, https://www.legalrobot.com [https://perma.cc/K3TL-WJTK]; LEGALSIFTER, https://www.legalsifter.com [https://perma.cc/Q2W9-HM7Q].

[119] Telephone Interview with Kira Systems Representative, *supra* note 115; Telephone Interview with Contract Standards Representative, *supra* note 58; Telephone Interview with Beagle Representative (Mar. 9, 2018); Telephone Interview with eBrevia Representative (Apr. 6, 2018).

[120] Telephone Interview with Legal Robot Representative (Mar. 14, 2018).

[121] Telephone Interview with LawGeex Representative (Apr. 11, 2018).

[122] *Id.*

[123] *Id.*; *see also* Telephone Interview with LawGeex Representative (Mar. 8, 2018).

sales and nondisclosure agreements.[124] For example, Contract Standards trained its pretrained model on publicly available contracts obtained through the Securities and Exchange Commission's EDGAR database.[125] The advantage of pretrained models is that users can apply them immediately without having to train the models themselves.[126] The downside, however, is that pretrained models cannot be used to understand types of contracts and terms that are not contained within the supplied training set.[127] As a result, pretrained models are not applicable for more niche and complex types of contracts. User-trained models, on the other hand, can be applied to any type of contract, but the user must supply the contracts that make up the training set.[128] The number of contracts needed for a user to train a model with sufficient accuracy depends on the complexity and variability of the contract—the more complex and variable the terms in the contract, the larger the required training set.[129] For example, Kira Systems offers a user-trained model that can be applied to any type of contract.[130] To use the model, the

[124] Telephone Interview with Contract Standards Representative, *supra* note 58; Telephone Interview with eBrevia Representative, *supra* note 119; Telephone Interview with Kira Systems Representative, *supra* note 115; Telephone Interview with Kira Systems Representative, *supra* note 60; Telephone Interview with LawGeex Representative, *supra* note 123; Telephone Interview with LawGeex Representative, *supra* note 121; Telephone Interview with LegalSifter Representative, *supra* note 116.

[125] Telephone Interview with Contract Standards Representative, *supra* note 58; *see Filings & Forms*, U.S. SEC. & EXCHANGE COMMISSION, https://www.sec.gov/edgar.shtml [https://perma.cc/GK9T-6539].

[126] Telephone Interview with LawGeex Representative, *supra* note 123; Telephone Interview with LawGeex Representative, *supra* note 121.

[127] Telephone Interview with LawGeex Representative, *supra* note 123; Telephone Interview with LawGeex Representative, *supra* note 121.

[128] Telephone Interview with Beagle Representative, *supra* note 119; Telephone Interview with eBrevia Representative, *supra* note 119; Telephone Interview with Kira Systems Representative, *supra* note 115; Telephone Interview with Kira Systems Representative, *supra* note 60; Telephone Interview with LegalSifter Representative, *supra* note 116.

[129] Telephone Interview with eBrevia Representative, *supra* note 119.

[130] Telephone Interview with Kira Systems Representative, *supra* note 115; Telephone Interview with Kira Systems Representative, *supra* note 60.

user must provide at least fifty contracts in which the terms of interest have been pre-labeled by the user.[131] The user then clicks a button labeled "Train," which trains the model on the contracts provided.[132] After the model has finished training, the system displays the model's accuracy.[133] One legal NLP company, LegalSifter, has developed a hybrid NLP product that resembles both a pretrained and a user-trained model.[134] LegalSifter will work with users to develop user-trained NLP models specifically for a user's niche contracts and terms.[135] LegalSifter then makes these models available to other users with similar niche contracts.[136] The models are retrained every week to take into account feedback and new data from all users.[137] Through this process, LegalSifter can effectively crowdsource the training of new models for any type of contract.[138] Legal NLP products—including pretrained, user-trained, and hybrid models—will increase the availability and quality of data on contract terms.

3. Computable Contracts

While less developed than contract management systems and natural language processing, computable contracts present a compelling opportunity for expanding the availability and quality of contract data. A contract is "computable" if it is both machine-readable and machine-executable.[139] A contract is machine-readable if it is expressed

---

[131] Telephone Interview with Kira Systems Representative *supra* note 115.

[132] *Id.*

[133] Telephone Interview with Kira Systems Representative, *supra* note 115; Telephone Interview with Kira Systems Representative, *supra* note 60.

[134] Telephone Interview with LegalSifter Representative, *supra* note 116.

[135] *Id.*

[136] *Id.*

[137] *Id.*

[138] *Id.*

[139] *See* Surden, *supra* note 108, at 634–36. Computable contracts are often referred to as "smart" contracts. This Article uses the term "computable" rather than "smart" because "computable" addresses the

in a format that can be processed by a computer.[140] As discussed above, most contracts are written in a natural language, such as English, that is not inherently interpretable by a computer. While natural language processing techniques are starting to enable computers to understand natural languages,[141] researchers in a related field of computer science are developing computer programming languages that can be used to express contracts in a fully machine-readable format.[142] For example, Sudhir Agarwal, Kevin Xu, and John Moghtader recently developed a computable contracting language they refer to as Contract Definition Language, which they used to model HIPAA regulations.[143] A computable contract is also machine-executable, which means the contract can be automatically executed when supplied

---

machine-interpretability that is at the heart of computable contracts whereas "smart" can mean many different things in different contexts.

[140] *Id.* at 639.

[141] *See supra* Section II.C.2.

[142] *See generally* Roach, *supra* note 19, at 54–59; Mark D. Flood & Oliver R. Goodenough, *Contract as Automaton: The Computational Representation of Financial Agreements* (Office of Fin. Research Working Paper No. 15-04, 2017), https://www.financialresearch.gov/working-papers/2015/03/26/contract-as-automation [https://perma.cc/4C9T-LPS2]; Tom Hvitved, Contract Formalisation and Modular Implementation of Domain-Specific Languages (2012) (unpublished Ph.D. dissertation, University of Copenhagen), https://pdfs.semanticscholar.org/33db/bb29ed7e 5b7c58dc651a8 c3223bc9711a863.pdf (on file with the *Columbia Business Law Review*); Ronald M. Lee, Candid: A Formal Language for Electronic Contracting (Aug. 1998) (unpublished manuscript), https://papers.ssrn.com /sol3/papers.cfm?abstract_id=2071383 [https://perma.cc/Z2BZ-TTA7]; Seyed Morteza Montazeri, Nivir Kanti Singha Roy & Gerardo Schneider, From Contracts in Structured English to CL Specifications (Sept. 13, 2011) (unpublished manuscript), https://arxiv.org/abs/1109.2657 [https://perma. cc/LR9D-YUK7]; Nick Szabo, A Formal Language for Analyzing Contracts (2002) (unpublished manuscript), http://nakamotoinstitute.org/contract-language [https://perma.cc/9ZC2-MVKY].

[143] *See* Sudhir Agarwal, Kevin Xu & John Moghtader, *Toward Machine-Understandable Contracts*, *in* A14J – Artificial Intelligence for Justice, Workshop at the 22nd European Conference on Artificial Intelligence (Aug. 30, 2016), http://www.ecai2016.org/content/uploads/ 2016/08/W2-ai4j-2016.pdf [https://perma.cc/KYR2-PC6T].

with real-world performance data.[144] For example, a computable weather derivative contract could automatically transfer money between the parties based on real-world weather data. While Sirion Labs does not automatically execute contracts between parties, their feature that allows a user to calculate contractual obligations based on real-world performance data is a simplified version of a machine-executable contract.[145]

Once fully developed, computable contracts will be the ideal mechanism for collecting contract data.[146] Data on contract terms can be easily collected from a computable contract because the contract is already written in a structured, machine-readable format. This is a substantial advantage over manually collecting term data from a natural language contract or even automatically extracting the data via natural language processing. Rather than needing to collect term data from a contract ex post, the terms of a computable contract are available for computational analysis throughout the contract's entire life. Computable contracts will also improve the collection of contract outcome data. Because computable contracts need real-world performance data to self-execute, this back-end outcome data can also be captured. For example, a computable contract for the delivery of widgets could collect performance data such as when the widgets are delivered, how many widgets are delivered, and the quality of the widgets delivered. Based on this information, the contract could determine whether the delivering party properly performed and, if so, how much the party should be paid under the contract. If the delivering party does not properly perform, the contract could flag that

---

[144] *See* Surden, *supra* note 108, at 658–59. Computable contracts can be used to collect outcome data regardless of whether the contract is completely self-executing or whether the contract merely produces a prima facie assessment that is then reviewed by the parties. *Id.* at 636. As a result, this Article does not take a position on whether computable contracts should be completely self-executing.

[145] *See supra* note 7 and accompanying text.

[146] *See* Surden, *supra* note 108, at 690–94 (noting that computable contracts can be used as inputs for other systems).

there is a dispute. Meanwhile, all of this outcome data could be collected and made available for analysis.

Although computable contracts are not yet available for widespread use, a critical technological innovation has substantially increased their feasibility: blockchain. Initially introduced in 2008, blockchain is best known as the technology that underlies digital currencies such as bitcoin.[147] A blockchain is a continuously growing public ledger of transactions (known as "blocks") supported by a distributed, peer-to-peer network that uses cryptography to ensure the validity of the blocks in the overall chain.[148] The key feature of a blockchain network is that it does not require a trust intermediary such as a bank or clearinghouse to validate transactions on the network.[149] Instead, the integrity of a blockchain network is maintained by what is commonly referred to as a "proof-of-work" system.[150] A proof-of-work system functions by having a distributed network of "miners"

---

[147] For the original Bitcoin paper that introduced blockchain technology, see Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System* BITCOIN.ORG, https://bitcoin.org/bitcoin.pdf [https://perma.cc/N55D-LTXA]. The paper was published under the pseudonym Satoshi Nakamoto. The identity of the person or persons who created Bitcoin and blockchain remains a mystery, though there are many theories. *See Satoshi Nakamoto*, WIKIPEDIA, https://en.wikipedia.org/wiki/Satoshi_Nakamoto [https://perma.cc/2CKU-K6WK].

