Washington Law Review

Volume 89
Number 1 *Symposium: Artificial Intelligence and the Law*

3-1-2014

Machine Learning and Law

Harry Surden

Follow this and additional works at: https://digitalcommons.law.uw.edu/wlr

 Part of the Science and Technology Law Commons

Recommended Citation
Harry Surden, Essay, *Machine Learning and Law*, 89 Wash. L. Rev. 87 (2014).
Available at: https://digitalcommons.law.uw.edu/wlr/vol89/iss1/5

This Essay is brought to you for free and open access by the Law Reviews and Journals at UW Law Digital Commons. It has been accepted for inclusion in Washington Law Review by an authorized editor of UW Law Digital Commons. For more information, please contact cnyberg@uw.edu.

MACHINE LEARNING AND LAW

Harry Surden*

INTRODUCTION

What impact might artificial intelligence (AI) have upon the practice of law? According to one view, AI should have little bearing upon legal practice barring significant technical advances.[1] The reason is that legal practice is thought to require advanced cognitive abilities, but such higher-order cognition remains outside the capability of current AI technology.[2] Attorneys, for example, routinely combine abstract reasoning and problem solving skills in environments of legal and factual uncertainty.[3] Modern AI algorithms, by contrast, have been unable to replicate most human intellectual abilities, falling far short in advanced cognitive processes—such as analogical reasoning—that are basic to legal practice.[4] Given these and other limitations in current AI technology, one might conclude that until computers can replicate the higher-order cognition routinely displayed by trained attorneys, AI would have little impact in a domain as full of abstraction and uncertainty as law.[5]

Although there is some truth to that view, its conclusion is overly broad. It misses a class of legal tasks for which current AI technology

---

* Associate Professor of Law, University of Colorado Law School; B.A. Cornell University; J.D. Stanford University; Affiliated Faculty, Stanford Codex Center for Legal Informatics. I would like to thank my colleagues at the University of Colorado for their insightful comments, and Ted Sichelman, Seema Shah, and Dan Katz for their helpful observations and suggestions.

1. *See, e.g.*, Symposium, *Legal Reasoning and Artificial Intelligence: How Computers "Think" Like Lawyers*, 8 U. CHI. L. SCH. ROUNDTABLE 1, 19 (2001) (Cass Sunstein argues that, "[A]t the present state of the art artificial intelligence cannot engage in analogical reasoning or legal reasoning").

2. *See, e.g.*, Karl Okamoto, *Teaching Transactional Lawyering*, 1 DREXEL L. REV. 69, 83 (2009) ("The essence of lawyering is 'creative problem solving' under conditions of uncertainty and complexity. This conception of lawyering as problem solving has become commonplace.").

3. *Id.* at 83.

4. *Id.*

5. *See* Harry Surden, *Computable Contracts*, 46 U.C. DAVIS L. REV. 629, 646 (2012) (discussing how language changes that are typically trivial for humans to decipher may confuse computer algorithms).

can still have an impact even given the technological inability to match human-level reasoning. Consider that outside of law, non-cognitive AI techniques have been successfully applied to tasks that were once thought to necessitate human intelligence—for example language translation.[6] While the results of these automated efforts are sometimes imperfect, the interesting point is that such computer generated results have often proven useful for particular tasks where strong approximations are acceptable.[7] In a similar vein, this Article will suggest that there may be a limited, but not insignificant, subset of legal tasks that are capable of being partially automated using current AI techniques despite their limitations relative to human cognition.