[148] *See* Nakamoto, *supra* note 147; *see also* Catherine Martin Christopher, *The Bridging Model: Exploring the Roles of Trust and Enforcement in Banking, Bitcoin, and the Blockchain*, 17 NEV. L.J. 139, 152 (2016); Joshua Fairfield, *Smart Contracts, Bitcoin Bots, and Consumer Protection*, 71 WASH. & LEE L. REV. ONLINE 35, 36–37 (2014); Shahla Hazratjee, Note, *Bitcoin: The Trade of Digital Signatures*, 41 T. MARSHALL L. REV. 55, 58–60 (2015); Trevor Kiviat, Note, *Beyond Bitcoin: Issues in Regulating Blockchain Transactions*, 65 DUKE L.J. 569, 578–80 (2015); Marco Iansiti & Karim R. Lakhani, *The Truth About Blockchain*, HARV. BUS. REV., Feb. 2017; Emile Loza de Siles, *Blockchain: A Short Primer for Lawyers*, L.J. NEWSLETTERS, Mar. 2017, http://www.lawjournalnews letters.com/sites/lawjournalnewsletters/2017/03/01/blockchain-a-short-primer-for-lawyers/?slreturn=20180408202029[https://perma.cc/P446-KVT8].

[149] *See* Nakamoto, *supra* note 147; *see also* Christopher, *supra* note 148, at 17; Fairfield, *supra* note 148, at 40; Kiviat, *supra* note 148, at 574.

[150] *See* Nakamoto, *supra* note 147, at 3.

complete computationally difficult cryptographic tasks to verify transactions on the blockchain network.[151] Each verified transaction block is added to a chain of previously verified blocks and the "longest" chain (i.e. the sequence with the most verified transactions) is treated as the official chain.[152] Miners are rewarded for their work with units of value on the network, of which Bitcoin is an example.[153] To disrupt the validity of a blockchain network, a bad actor would need to amass a majority of computing power on the network, which could be incredibly expensive.[154] Such an actor would not have an incentive to do so, however, because trust in the network would deteriorate and the actor's units of value on the network would become worthless.[155] While public attention has largely focused on blockchain's cryptocurrency applications, many commentators have noted that blockchain's ability to verify transactions without needing a trust intermediary has the potential to enable computable contracts.[156]

Companies and organizations have begun to experiment with blockchain-enabled computable contracts, the most notable of which is Ethereum.[157] Ethereum has created a blockchain-enabled platform on top of which users can build

---

[151] *Id.*

[152] *Id.*

[153] *Id.* at 4.

[154] *Id.* at 3. This is commonly referred to as a fifty-one percent attack.

[155] *Id.* at 4.

[156] *See* Christopher, *supra* note 148, at 16; Karen E.C. Levy, *Book-Smart, Not Street-Smart: Blockchain-Based Smart Contracts and the Social Workings of Law*, 3 ENGAGING SCI., TECH., & SOC'Y 1, 1–3 (2017); Max Raskin, *The Law and Legality of Smart Contracts*, 1 GEO. L. TECH. REV. 305, 317–19 (2017); Lauren Henry Scholz, *Algorithmic Contracts*, 20 STAN. TECH. L. REV. 128, 145–46 (2017); Hazratjee, *supra* note 148, at 83; Kiviat, *supra* note 148, at 573, 603, 605–06.

[157] *See A Next Generation Smart Contract and Decentralized Application Platform*, GITHUB, https://github.com/ethereum/wiki/wiki/White-Paper [https://perma.cc/9YT8-KRPP]; *see also* Ethereum, https://www.ethereum.org [https://perma.cc/NM4H-DXRF].

their own computable contract applications.[158] The platform even includes a "contract-oriented" programming language for computable contracts, known as Solidity.[159] Using Solidity, computable contract developers can implement any number of innovative contract designs. As platforms such as Ethereum continue to grow and improve, computable contracts will likely begin to move into the mainstream of contracting, thereby creating new opportunities for collecting high-quality, structured contract data at scale. Future research will be needed to explore how to best design computable contracts to function as mechanisms for data collection.

D.   Beyond Automation

Predictive contracting differs from traditional contract automation both in its primary objective and in its technological foundation. Introduced in the 1970s,[160] contract automation technology has traditionally focused on reducing the cost and time associated with drafting a contract.[161] The classic example of this type of contract automation is LegalZoom, a web-based service that allows users to generate legal documents covering issues ranging from employment agreements to wills and trusts to basic intellectual property matters.[162] The target user is a non-lawyer who wants a fast and cheap way to generate a legal document without having

---

[158]   *See A Next Generation*, *supra* note 157, at 14 (describing Ethereum as "a blockchain with a built-in Turing-complete programming language, allowing anyone to write smart contracts and decentralized applications where they can create their own arbitrary rules for ownership, transaction formats and state transition functions").

[159]   *See* SOLIDITY, https://solidity.readthedocs.io/en/v0.4.23 [https://perma.cc/W3BJ-8JSZ].

[160]   *See* Kathryn D. Betts & Kyle R. Jaep, *The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life into a Decades-Old Promise*, 15 DUKE L. & TECH. REV. 216, 218 (2017).

[161]   *See supra* note 30.

[162]   *See* LEGALZOOM, https://www.legalzoom.com [https://perma.cc/4VNV-EH7N].

to hire a lawyer.[163] The documents involved tend to be highly standardized and require relatively little customization. There are also automation products targeted at lawyers that help them assemble routine documents in a short amount of time.[164] In both cases, the automation technology is intended to reduce the time and cost of producing a "good enough" contract. Predictive contracting, on the other hand, is focused on making contracts substantively better by providing contract drafters with statistical insights into the connections between contract terms and outcomes.

Predictive contracting also employs fundamentally different technology than traditional contract automation products. Most contract automation tools are built on top of pre-coded, rules-based logic systems.[165] These tools ask the user a variety of questions, and then, based on the user's responses and the system's internal logic (developed with the input of a subject matter expert such as a lawyer), present the user with additional questions until the tool has worked through the entire logic tree.[166] The tool then generates a document for the user from a set of pre-coded terms.[167] Unlike traditional automation tools, the internal logic of a predictive contracting system does not have to be explicitly defined ex ante.[168] Instead, predictive contracting relies on statistical machine learning models that develop their own logic over

---

[163] LegalZoom, however, does offer a premium service to connect users with a lawyer for more complex matters.

[164] Wilson Sonsini Goodrich & Rosati, a large Silicon Valley law firm, developed an internal automation tool that allows associates to generate a full set of "startup documents" for a new company, including a certificate of incorporation, founder stock purchase agreements, and company bylaws, by answering a set of pre-generated questions. *See WSGR Term Sheet Generator*, WILSON SONSINI GOODRICH & ROSATI, https://www.wsgr.com/ WSGR/Display.aspx?SectionName=practice/termsheet.htm [https://perma .cc/P4RP-BLZF].

[165] *See* Betts & Jaep, *supra* note 160, at 218–19.

[166] *Id.*

[167] *Id.*

[168] *See* Surden, *supra* note 35, at 93–95; Remus & Levy, *supra* note 44, at 9 (distinguishing between deductive instructions and data-driven instructions).

time based on a training data set.[169] Two primary implications arise from this technological difference. First, because a traditional automation tool derives its logic from a real-world expert, such as a lawyer, it will never be able to generate a better contract than the expert could have generated on her own (though the tool will often be faster and cheaper). Predictive contracting, on the other hand, generates its insights by analyzing large amounts of data in a way that a lawyer cannot. As a result, a predictive contracting system is complimentary to a lawyer's experience as opposed to merely replicative. Second, traditional automation tools are static whereas predictive contracting is dynamic. A traditional tool has to be pre-coded and therefore its logic cannot change over time unless it is recoded. On the other hand, a predictive contracting system can continuously update its logic as it is supplied with additional data from subsequent contracts, thereby improving its accuracy and generating better insights.

III. DISCUSSION

Predictive contracting has theoretical and practical implications. At the same time, predictive contracting faces a variety of limitations and risks. This Part examines these issues and proceeds as follows. Section III.A discusses the theoretical implications of predictive contracting, Section III.B discusses the practical implications, and Section III.C discusses the limitations and risks.

A. Theoretical Implications

Predictive contracting has multiple implications for the theory of contract design. The Sections below examine how predictive contracting can lead to (1) greater customization, (2) increased innovation, (3) more complete contract design, (4) more effective balancing of front-end and back-end costs, (5) better risk assessment and allocation, and (6) more accurate term pricing for negotiation.

---

[169] *See supra* Section II.A; *see also* Surden, *supra* note 35, at 93–95.

1. Customization

Contracts (both business-to-business and business-to-consumer) frequently display a high degree of standardization.[170] This can be seen in the widespread use of boilerplate provisions in commercial contracts.[171] These boilerplate terms often prove quite resistant to change and will sometimes remain in use despite adverse shocks.[172] This prevalence of standardization runs counter to the traditional efficient contracting view that predicts that parties to a contract will select the set of terms that maximizes the joint value generated by the contract.[173] Assuming that contracting scenarios display some degree of heterogeneity from one scenario to the next, contracts should reflect this

---

[170] *See generally* GULATI & SCOTT, *supra* note 18; Boardman, *supra* note 15; Choi & Gulati, *supra* note 17; Choi & Gulati, *supra* note 82; Choi, Gulati & Posner, *supra* note 17; Goetz & Scott, *supra* note 15; Hill, *supra* note 18; Kahan & Klausner, *supra* note 17; Kahan & Klausner, *supra* note 15; Klausner, *supra* note 15; Korobkin, *supra* note 18; Richman, *supra* note 18; Schwartz & Scott, *supra* note 9; Triantis, *supra* note 21.

[171] *See* GULATI & SCOTT, *supra* note 18, at 2–3 (describing the use of a "pari passu" clause in over ninety percent of cross-border sovereign bond contracts); Choi & Gulati, *supra* note 82, at 1130 (discussing the use of form contracts in business-to-consumer contracting); Hill, *supra* note 18, at 59, 63 (highlighting the use of forms in transactional law practice and describing how law firms use forms during the drafting process); Kahan & Klausner, *supra* note 15, at 718 (describing evidence of contractual boilerplate, such as bond indentures and corporate charters); Klausner, *supra* note 15, at 762 (discussing the use of form documentation in corporate contracting).

[172] *See* Choi & Gulati, *supra* note 17, at 934–35 (providing empirical evidence of a slow shift in standardized terms following an external shock); Choi, Gulati & Posner, *supra* note 17, at 3, 7–10 (proposing a model of boilerplate evolution in which an external shock disrupts pre-shock standardization causing a period of innovation that ultimately leads to post-shock standardization). The authors provide empirical evidence for this model of evolution from New York and English sovereign bond markets. *Id.* at 27, 35; *see also* GULATI & SCOTT, *supra* note 18, at 2–3 (discussing how a boilerplate "pari passu" clause used in cross-border sovereign bond contracts failed to be modified or discontinued even after an adverse judicial interpretation).

[173] *See supra* note 17 and accompanying text.

heterogeneity in their design. Yet this is typically not observed.

The contracts literature has posited numerous explanations for the prevalence of standardized terms in commercial contracts.[174] From these varied explanations, two common themes arise. First, standardized terms are faster and cheaper to use than nonstandard terms and therefore standardization reduces transaction costs.[175] Second,

[174] For a good overview of proposed reasons for contract standardization, see GULATI & SCOTT, *supra* note 18, at 34–43 (discussing reasons including learning externalities, network externalities, negative signaling, hindsight bias, satisficing, drafting routinization, herd behavior, collective action and free riding, endowment effects, and a lack of understanding of boilerplate terms). For a discussion of learning and network externalities, see Choi & Gulati, *supra* note 17, at 934–36 (providing empirical evidence of network externalities in the sovereign bond market); Kahan & Klausner, *supra* note 15, at 718–27, 742–60 (describing the effects of learning and network externalities and providing empirical evidence from bond covenants). For a discussion of cognitive biases, see Kahan & Klausner, *supra* note 17, at 359–64 (discussing status quo bias, anchoring bias, and conformity bias); Korobkin, *supra* note 18, at 1586–87 (describing how status quo and endowment bias lead to a higher prevalence of standardized terms). For a discussion of herd behavior, see GULATI & SCOTT, *supra* note 18, at 149; Kahan & Klausner, *supra* note 17, at 356–58 (noting that contract drafters are incentivized to use standardized, widely-used terms because if these terms fail, then the drafters are failing as a group rather than individually). For a discussion of agency costs, see GULATI & SCOTT, *supra* note 18, at 6 (arguing that inefficient standardization in the sovereign bond market arose because contract drafters were incentivized to promote volume-based, "cookie-cutter" transactions); Hill, *supra* note 18, at 77–78 (discussing how law firms frequently lack an incentive to produce a better form); Kahan & Klausner, *supra* note 17, at 353–55; Richman, *supra* note 18, at 79–82 (identifying that contract drafters are incentivized to use standardized terms that have previously been used and that law firms are incentivized to develop routines for the mass production of homogenous contracts).

[175] *See* Goetz & Scott, *supra* note 15, at 262–64, 290 (describing how implied contract terms reduce transaction costs by providing contracting parties with standardized "preformulations"); Klausner, *supra* note 15, at 782–84 (discussing how it is cheaper and faster for contract drafters to use standardized terms); Triantis, *supra* note 21, at 186–87 (noting that standardized terms reduce costs associated with contract drafting primarily because they can easily be redeployed).

standardized terms are more familiar to the contracting parties, lawyers, third parties, and courts and are therefore more predictable than nonstandard terms.[176]

Despite these apparent benefits, contracts can be inefficiently over-standardized. For example, assume that a company can choose between multiple different available versions of a particular contract term (Versions A, B, C, etc.). Over time, the company has settled on Version A as the standard version of the term. Version A is cheap and reliable. Assume that the cost to include Version A in the contract is effectively zero and the joint value generated by Version A in any contracting scenario is $10. In certain contracting scenarios, however, other versions of the term, while costlier to include, generate a net joint value of greater than $10. For example, in certain scenarios, Version B, which costs $10 to include, generates a joint value of $30 for a net joint value of $20, greater than the joint value generated by Version A. As a result, the standardized practice of always using Version A leads to a loss of total joint value across all contracting scenarios.

Predictive contracting can lead to a more efficient mix of standardization and customization in three ways. First, predictive contracting can reduce the cost and time associated with drafting and incorporating nonstandard terms. A predictive contracting system can supply contract drafters with a library of available contract terms. Instead of having to craft bespoke contract terms from scratch (a costly and time-

---

[176] *See* Boardman, *supra* note 15, at 1107 (proposing that contract drafters may be willing to accept a standardized term that has an inefficient judicial interpretation as long as that interpretation is fixed and therefore produces little to no uncertainty); Choi & Gulati, *supra* note 17, at 931 (highlighting that standardized terms reduce uncertainty); Goetz & Scott, *supra* note 15, at 263–64 (noting that courts have a preference for standardized implied terms and often disapprove of attempts to modify these terms with nonstandard express terms); Kahan & Klausner, *supra* note 17, at 353–55 (describing how contract drafters—who are often lawyers—are typically more risk averse than parties and therefore are more likely to use standardized terms than would otherwise be efficient due to the low uncertainty of these terms); Klausner, *supra* note 15, at 776–79 (discussing the judicial interpretative benefits of using standardized terms).

consuming process), contract drafters can select from different versions of terms available in the predictive contracting system and customize them as necessary. For example, Sirion Labs provides contract drafters with a contract creation tool that uses historical contract data.[177] The tool contains a searchable, filterable library of all of the user's past contract terms.[178] For a given term, the tool displays different versions of that term that the user has used before.[179] A contract drafter can build a draft of a contract by dragging and dropping terms from the term library into the draft, modifying them as necessary and supplying condition data such as party identities and dates.[180] Second, predictive contracting can make nonstandard terms more predictable, thereby reducing the uncertainty associated with these terms. A predictive contracting system can supply contract drafters with information on the statistical connections between contract terms and outcomes, including the likelihood of litigation and potential adverse judicial interpretation. Equipped with this information, contract drafters can make better-informed decisions about the inclusion of nonstandard terms.[181] Third, predictive contracting can facilitate the use of highly-tailored, context-specific contract terms. A predictive contracting model can take into account exogenous conditions, such as party and transaction characteristics, and provide the contract drafter with insights into how these external factors

---

[177] Telephone Interview with Sirion Labs Representative, *supra* note 96.

[178] *Id.* For example, a user can filter the term library based on characteristics such as the type of contract, the name of the counterparty, the counterparty's industry, and the jurisdiction of the contract. *Id.*

[179] *Id.* Contract drafters can label these terms for negotiation purposes as either "preferred," "fall-back," or "walk-away." *Id.*

[180] *Id.*

[181] This requires the predictive contracting system to have been trained on contracts containing nonstandard terms. For some nonstandard terms, it may be the case that they have been used so infrequently that there are insufficient data to train the model. From a prediction perspective, nonstandard terms such as these are similar to novel terms that have never been used. For a discussion of how predictive contracting can assist with the generation of novel terms, see *infra* Section III.A.2.

affect the connections between terms and outcomes. Contract drafters can use this information to select the best set of terms given the exogenous conditions present in the contracting scenario.

Returning to the hypothetical example discussed above, a predictive contracting system could enable the company to identify the contracting scenarios in which alternative versions of the term in question generate greater joint value than Version A. As a result, the company would be able to select the value-maximizing version of the term in each contracting scenario. In addition, an easily-accessible library of terms containing examples of alternative versions would likely reduce the cost associated with using these alternative versions. As this example demonstrates, predictive contracting can lead to a more efficient selection of contract terms given a set of exogenous conditions.

2. Innovation

Closely related to the issue of standardization versus customization is the issue of contract innovation. Whereas customization is concerned with the selection of efficient contract terms from a set of available terms and the tailoring of those terms to specific contracting scenarios, innovation is concerned with the generation of new terms.[182] Contract innovation is critically important to contract design because it is the source of novel terms.[183] Contract innovation expands the option set of available terms from which contract drafters can select when designing a contract. Yet despite the importance of contract innovation, contract drafters rarely innovate.[184]

---

[182] *See* Triantis, *supra* note 21, at 192 (discussing the difference between customization and innovation and highlighting the scalability of innovative terms).

[183] *See* GULATI & SCOTT, *supra* note 18, at 164 (stressing the importance of contract innovation).

[184] *See generally* GULATI & SCOTT, *supra* note 18; Boardman, *supra* note 15; Choi & Gulati, *supra* note 17; Goetz & Scott, *supra* note 15; Kahan & Klausner, *supra* note 17; Kahan & Klausner, *supra* note 15; Triantis, *supra* note 21.

The contracts literature suggests two primary reasons for the inefficiently low level of contract innovation observed in the market.[185] First, innovative terms have not been previously assessed by the contracting parties, lawyers, third parties, and courts, and are therefore more uncertain than existing terms.[186] Second, parties typically have a difficult time identifying the effects of specific terms, and therefore, contract drafters (who are primarily evaluated on whether a contract performs poorly) have little incentive to take on the cost and risk of innovating if success from their innovation is unlikely to be recognized.[187] For example, assume a contract drafter is faced with a decision between using an existing term or a new term. The existing term is familiar and has a 100% chance of generating $10 of value for the drafter's client. The new term, on the other hand, has a 50% chance of generating $50 and a 50% chance of losing $10, for an expected value of $20.[188] Assuming the client is risk neutral,[189] it would prefer

[185] A third, related reason for the lack of innovation flows from the two primary reasons: organizational impediments faced by contract drafters. *See* GULATI & SCOTT, *supra* note 18, at 145–49, 161 (describing barriers to innovation within law firms, including examples of contract drafters being reprimanded for attempting to innovate); Smith & King, *supra* note 82, at 31 (noting how contracts can become intertwined with other organizational processes, which leads to innovation inertia); Triantis, *supra* note 21, at 186 (discussing structural impediments that prevent contract drafters from having an incentive to innovate).

[186] *See* Goetz & Scott, *supra* note 15, at 263 (discussing how innovative terms are more likely to be misinterpreted by courts).