In particular, this Article focuses upon a class of AI methods known as "machine learning" techniques and their potential impact upon legal practice. Broadly speaking, machine learning involves computer algorithms that have the ability to "learn" or improve in performance over time on some task.[8] Given that there are multiple AI approaches, why highlight machine learning in particular? In the last few decades, researchers have successfully used machine learning to automate a variety of sophisticated tasks that were previously presumed to require human cognition. These applications range from autonomous (i.e., self-driving) cars, to automated language translation, prediction, speech recognition, and computer vision.[9] Researchers have also begun to apply these techniques in the context of law.[10]

To be clear, I am not suggesting that all, or even most, of the tasks routinely performed by attorneys are automatable given the current state of AI technology. To the contrary, many of the tasks performed by attorneys do appear to require the type of higher order intellectual skills that are beyond the capability of current techniques. Rather, I am suggesting that there are subsets of legal tasks that are likely

---

6. *See* DAVID BELLOS, IS THAT A FISH IN YOUR EAR?: TRANSLATION AND THE MEANING OF EVERYTHING 253–57 (2011); *Find Out How Our Translations Are Created*, GOOGLE, http://translate.google.com/about (last visited Feb. 24, 2014).

7. *See* BELLOS, *supra* note 6.

8. PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA 3 (2012).

9. Burkhard Bilger, *Auto Correct: Has the Self-Driving Car at Last Arrived?*, NEW YORKER, Nov. 25, 2013, at 96, 106; PARAG KULKARNI, REINFORCEMENT AND SYSTEMIC MACHINE LEARNING FOR DECISION MAKING 1–2 (2012) (discussing computer vision).

10. *See, e.g.*, 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, 936 (2013) (discussing legal applications such as automation in document discovery and quantitative legal prediction).

automatable under the current state of the art, provided that the technologies are appropriately matched to relevant tasks, and that accuracy limitations are understood and accounted for. In other words, even given current limitations in AI technology as compared to human cognition, such computational approaches to automation may produce results that are "good enough" in certain legal contexts.

Part I of this Article explains the basic concepts underlying machine learning. Part II will convey a more general principle: non-intelligent computer algorithms can sometimes produce intelligent results in complex tasks through the use of suitable proxies detected in data. Part III will explore how certain legal tasks might be amenable to partial automation under this principle by employing machine learning techniques. This Part will also emphasize the significant limitations of these automated methods as compared to the capabilities of similarly situated attorneys.

I. OVERVIEW OF MACHINE LEARNING

A. What Is Machine Learning?

"Machine learning" refers to a subfield of computer science concerned with computer programs that are able to learn from experience and thus improve their performance over time.[11] As will be discussed, the idea that the computers are "learning" is largely a metaphor and does not imply that computers systems are artificially replicating the advanced cognitive systems thought to be involved in human learning.[12] Rather, we can consider these algorithms to be learning in a *functional* sense: they are capable of changing their behavior to enhance their performance on some task through experience.[13]

Commonly, machine learning algorithms are used to detect patterns in data in order to automate complex tasks or make predictions.[14] Today, such algorithms are used in a variety of real-world commercial applications including Internet search results, facial recognition, fraud

11. STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 693 (3d ed. 2010).

12. I. H. WITTEN, DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES § 1.3 (3d ed. 2011).

13. *Id.*

14. David E. Sorkin, *Technical and Legal Approaches to Unsolicited Electronic Mail*, 35 U.S.F. L. REV. 325, 326 (2001).

detection, and data mining.[15] Machine learning is closely associated with the larger enterprise of "predictive analytics" as researchers often employ machine learning methods to analyze existing data to predict the likelihood of uncertain outcomes.[16] If performing well, machine learning algorithms may produce automated results that approximate those that would have been made by a similarly situated person. Machine learning is thus often considered a branch of artificial intelligence, since a well-performing algorithm may produce automated results that appear "intelligent."[17]

The goal of this Part is to convey some basic principles of machine learning in a manner accessible to non-technical audiences in order to express a larger point about the potential applicability of these techniques to tasks within the law.