[187] *See* Gardner, *supra* note 22, at 43–44 (providing examples of the difficulty parties face in determining the value of services provided by contract drafters); Goetz & Scott, *supra* note 15, at 291 (noting that contract drafters considering innovating must incur the cost of identifying terms that are superior to existing terms); Kahan & Klausner, *supra* note 17, at 353–55 (discussing how contract drafters are frequently judged based on whether a contract fails rather than whether they identify the optimal set of terms and that the success of a new term may not be attributed to the contract drafter's innovation); Triantis, *supra* note 21, at 180, 194 (arguing that contract innovation is stymied by the inability of parties to evaluate the effects of specific contract terms).

[188] Expected value = (0.5 * $50) + (0.5 * -$10) = $20.

[189] An entity is risk neutral if it prefers the option with the greatest expected value, regardless of risk.

that the contract drafter use the new term. Assume, however, that the client cannot observe the magnitude of the result, only whether the result is positive or negative.[190] In this example, the contract drafter will always choose the existing term, which has a 100% chance of generating a positive result, even though the new term is better for the client. This principal-agent problem can lead contract drafters to select terms that serve their own interests, but not necessarily those of their clients.

Predictive contracting can lead to a greater level of contract innovation in three ways. First, predictive contracting can track the effects of innovative terms by identifying ex-post connections between these terms and various contract outcomes. If contract drafters have an effective means of demonstrating to parties the effects of their innovations, they will have a much stronger incentive to innovate in the first place. In the above example, the contract drafter could use a predictive contracting system to show the client that the new term generates a greater expected value than the existing term over a sufficiently large number of contracting instances. In this example, the predictive contracting system aligns the incentives of the client and the contract drafter, thereby helping to mitigate the principal-agent problem.

Second, unlike traditional automation technologies, predictive contracting can assist with the generation of new terms.[191] Natural language processing ("NLP") can break contract terms down into their constituent conceptual subparts, known as ontologies.[192] For example, Legal Robot is

---

[190] This is a simplified representation of the fact that parties typically evaluate a contract drafter based on whether the contract did "poorly" or not.

[191] *See* Betts & Jaep, *supra* note 160, at 227 (noting that historically, contract automation tools have been unable to produce novel contract language).

[192] Telephone Interview with Legal Robot Representative, *supra* note 120; *see* Dominique Estival, Chris Nowak & Andrew Zschorn, *Towards Ontology-Based Natural Language Processing*, *in* PROCEEDINGS OF THE WORKSHOP ON NLP AND XML 59–66 (2004).

developing NLP technology that deconstructs contract terms into the fundamental building blocks of contract language.[193] A predictive contracting system equipped with this technology could identify connections between these contract subparts and various contract outcomes. Contract drafters could use this information during the contract innovation process to have an understanding ex ante of the potential effects of a new term. Figure 2 shows an example of this process.

Figure 2: Innovation Example
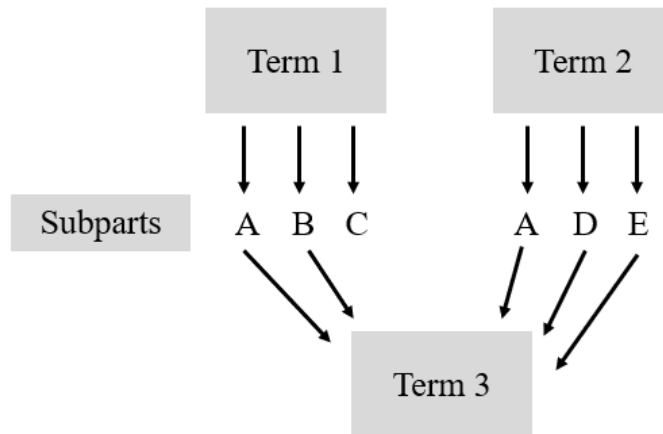
As can be seen in Figure 2, a predictive contracting system equipped with ontological NLP technology can deconstruct two terms, Term 1 and Term 2, into their constituent subparts. Term 1 is comprised of Subparts A, B, and C and Term 2 is comprised of Subparts A, D, and E. A contract drafter is considering using a new term, Term 3, that is comprised of Subparts A, B, D, and E. Based on prior use of Terms 1 and 2, the predictive contracting system can provide the contract drafter with information on connections between the subparts that make up Term 3 and various contract

---

[193] Telephone Interview with Legal Robot Representative, *supra* note 120.

outcomes. This information can help the contract drafter understand the likely effects of Term 3 on relevant outcomes. For example, the drafter may be particularly concerned with the likelihood of Term 3 resulting in litigation and can therefore examine the connections between Subparts A, B, D, and E and litigation outcomes. The predictive contracting system enables the contract drafter to predict potential effects of an innovative term.

Third, predictive contracting can enable contractual experimentation. If a company engages in a sufficiently large volume of contracting, it can use predictive contracting to "A/B test" a new contract term or set of terms. For example, assume a company currently uses Version A of a term but is interested in potentially using Version B. Using predictive contracting technology, the company could run a controlled experiment in which it randomly assigns Version A to one set of contracts and Version B to another set. The company can then track the effects of Versions A and B on outcomes of interest.

3. Completeness

Contracts are frequently described as a set of conditional directions that specify the obligations of the parties with respect to one another in a variety of contingent future states.[194] Based on this view, the ideal contract is a "complete contingent" contract, one that contains directions for every possible future state of the world.[195] Despite the appeal of contractual completeness, real-world contracts are incomplete because they do not account for all potential contingencies.[196]

---

[194] *See* Anderlini & Felli, *supra* note 18, at 4 (modeling contracts as algorithmic maps that connect contingent states with the actions to be taken if those states occur); Goetz & Scott, *supra* note 15, at 264 (noting that the key feature of a contract is the ability to specify directions for future contingencies); Schwartz & Scott, *supra* note 9, at 557 (highlighting the "intertemporal" nature of contracts); Scott & Triantis, *supra* note 16, at 188 (defining a contract as a legally binding promise to act in the future).

[195] *See* Goetz & Scott, *supra* note 15, at 267; Williamson, *supra* note 25, at 236.

[196] *See generally* Anderlini & Felli, *supra* note 18; Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of*

While the contracts literature provides multiple explanations for contractual incompleteness,[197] by far the most common explanation is that there are too many potential contingencies for contract drafters to feasibly design contracts that account for all future states.[198] As a result, contracts are often incomplete, especially with respect to low-likelihood contingencies.[199]

Default Rules, 99 YALE L.J. 87 (1989); Choi & Gulati, *supra* note 82; Claire A. Hill, *Bargaining in the Shadow of the Lawsuit: A Social Norms Theory of Incomplete Contracts*, 34 DEL. J. CORP. L. 191 (2009); Choi & Triantis, *supra* note 81; Katz, *supra* note 18; Schwartz & Scott, *supra* note 9; Scott & Triantis, *supra* note 19; Scott & Triantis, *supra* note 16.

[197] For a good literature summary of contractual incompleteness, see Scott & Triantis, *supra* note 19, at 816; *see also* Ayres & Gertner, *supra* note 196, at 94 (describing how a party may strategically withhold information that would make a contract more complete so as to increase its percentage share of the value generated by the contract); Choi & Triantis, *supra* note 81, at 859–60 (arguing that contractual incompleteness "can facilitate the provision of efficient incentives and the signaling of private information at the time of contracting and of renegotiation"); Hill, *supra* note 196, at 208–12 (discussing how contractual incompleteness increases the uncertainty and potential cost of litigation and can therefore serve as an ex ante bonding mechanism to deter ex post litigation); Katz, *supra* note 18, at 172–74.

[198] *See* Anderlini & Felli, *supra* note 18, at 8 (defining incomplete contracts as "contracts that show *evidence* that the contracting parties were *constrained* in their *ability to distinguish between states* when the contract was drawn up") (emphasis in original); Ayres & Gertner, *supra* note 196, at 92–94; Choi & Gulati, *supra* note 82, at 1159 (describing how parties rationally choose to neither contract for every contingency nor clarify every potential meaning of a term and therefore accept contractual incompleteness); Schwartz & Scott, *supra* note 9, at 594–95; Scott & Triantis, *supra* note 16, at 189–90 (arguing that complete contingent contracts are impossible given the costs associated with planning for all potential future states); Smith & King, *supra* note 82, at 7, 17.

[199] *See* Schwartz & Scott, *supra* note 9, at 559 (noting that parties will often choose not to bear the costs associated with contracting for low likelihood contingencies); Choi & Gulati, *supra* note 82, at 1155–56 (citing Lisa Bernstein, *The Questionable Empirical Basis of Article 2's Incorporation Strategy: A Preliminary Study*, 66 U. CHI. L. REV. 710, 747 (1999) (discussing how incompleteness in industry customs is often associated with low probability contingencies that trade associations choose not to expend resources on)).

Predictive contracting can increase contractual completeness by providing contract drafters with information on the set of potential future states and their associated likelihoods. For example, assume a contract drafter is designing a term for a contract and is deciding what potential contingencies to plan for. From the drafter's personal experience and that of her colleagues, the drafter knows that Contingencies A and B are by far the most common and that Contingency C occasionally happens as well. Traditionally, the drafter would likely design the term to account for Contingencies A, B, and C. Assume instead, however, that the drafter has access to a predictive contracting system trained on a dataset of thousands of similar contracts. The system shows the drafter that Contingency A occurs fifty percent of the time, Contingency B occurs forty percent of the time, Contingency C occurs five percent of the time, and Contingencies D through H each occur approximately one percent of the time. The drafter can then use this information to design a more complete term. While complete contingent contracts are still infeasible,[200] predictive contracting can increase contractual completeness relative to traditional contracting.

4. Cost Balancing

The costs associated with a contract can be divided between the front-end and the back-end of the contract's life.[201] Front-end costs include any costs to negotiate the contract, conduct due diligence, design and draft the contract, and execute the contract.[202] Back-end costs include any costs

---

[200] *But see* Anthony J. Casey & Anthony Niblett, *Self-Driving Contracts*, 43 J. CORP. L. 1, 1, 12–15 (2017) (suggesting that advances in predictive analytics will eventually lead to contracts that are perfectly complete because they will use technology and data to convert ex ante objectives into ex post directives for the parties given any set of contingencies).

[201] *See supra* note 81 and accompanying text.

[202] *See* Choi & Triantis, *supra* note 81, at 851; Posner, *supra* note 81, at 1583–84; Scott & Triantis, *supra* note 19, at 814, 817, 822–24; Scott & Triantis, *supra* note 16, at 190–91.

to perform obligations under the contract, monitor performance, and potentially renegotiate, arbitrate, or litigate the contract.[203] Between the front-end and back-end of a contract's life, uncertainty is resolved regarding which contingent state of the world will occur.[204] According to the contracts literature, parties should (and do) balance front-end and back-end costs.[205] The general view is that investing in front-end costs reduces back-end costs.[206] For example, spending additional resources during the front-end to make a contract more complete and precise reduces the likelihood that the contract will result in litigation, and if it does, reduces the likelihood that a judge will misinterpret the contract.[207]

In order for contract drafters to efficiently balance front-end and back-end costs, they need to understand the connections between these costs. Without this information,

[203] *See* Choi & Triantis, *supra* note 81, at 851; Posner, *supra* note 81, at 1583–84; Scott & Triantis, *supra* note 19, at 814, 817, 822–24; Scott & Triantis, *supra* note 16, at 190–91.

[204] *See* Scott & Triantis, *supra* note 19, at 823 (drawing the distinction between the front-end and back-end of a contract's life as the resolution of uncertainty).

[205] *See* Posner, *supra* note 81, at 1583–84; Scott & Triantis, *supra* note 19, at 814, 817, 836 (proposing the use of indifference curves to model the efficient tradeoff between front-end and back-end costs); Scott & Triantis, *supra* note 16, at 196–98.

[206] *See* Posner, *supra* note 81, at 1608; Scott & Triantis, *supra* note 19, at 814, 817, 822–24, 836; Scott & Triantis, *supra* note 16, at 196–98. *But see* Hill, *supra* note 196, at 208–12 (arguing that less specific contract terms, which lead to lower front-end costs, can serve as a bonding mechanism that deters parties from engaging in litigation, thereby reducing back-end costs as well); Choi & Triantis, *supra* note 81, at 852–55, 859–60, 883, 885, 921–22 (arguing that vague terms with lower front-end costs can incentivize an efficient provision of information ex ante to deter adverse selection, thereby reducing back-end litigation costs). Choi & Triantis provide a series of numerical examples for their proposed theory using the context of a material adverse change clause in an acquisition agreement. *See* Choi & Triantis, *supra* note 81, at 896–920.

[207] The tradeoff between vague (less costly) and precise (costlier) contract terms during the front-end and the effects of contractual preciseness on back-end litigation costs has been compared to the tradeoff between rules and standards from the literature on legislation and regulation. *See* Scott & Triantis, *supra* note 19, at 820.

parties may engage in inefficient balancing. For example, assume that a contract drafter is considering how much time and effort to invest in the design of a term during the front-end of a contract's life. The drafter can either design a vague term that is less expensive (Term 1) or a precise term that is more expensive (Term 2). Assume Term 2 is $10 more expensive in the front-end than Term 1. The drafter should therefore select Term 2 if Term 2 will save more than $10 in expected value on the back-end relative to Term 1. Assume the relevant back-end cost in this scenario is potential litigation and that litigation carries a cost of $100. The drafter should therefore use Term 2 if Term 2 reduces the likelihood of litigation relative to Term 1 by more than 10%.[208] To make this determination, the drafter needs information on the front-end cost of Term 2 relative to Term 1, the cost of litigation, and the reduction in the likelihood of litigation associated with Term 2 relative to Term 1. If the drafter mistakenly believes that Term 2 reduces the likelihood of litigation by 15% but Term 2 actually only reduces the likelihood by 5%, the drafter would inefficiently invest in Term 2 during the front-end, thereby increasing overall contracting costs.

Predictive contracting can lead to more efficient cost balancing by providing contract drafters with key information regarding front-end and back-end costs. A predictive contracting system can track both front-end and back-end costs associated with a particular type of contract. The system can then use these data (along with data on relevant exogenous conditions) to identify connections between front-

---

[208] The drafter should use Term 2 if the benefit of Term 2 relative to Term 1 is greater than the cost of Term 2 relative to Term 1, which is $10. The benefit of Term 2 relative to Term 1 can be represented as the reduction in the likelihood of litigation multiplied by the cost of litigation, which is $100. The balancing equation is therefore: Cost of Term 2 = $10 < Benefit of Term 2 = Reduction in Litigation Likelihood * $100. The benefit of Term 2 is therefore greater than the cost of Term 2 if Term 2 reduces the likelihood of litigation by more than ten percent. This is a contractual application of the famous formula for determining negligence set forth by Judge Learned Hand in United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947).

end and back-end costs. In the previous example, a predictive contracting system could provide the contract drafter with the expected front-end costs of Terms 1 and 2 and the range of expected litigation outcomes (including likelihoods and costs) associated with these terms. The contract drafter would therefore be able to make a better-informed (and likely more efficient) cost balancing decision. In addition, as discussed previously, predictive contracting systems can reduce the cost of designing and drafting customized, context specific terms.[209] This enables contract drafters to include precise terms in the front-end at lower cost, thereby increasing the number of situations in which front-end investment can reduce overall contracting costs.

5. Risk Assessment and Allocation

Most traditional contracting occurs under uncertainty.[210] At the time of designing a contract, the parties and their contract drafters typically do not know which contingent state of the world will occur, nor do they know the full set of possible contingent states or the probability distribution of these states. Contract drafters, like most humans, are bad at making decisions under uncertainty, especially when dealing with very low-probability contingencies.[211] Drafters are much

---

[209] *See supra* Section II.A.1.

[210] *See* Boardman, *supra* note 15, at 1107; Choi & Gulati, *supra* note 17, at 931; Choi & Triantis, *supra* note 81, at 882; Hill, *supra* note 196, at 208–12; Korobkin, *supra* note 18, at 1622; Scott & Triantis, *supra* note 19, at 823; Scott & Triantis, *supra* note 16 (describing the difficulty contract drafters face in attempting to achieve both ex ante and ex post efficiency under uncertainty); Smith & King, *supra* note 82, at 7; Williamson, *supra* note 18, at 555 (listing uncertainty as one of the three key features of contracts along with frequency and asset specificity); Williamson, *supra* note 25, at 259 (discussing the importance of transaction-specific contract design under uncertainty).

[211] *See* Hill, *supra* note 18, at 73; Christine Jolls, Cass R. Sunstein & Richard Thaler, *A Behavioral Approach to Law and Economics*, *in* BEHAVIORAL LAW AND ECONOMICS 13, 38 (Cass R. Sunstein ed., 2000) (discussing the difficulty humans face in estimating low-probability outcomes); Cass R. Sunstein, *Introduction*, *in* BEHAVIORAL LAW AND ECONOMICS 1, 4–5 (Cass Sunstein ed., 2000). Much of the literature on law

better equipped, however, to make decisions under risk.[212] The distinction between risk and uncertainty is that unlike uncertainty, risk involves a known set of possible outcomes and the probability distribution of those outcomes.[213] A coin flip is the classic example of risk: the coin flipper knows that the two possible outcomes are heads or tails and that the likelihood of each is fifty percent. Most people would much rather make a decision using a coin flip as opposed to a method with unknown outcomes and probabilities. While real-world situations are generally far more complex and often involve a mix of uncertainty and risk, the core observation is that humans are better equipped to make decisions when they have information on possible outcomes and their likelihoods. Accurately assessing risk is especially important in the context of contracting because risk allocation is a key feature of many contracts.[214]

and behavioral economics builds on the work of Amos Tversky and Daniel Kahneman. *See generally* Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979); Amos Tversky & Daniel Kahneman, *Advances in Prospect Theory: Cumulative Representation of Uncertainty*, 5 J. RISK & UNCERTAINTY 297 (1992); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974); Amos Tversky & Daniel Kahneman, *Rational Choice and the Framing of Decisions*, 59 J. BUS. S251 (1986).

[212] *See* Hill, *supra* note 18, at 74–75.

[213] *See* FRANK KNIGHT, RISK, UNCERTAINTY, AND PROFIT, 198–99 (1921) (proposing the original distinction between risk and uncertainty); *see also* Craig R. Fox & Amos Tversky, *Ambiguity Aversion and Comparative Ignorance*, 110 Q.J. ECON. 585, 585 (1995); Hill, *supra* note 18, at 74–75.

[214] *See* Afra Afsharipour, *Transforming the Allocation of Deal Risk Through Reverse Termination Fees*, 63 VAND. L. REV. 1161, 1163–68 (2010) (discussing the role of reverse termination fees in the allocation of risk in corporate acquisitions); Choi & Triantis, *supra* note 81, at 851 (arguing that risk allocation is a significant component of contract design that enables efficient decisions regarding investment, contracting, and trade); Robert T. Miller, *The Economics of Deal Risk: Allocating Risk Through MAC Clauses in Business Combination Agreements*, 50 WM. & MARY L. REV. 2007, 2007–09 (2009) (presenting the results of an empirical study of risk allocation via material adverse change ("MAC") clauses in corporate acquisitions, finding that MAC clauses typically allocate four types of risk: systematic risks, indicator risks, agreement risks, and business risks).

Companies are currently using contract data to facilitate more effective contract risk assessment and allocation. For example, Microsoft tracks contract risks identified during the contract design stage to determine how frequently these risks occur. The company then uses this information to inform the design of future contracts.[215] When Microsoft negotiates a contract, the contract team puts together a risk profile of the expected risks associated with the contract.[216] The company uses this risk profile when determining whether to enter into the contract and whether to negotiate specific nonstandard terms.[217] After the contract is signed, Microsoft tracks which of the identified risks (if any) occur during the performance of the contract as well as any unidentified risks.[218] The company then uses these data to update its standard terms and improve its risk assessment process.[219] For example, if Microsoft always negotiates against the inclusion of a particular type of penalty provision, but then the data show that the penalty in question never occurs, the company can consider altering its assessment of the riskiness of such a provision.[220] Microsoft's system for tracking contract risk data allows it to make better-informed risk assessment and allocation decisions.

Predictive contracting can enable parties to convert contract uncertainty into contract risk. By tracking contract outcomes, a predictive contracting system can provide contract drafters with information on the set of potential contingent states and their respective likelihoods, thereby quantifying uncertainty. Equipped with distributional information on contract outcomes, contract drafters can more effectively assess contract risk. In addition, a predictive contracting system can identify connections between contract design and contract risks. This allows contract drafters to better evaluate the risk profile of a contract given its terms

---

[215] Telephone Interview with Microsoft Representatives, *supra* note 87.

[216] *Id.*

[217] *Id.*

[218] *Id.*

[219] *Id.*

[220] *Id.*

and relevant exogenous conditions. Predictive contracting also creates opportunities for new term designs that use distributional data on contract outcomes. Contract drafters can use outcome probability distributions to craft terms that have contingent effects based on the relation of a given outcome to the expected probability distribution of that outcome. These terms could enable contract drafters to more effectively allocate contract risk.

6. Negotiation and Term Pricing

According to the efficient contracting literature, the goal of contracting parties is to maximize the joint value created by the contract.[221] Parties select the set of contract terms that maximize the joint value and then divide up the value via the price term based on their relative bargaining power.[222] Under this view, parties optimize for the size of the pie, not how the pie is sliced. This view relies on the unrealistic assumption, however, that parties contract under perfect conditions, including full information and no transaction costs.[223] In the presence of contracting imperfections, such as asymmetric information, parties are often unable to achieve the value-maximizing contract design.[224] Instead of focusing on joint

---

[221] *See supra* note 16 and accompanying text.

[222] This view has been described as an "irrelevance" theory of bargaining power because according to this theory, the relative bargaining power of the parties does not affect the design of the contract, but rather only affects the price. *See* Williams, *supra* note 2, at 106–07; Choi & Triantis, *supra* note 2, at 1670.

[223] *See* Williams, *supra* note 2, at 110.

[224] *See* OLIVER HART, FIRMS, CONTRACTS AND FINANCIAL STRUCTURE 32–33 (1995); LAX & SEBENIUS, *supra* note 222, at 38–40, 245–46; ROBERT H. MNOOKIN, SCOTT R. PEPPET & ANDREW S. TULUMELLO, BEYOND WINNING: NEGOTIATING TO CREATE VALUE IN DEALS AND DISPUTES 9 (2000); Ian Ayres & Robert Gertner, *Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules*, 101 YALE L.J. 729, 736–37, 742 (1992); Choi & Triantis, supra note 2, at 1687–89; Oliver Hart & John Moore, *Property Rights and the Nature of the Firm*, 98 J. POL. ECON. 1119, 1132 (1990); Jason Scott Johnston, *Strategic Bargaining and the Economic Theory of Contract Default Rules*, 100 YALE L.J. 615, 615–16, 636–37 (1990); Korobkin, *supra* note 18, at 1206 (arguing that costly information leads parties to choose

value maximization, parties typically use their bargaining power to jockey for contract terms that are allocatively advantageous but not necessarily value-maximizing.[225] In their efforts to secure a larger piece of the pie, parties often end up with a smaller pie.

Predictive contracting can improve contract design in the presence of contracting imperfections by providing parties with better information on the value of specific contract terms. This information enables parties to more accurately price contract terms during the negotiation process.[226] For example, assume two parties are negotiating which version of a term to include in a contract. Party A prefers Version 1 and Party B prefers Version 2. Under traditional contracting, in the presence of incomplete information, the parties would likely use their relative bargaining power to determine which version of the term to include. Assume, however, that the parties have access to a predictive contracting system that can

---

inefficient, allocatively advantageous terms as opposed to efficient terms); Williams, *supra* note 2, at 113–15.

[225] *See* Albert Choi & George Triantis, *Market Conditions and Contract Design: Variations in Debt Contracting*, 88 N.Y.U. L. REV. 51, 53–55 (2013) (discussing how lending covenants were primarily borrower-friendly from 2000–2007 and then shifted to be lender-friendly after the financial crisis); Choi & Triantis, *supra* note 2, at 1693 (discussing how material adverse change definitions in merger agreements shift between seller-friendly and buyer-friendly forms); de Fontenay, *supra* note 112, at 395–98, 405–13; Williams, *supra* note 2, at 151–58 (finding a statistically significant connection between bargaining power and a variety of venture capital financing terms in venture capital contracts from 2004–2015); Joseph W. Bartlett, *Sea Change*, VC EXPERTS, https://www.vcexperts.com/ reference/buzz/63 [https://perma.cc/Q539-7V2P] (discussing how venture financing terms became much more investor-friendly following the dot com crash of 2000–2001).

[226] Large law firms currently use privately compiled contract data to help their clients gain a bargaining advantage via more accurate term pricing. *See* de Fontenay, *supra* note 112, at 396–98. Law firms only use these data to engage in relative term pricing, however, as opposed to absolute pricing. *Id.* at 425–26 ("[T]ransactional lawyers are likely to be more accurate at ranking terms against one another ('Term A should be worth more to you than Term B') than at ascribing a specific value to each term ('Term A should be worth *x* dollars to you')."). Predictive contracting can enable term pricing that is far closer to absolute pricing.

identify the expected values of the terms for each party. For example, assume Version 1 has an expected value of $20 for Party A and $10 for Party B and Version 2 has an expected value of $10 for Party A and $30 for Party B. Using this information, the parties can see that the joint expected value of Version 2 ($40) is $10 greater than the joint expected value of Version 1 ($30). Even if Party A has greater relative bargaining power, Party A is still likely to agree to use Version 2 in exchange for a more favorable price term.[227] Predictive contracting moves real world contracting closer to the full information assumption of efficient contracting theory, thereby increasing the likelihood that parties select the set of contract terms that maximize the joint value of the contract.[228]

B.   Practical Implications

Predictive contracting has practical implications for the contracting ecosystem, particularly for the transactional lawyers (both at law firms and in-house) who frequently serve as the primary drafters of business contracts. Technological innovation in the legal industry has frequently been characterized as an existential threat to lawyers.[229] This is a familiar narrative in a world in which jobs long done by humans are being automated at a rapid pace.[230] Yet, despite proposing to automate many tasks traditionally performed by transactional lawyers, predictive contracting will not make transactional lawyers obsolete. To the contrary, predictive

---

[227]   This example is a contractual application of the Coase Theorem. *See supra* note 17.

[228]   This assumes that both parties have access to predictive contracting systems. For a discussion of the risks that arise when only one party has access to predictive contracting, see *infra* Section III.C.4.

[229]   *See* RICHARD SUSSKIND, THE END OF LAWYERS? RETHINKING THE NATURE OF LEGAL SERVICES 99–146 (2010); RICHARD SUSSKIND, TOMORROW'S LAWYERS: AN INTRODUCTION TO YOUR FUTURE 43–58 (2d ed. 2017).

[230]   *See generally* Natalie Kitroeff, *Robots Could Replace 1.7 Million American Truckers in the Next Decade*, L.A. TIMES (Sep. 25, 2016), http://www.latimes.com/projects/la-fi-automated-trucks-labor-20160924 [https://perma.cc/WLL5-NH9R].

contracting has the potential to make transactional lawyers more valuable than ever before by providing them with access to previously unavailable information on the statistical connections between contract terms and outcomes. This newly available information will likely significantly change the role of transactional lawyers.

In his pivotal article in 1984, Ronald Gilson characterized transactional lawyers as "transaction cost engineers."[231] According to Gilson, transactional lawyers add value to a transaction by reducing the costs associated with that transaction by more than the fee they charge.[232] Lawyers accomplish this goal through the use of contract mechanisms, such as representations, warranties, and indemnification provisions.[233] Since Gilson's pioneering work, legal commentators have proposed additional roles for transactional lawyers including reputational intermediaries,[234] regulatory compliance experts,[235] and enterprise architects.[236] More recently, Elizabeth de Fontenay has noted that transactional lawyers at large law firms add value by collecting data on private deal terms that they use to provide market insights to their clients.[237]

Predictive contracting will enable two additional roles for transactional lawyers to complement those discussed above.

---

[231] Ronald J. Gilson, *Value Creation by Business Lawyers: Legal Skills and Asset Pricing*, 94 YALE L.J. 239, 255 (1984); *see also* Lisa Bernstein, *The Silicon Valley Lawyer as Transaction Cost Engineer?*, 74 OR. L. REV. 239, 251–52 (1995); Ronald J. Gilson & Robert H. Mnookin, *Foreword: Business Lawyers and Value Creation for Clients*, 74 OR. L. REV. 1, 2 (1995).

[232] *See* Gilson, *supra* note 231, at 255.

[233] *See generally id.* at 256–93.

[234] *See* Reinier H. Kraakman, *Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy*, 2 J.L., ECON., & ORG. 53, 94 (1986); Karl S. Okamoto, *Reputation and the Value of Lawyers*, 74 OR. L. REV. 15, 18–19 (1995); Larry E. Ribstein, *Ethical Rules, Agency Costs, and Law Firm Structure*, 84 VA. L. REV. 1707, 1739–40 (1998).

[235] *See* Steven L. Schwarcz, *Explaining the Value of Transactional Lawyering*, 12 STAN. J.L., BUS. & FIN. 486, 500 (2007).

[236] *See* George W. Dent, Jr., *Business Lawyers as Enterprise Architects*, 64 BUS. LAW. 279, 317 (2009).

[237] *See* de Fontenay, *supra* note 112, at 395–98.

The first new role is outcome engineer. Similar to Gilson's transaction cost engineers who use contractual mechanisms to reduce transaction costs, outcome engineers will use predictive contracting to help clients achieve their desired outcomes. Outcome engineers will accomplish this by helping their clients understand the likely effects of contract terms on outcomes as identified by a predictive contracting system. In addition, outcome engineers will assist their clients in weighing tradeoffs between expected outcomes. For example, an outcome engineer could use a predictive contracting system to identify that Term A typically results in a lower quality of counterparty performance but also has a lower likelihood of causing a dispute, whereas Term B results in higher performance quality but causes more disputes. The outcome engineer could use this information to help her client balance performance quality against dispute likelihood. After discussing the potential effects of terms on outcomes and the tradeoffs between expected outcomes, an outcome engineer can design a contract that is tailored to achieve her client's desired set of outcomes.

The second new role for transactional lawyers enabled by predictive contracting is contract innovator. As was discussed in Section II.A.2, predictive contracting will promote contract innovation by helping contract drafters understand the potential effects of new terms on outcomes. These new terms, however, must still be created by humans. Transactional lawyers will therefore have a pivotal role to play in creating innovative terms that can then be used and analyzed by a predictive contracting system. The generation of new terms will begin with a creative contract innovator who designs an initial version of a new term. During the design process, the innovator can use a predictive contracting system to gain insights into the potential effects of the new term based on ontological similarities with existing terms.[238] Once the new term is put into use in contracts, the innovator can track the term's performance using predictive contracting and modify

---

[238]   *See supra* Section III.A.2.

subsequent versions of the term based on real-world outcome data.

Along with creating new roles, predictive contracting will also alter the set of skills that are important for the practice of transactional law. One of the hallmarks of a valuable transactional lawyer has traditionally been a substantial amount of experience and the anecdotal deal knowledge that comes with that experience. Business lawyers have long advised their clients based on their prior experiences with similar transactions. Yet, in the presence of a predictive contracting system trained on thousands of prior contracts, a lawyer's anecdotal knowledge of past deals will become less valuable. While experience will still be important, a lawyer's memory will no longer be needed to serve as a rudimentary database of deal terms and outcomes. In addition, a lawyer's ability to "guesstimate" the effects of terms on outcomes will not be necessary when a predictive contracting system can provide concrete statistical evidence of those effects. On the other hand, skills that cannot be replicated by a predictive contracting system will become even more valuable. This is especially true of skills oriented towards human interaction, including client counseling and negotiation. Furthermore, creativity and the ability to innovate will be incredibly valuable as these skills will enable a transactional attorney to design new terms.

Predictive contracting will also affect the relationship between law firms and their clients. As de Fontenay discusses, large law firms provide value to their clients partly through market knowledge derived from data they collect on private deal terms.[239] Predictive contracting systems, however, will likely provide more extensive, robust, and granular data on contract terms than the data that is currently collected by large law firms. As a result, law firms may lose some of the advantages that their term databases have traditionally provided to them. Clients are much better situated to collect data on contract outcomes because they are the entities that are directly affected by those outcomes. By making term data

---

[239]   *See* de Fontenay, *supra* note 112 and accompanying text.

more broadly available and increasing the importance of outcome data, predictive contracting will likely weaken the position of law firms relative to their clients.

C.   Limitations and Risks

Predictive contracting faces a number of limitations and risks including: (1) technical constraints, (2) concerns regarding data privacy and confidentiality, (3) the regulation of the unauthorized practice of law, and (4) the potential for exacerbating information inequality. These issues are addressed in the Sections below.

1.   Technical Constraints

Predictive contracting faces a number of technical constraints related to data sufficiency, representativeness, bias, and interpretability.

The primary technical constraint that predictive contracting faces is collecting sufficient data on contract terms, outcomes, and conditions with which to train a machine learning model. Many contracting scenarios (like much of law) are highly complex systems.[240] Models of complex systems require substantial amounts of data to generate accurate predictions. Despite developments in contract management, natural language processing, and computable contracts, there will still be challenges to contract data collection. First, some contracting scenarios will have so many relevant terms, outcomes, and conditions that the available data set will not be able to support the complexity of the scenario. As a result, initial applications of predictive

---

[240] *See* Katz, *supra* note 22, at 962 (describing legal systems as "complex adaptive systems with elaborate levels of complexity"). For a discussion of legal complexity, see generally RICHARD A. EPSTEIN, SIMPLE RULES FOR A COMPLEX WORLD (1995); Louis Kaplow, *A Model of the Optimal Complexity of Legal Rules*, 11 J.L., ECON., & ORG. 150 (1995); Peter H. Schuck, *Legal Complexity: Some Causes, Consequences, and Cures*, 42 DUKE L.J. 1 (1992); R. George Wright, *The Illusion of Simplicity: An Explanation of Why the Law Can't Just Be Less Complex*, 27 FLA. ST. U. L. REV. 715 (2000).

contracting will likely be narrow in scope to reduce complexity and thereby require less data. As data sets become larger and more robust, users can expand their scope of analysis to include a greater number of terms, outcomes, and conditions. Second, some outcomes may occur so infrequently (such as "bet the company" litigation) that a model cannot be trained to predict these outcomes.

Third, some data may be incredibly difficult or even impossible to collect, such as the effects of relational contracting.[241] This highlights one of the key limitations of predictive contracting: if a term, condition, or outcome cannot be represented as machine-interpretable data, it cannot be analyzed using predictive contracting. Fortunately, many of these variables can be represented in a predictive contracting model with proxy variables that are much easier to obtain. With respect to relational contracting, for example, the presence of relational contracting can be proxied by a variable, such as the length of any prior contracting relationship between the parties. As discussed above, early applications of predictive contracting will focus on relatively simple contracting scenarios and will likely not include difficult-to-obtain variables such as the effects of relational contracting. As predictive contracting systems and their associated data sets improve and start to analyze more complex contracting scenarios, the ability to identify effective proxies for difficult-to-obtain variables will become a key feature of model design.

Another technical concern is whether the contracts in the training set are representative of future contracts. Machine

---

[241] *See* Surden, *supra* note 108, at 683–84 (defining a relational contract as one in which "the parties are incapable of reducing important terms of the arrangement to well-defined obligations") (quoting Charles J. Goetz & Robert E. Scott, *Principles of Relational Contracts*, 67 VA. L. REV. 1089, 1091 (1981)); Surden, *supra* note 35, at 106–07. For a discussion of relational contracting and the role of norms and institutions, see generally Lisa Bernstein, *Beyond Relational Contracts: Social Capital and Network Governance in Procurement Contracts*, 7 J. LEGAL ANALYSIS 561 (2015); Lisa Bernstein, *Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry*, 21 J. LEGAL STUD. 115 (1992); Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions*, 99 MICH. L. REV. 1724 (2001).

learning models struggle at predicting outcomes for scenarios that differ substantially from their training data.[242] For example, a contract drafter should not use a predictive contracting model trained on sales agreements to predict the cost of drafting a licensing agreement. A predictive contracting model will produce far more accurate predictions if the contracting scenarios being analyzed are similar to those on which the model was trained. Users can help ensure the representativeness of a predictive contracting model by periodically updating the training set with new contracts and using this updated data set to retrain the model.

A predictive contracting model can also produce inaccurate predictions if the training set is systematically biased.[243] For example, if a predictive contracting model is trained on a set of contracts all drafted by the same law firm, and the firm has an idiosyncratic preference for a particular set of terms, the model will be biased in favor of these terms. When training a model, contract drafters should be aware of potential biases in the training set. In addition, machine learning researchers are developing methods to debias data sets.[244]

Predictive contracting systems will also face challenges regarding the interpretability of their results. Like other prediction systems based on machine learning, predictive contracting systems will not necessarily be able to tell contract drafters *why* the identified connections between contract terms and outcomes exist. While "black box" models are useful for identifying connections that are otherwise difficult or impossible to identify, understanding the source of these

---

[242] *See* Dolin, *supra* note 82, at 2; Ng, *supra* note 72, at 30–31; Remus & Levy, *supra* note 44, at 511 (using the example of autonomous vehicles having difficulty driving on roads that contain hazards not contained in their training sets); Surden, *supra* note 35, at 105. *But see* McKamey, *supra* note 40, at 52–54.

[243] *See* Surden, *supra* note 35, at 106.

[244] *See* Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama & Adam Kalai, *Quantifying and Reducing Stereotypes in Word Embeddings*, 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications, at 43 (2016), https://arxiv.org/pdf/1606.06121.pdf [https://perma.cc/UB8W-ZDVK] (developing an algorithm to reduce gender stereotyping in text data).

connections is important for contract drafters to make informed contract design decisions.[245] Contract drafters will need to use their knowledge and expertise to interpret the results of predictive contracting models. In addition, machine learning researchers are developing methods to increase the interpretability of machine learning models.[246]

2. Privacy and Confidentiality

Concerns about data privacy and confidentiality have increased substantially in recent years. Numerous large companies have had customer data stolen or misused.[247] Governments have responded by stepping up scrutiny of how companies collect, store, and use customer data, most notably the General Data Protection Regulation in the European

[245] *See* Katz, *supra* note 22, at 950 n.198 and accompanying text; Scholz, *supra* note 156, at 160 (describing how many companies use machine learning algorithms to "poke around looking for patterns" in data without understanding why those patterns exist).

[246] *See* Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas & Rory Sayres, Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV) 1 (June 7, 2018) (unpublished manuscript), https://arxiv.org/abs/1711.11279 [https://perma.cc/ARA9-BX8N] (introducing Concept Activation Vectors, which can be used to assist in the interpretation of machine learning models).

[247] In 2017, Equifax, one of the major consumer credit agencies in the United States, experienced a massive security breach that compromised the personal information of tens of millions of Americans. *See* Tara Siegel Bernard, Tiffany Hsu, Nicole Perlroth & Ron Lieber, *Equifax Says Cyberattack May Have Affected 143 Million in the U.S.*, N.Y. TIMES (Sep. 7, 2017), https://www.nytimes.com/2017/09/07/business/equifax-cyberattack.html [https://perma.cc/W7RM-D97L]. In 2018, Facebook announced that the data of millions of users had been improperly accessed by Cambridge Analytica, a political consulting firm associated with the Trump campaign during the 2016 U.S. presidential election. *See* Cecilia Kang & Sheera Frenkel, *Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html [https://perma.cc/A5QQ-Y575]. The Cambridge Analytica scandal resulted in Mark Zuckerberg, Facebook's founder and CEO, being called to testify before Congress about Facebook's data security and privacy controls. *Id.*

Union.[248] Predictive contracting, like other technology systems that use data to predict outcomes, will encounter privacy challenges related to the collection, storage, and use of contract data.[249] In addition, there will likely be added complexity due to professional regulations that restrict how lawyers may share confidential client information.[250] Contract technology companies frequently cite data privacy as one of their clients' main concerns.[251] For contract technology companies developing machine learning systems, user privacy concerns are the main hurdle to pooling contract data from

[248] *See* EUGDPR, https://www.eugdpr.org [https://perma.cc/8WBN-XNU7] ("The EU General Data Protection Regulation (GDPR). . .[was] designed to [h]armonize data privacy laws across Europe, [to] protect and empower all EU citizens data privacy, [and to] [r]eshape the way organizations across the region approach data privacy."). Violations of GDPR can result in potentially serious penalties for offending companies. *See GDPR Key Changes,* EUGDPR, https://www.eugdpr.org/the-regulation.html [https://perma.cc/4ZW9-6WT7] (noting that under GDPR, "[o]rganizations in breach of GDPR can be fined up to 4% of annual global turnover or €20 Million (whichever is greater).").

[249] *See* Casey & Niblett, *supra* note 42, at 50–51. For a discussion of the intersection of privacy law and data collection, see generally Lisa Austin, *Privacy and the Question of Technology*, 22 L. & PHIL. 119 (2003); Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63 (2012); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH & INTELL. PROP. 239 (2013); Paul M. Schwartz, *Information Privacy in the Cloud*, 161 U. PA. L. REV. 1623 (2013); PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT (Julia Lane, Victoria Stodden, Stefan Bender & Helen Nissenbaum eds., 2014).

[250] *See* MODEL RULES OF PROF'L CONDUCT r. 1.6(a) (AM. BAR ASS'N 2008) ("A lawyer shall not reveal information relating to the representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation or the disclosure is permitted by paragraph (b)"); *see also* de Fontenay, *supra* note 112, at 428–30.

[251] Interview with Contract Room Representative, in Palo Alto, Cal., *supra* note 96; Telephone Interview with Beagle Representative, *supra* note 119; Telephone Interview with Contract Assistant Representative, *supra* note 111; Telephone Interview with Contract Standards Representative (Mar. 6, 2018); Telephone Interview with Kira Systems Representative, *supra* note 60.

multiple customers to increase the size of training sets.[252] Companies that build predictive contracting systems will need to be highly diligent with respect to how they handle user data. A security breach or other misuse of user data could mean the end of an otherwise promising predictive contracting company.

Privacy also presents a problem for computable contracts, one of the key sources of contract data for predictive contracting systems. Due to the peer-to-peer nature of blockchain systems, the transactions on many computable contracts platforms are public. For sensitive contracting scenarios, parties are unlikely to use publicly viewable computable contracts due to privacy concerns. In response to this concern, computable contracts companies such as Oasis Labs are beginning to develop platforms that support "privacy-preserving" computable contracts.[253]

3. Unauthorized Practice of Law

A core feature of the regulation of the legal profession is that non-lawyers may not practice law.[254] If a non-lawyer (or a company owned or managed by a non-lawyer) renders legal services, they run the risk of incurring liability for what is generally referred to as the "unauthorized practice of law" ("UPL").[255] While originally intended to prevent human non-lawyers from practicing law, UPL regulations have been applied to legal technology companies owned and operated by non-lawyers.[256] Legal technology commentators have noted

---

[252] Telephone Interview with Beagle Representative, *supra* note 119; Telephone Interview with Contract Standards Representative, *supra* note 251; Telephone Interview with Kira Systems Representative, *supra* note 60.

[253] *See Privacy-First Cloud Computing on Blockchain: A Non-Technical Primer*, OASIS LABS, Nov. 2018, https://docsend.com/view/fsdz4hv [https://perma.cc/F5Z2-YKUY].

[254] *See e.g.*, CAL. BUS. & PROF. CODE §§ 6125–6133 (2018).

[255] *Id.*

[256] In 2017, three New Jersey Supreme Court Committees (Advisory Committee on Professional Ethics, Committee on Attorney Advertising, and Committee on the Unauthorized Practice of Law) issued a joint opinion finding that AVVO, LegalZoom, and Rocket Lawyer were all in violation of

that UPL regulations could create challenges for the use of machine learning in the provision of legal services.[257]

The use of predictive contracting systems could potentially be limited by UPL regulations. Whether UPL regulations will apply to predictive contracting will depend in large part on whether the contract drafter using the predictive contracting system is a lawyer. The ABA has largely embraced the use of machine learning technology by lawyers.[258] The ABA views the use of such technology by lawyers similarly to employing a non-lawyer assistant, which is permissible under UPL regulations.[259] This includes legal technology provided to a lawyer by a third-party company owned and operated by non-lawyers.[260] As a result, the use of predictive contracting systems by contract drafters who are lawyers will likely be permissible under UPL regulations. The permissibility of non-lawyer contract drafters using predictive contracting, on the other hand, is much less clear. The general view has been that a company owned and operated by non-lawyers violates UPL regulations by providing legal services via a technology system.[261] A recent court case, however, has called into question whether services provided by a technology system qualify as practicing law. In *Lola v. Skadden, Arps, Slate, Meagher & Flom, LLP*, the Second Circuit stated that "an

UPL regulations (AVVO was found to offer an impermissible referral service, whereas LegalZoom and Rocket Lawyer were found to be unregistered with the New Jersey Supreme Court). *See* N.J. Advisory Comm. on Prof'l Ethics, ACPE Joint Opinion 732, at 8 (June 21, 2017), https://www.themodernfirm.com/wp-content/uploads/2017/09/ACPE-732-Avvo-LegalZoom-Rocket-Lawyer-6.21.17.pdf [https://perma.cc/2NEQ-UNUW].

[257] *See* Betts & Jaep, *supra* note 160, at 232; Dolin, *supra* note 82, at 3; McGinnis & Pearce, *supra* note 40, at 3059–64; Larry E. Ribstein, *The Death of Big Law*, 2010 WIS. L. REV. 749, 807–08 (2010); Ray Worthy Campbell, *Rethinking Regulation and Innovation in the U.S. Legal Services Market*, 9 N.Y.U. J.L. & BUS. 1, 45–51 (2012).

[258] *See* McGinnis & Pearce, *supra* note 40, at 3059–61.

[259] *Id.* at 3060.

[260] *Id.* at 3060–61 (citing MODEL RULES OF PROF'L CONDUCT r. 5.3 cmt. 3 (AM. BAR ASS'N 2013)).

[261] *Id.* at 3061–64.

individual who . . . undertakes tasks that could otherwise be performed entirely by a machine cannot be said to engage in the practice of law."[262] This reasoning suggests that the use of a technology system such as predictive contracting may not qualify as practicing law even if the individuals using the system are non-lawyers.

4. Information Inequality

Predictive contracting provides contract drafters with better information than they would otherwise have access to under traditional contracting. This information enables contract drafters to craft more effective contracts that increase the joint value generated by a transaction.[263] The analysis of predictive contracting thus far, however, has assumed that all parties to a contract have access to predictive contracting systems. If, on the other hand, only one party has access to predictive contracting, that party will have a substantial information advantage during the negotiation and design of the contract. When one party has a large information advantage, that party tends to use its advantage to extract value from its counterparty via potentially inefficient but allocatively advantageous terms.[264] In the context of business-to-business contracting, it is reasonable to assume that as predictive contracting systems demonstrate their value in the market, more businesses will start to use such systems, thereby decreasing the likelihood that a party will have an information advantage relative to its counterparty. This assumption cannot be made, however, in the context of business-to-consumer contracting. There is a large body of literature on consumer contracts that highlights the information and bargaining power disparity between

---

[262] Lola v. Skadden, Arps, Slate, Meagher & Flom L.L.P., 620 F. App'x. 37, 45 (2d Cir. 2015) (holding that document review did not per se meet North Carolina's definition of practicing law). This case is the first to use this reasoning.

[263] *See supra* Section III.A.6.

[264] *See supra* Section III.A.6.

businesses and consumers.[265] Unlike businesses, consumers are highly unlikely to have access to predictive contracting systems. As a result, predictive contracting has the potential to exacerbate the information inequality that already exists in the consumer contracting context. For example, a business could use a predictive contracting system to identify the set of contract terms that best advantage the business at the expense of the consumer.[266] The potential effects of predictive contracting on consumer contracts will need to be closely monitored.[267]

IV. CONCLUSION

This Article examined how contract drafters can use data on contract outcomes to inform contract design. Building on recent developments in contract data collection and analysis, the Article proposed predictive contracting, a new method of contracting in which contract drafters can design contracts using a technology system that helps predict the connections between contract terms and outcomes. The Article then discussed the theoretical and practical implications of predictive contracting. On a theoretical level, predictive contracting can lead to greater customization, increased innovation, more complete contract design, more effective balancing of front-end and back-end costs, better risk assessment and allocation, and more accurate term pricing for negotiation. On a practical level, predictive contracting has the potential to significantly alter the role of transactional

---

[265] *See supra* note 74.

[266] *But see* G. Marcus Cole, *Rational Consumer Ignorance: When and Why Consumers Should Agree to Form Contracts Without Even Reading Them*, 11 J.L., ECON. & POL'Y 413, 413–16 (2015) (arguing that consumer contracts in competitive markets should be consumer-favorable due to competition between businesses with respect to contract terms. In less competitive markets, however, businesses with monopoly or oligopoly market positions are more likely to include terms harmful to consumers).

[267] Consumer protection organizations, such as CLAUDETTE, are starting to use machine learning to identify unlawful and/or harmful terms in consumer contracts. *See About*, CLAUDETTE, http://claudette.eui. eu/about/index.html [https://perma.cc/EM8Z-WZXG].

lawyers by providing them with access to previously unavailable information on the statistical connections between contract terms and outcomes. The Article also discussed a number of risks and limitations faced by predictive contracting, including technical constraints, concerns regarding data privacy and confidentiality, the regulation of the unauthorized practice of law, and the potential for exacerbating information inequality.

Further research is required to more fully develop predictive contracting. The first step is to build a working prototype of a predictive contracting system using real-world contract data. In addition to developing a prototype, future research should examine how computable contracts can best be used as mechanisms for contract data collection. Of the three sources of contract data discussed in this Article, computable contracts have the greatest potential to collect machine-readable contract data at scale. While computable contracts are still in a nascent stage of development, blockchain companies, such as Ethereum, are starting to push computable contracts towards mainstream use. Designing computable contracting systems with data collection in mind is critical for the long-term success of predictive contracting. Lastly, subsequent research should explore the risks posed by predictive contracting, particularly the potential for exacerbating information inequality between asymmetrically situated parties such as in consumer contracting.