1. Email Spam Filters as an Example of Machine Learning

Consider a familiar example—email "spam" filters—that will illustrate some basic features common to machine learning techniques. "Spam" emails are unsolicited, unwanted commercial emails that can interfere with a user accessing more important communications.[18] In principle, an email user could manage spam manually by reading each email, identifying whether a given email is spam, and deleting those determined to be spam. However, given that this task is labor intensive, it would be desirable to automate spam identification. To perform such automated filtering of spam, email software programs frequently use machine learning algorithms.[19]

How do machine learning algorithms automatically identify spam? Such algorithms are designed to detect patterns among data. In a typical process, a machine learning algorithm is "trained" to recognize spam emails by providing the algorithm with known examples of spam for pattern analysis. For instance, imagine that a person determines that a particular email is spam and flags it as such using her email reading software. We can think of this act of flagging as an indication to the computer algorithm that this is a verified example of a spam email that

---

15. WITTEN, *supra* note 12, at § 1.3.

16. *See, e.g.*, LAWRENCE MAISEL, PREDICTIVE BUSINESS ANALYTICS: FORWARD LOOKING CAPABILITIES TO IMPROVE BUSINESS PERFORMANCE, 27–30 (2014).

17. RUSSEL & NORVIG, *supra* note 11, at 3–5.

18. Sorkin, *supra* note 14, at 325–30.

19. *Id.*

should be assessed for patterns.[20]

In analyzing the spam email, the machine learning algorithm will attempt to detect the telltale characteristics that indicate that a given email is more likely than not to be spam. After analyzing several such examples, the algorithm may detect a pattern and infer a general "rule"[21]—for instance that emails with the phrase "Earn Extra Cash" tend to be statistically more likely to be spam emails than wanted emails. It can then use such learned indicia to make automated assessments about the likelihood that a new incoming email is or is not spam.[22]

In general, machine learning algorithms are able to automatically build such heuristics by inferring information through pattern detection in data. If these heuristics are correct, they will allow the algorithm to make predictions or automated decisions involving future data.[23] Here, the algorithm has detected a pattern within the data provided (i.e., the set of example spam emails) that, of the emails that were flagged as spam, many of them contained the phrase "Earn Extra Cash." From this pattern, it then inferred a heuristic: that emails with the text "Earn Extra Cash" were more likely to be spam. Such a generalization can thus be applied going forward to automatically categorize new incoming emails containing "Earn Extra Cash" as spam. The algorithm will attempt to detect other similar patterns that are common among spam emails that can be used as a heuristic for distinguishing spam from wanted emails.

Importantly, machine learning algorithms are designed to improve in performance over time on a particular task as they receive more data. The goal of such an algorithm is to build an internal computer model of some complex phenomenon—here spam emails—that will ultimately allow the computer to make automated, accurate classification decisions.

---

20. In many cases, machine learning algorithms are trained through carefully validated training sets of data, in which the data has been carefully screened and categorized by people. *See, e.g.*, DAVID BARBER, BAYESIAN REASONING AND MACHINE LEARNING 290–96 (2011).

21. The term "rule" is used approximately in the sense of "rule of thumb." This is important, because machine learning is an *inductive* rather a *deductive* technique. In a deductive approach, general logical rules (statements) characterizing the state of the world are expressly articulated, and information is extracted by combining statements according to logical operations. By contrast, in an *inductive* approach, models of the world are developed upon observing the past and expressing the state of the world (often) in probabilities induced from observation, rather than as general rules. *See generally* Katz, *supra* note 10, at 946.

22. To be clear, this is an extreme over-simplification of machine learning for illustrative purposes. Moreover, there are many different machine learning algorithmic strategies other than the particular one illustrated here. *See generally* MEHRYAR MOHRI ET AL., FOUNDATIONS OF MACHINE LEARNING (2012).

23. TOBY SEGARAN, PROGRAMMING COLLECTIVE INTELLIGENCE: BUILDING SMART WEB 2.0 APPLICATIONS 3 (2007).

In this case, the internal model would include multiple rules of thumb about the likely characteristics of spam induced over time—in addition to the "Earn Extra Cash" heuristic just described—that the computer can subsequently follow to classify new, incoming emails.

For instance, such an algorithm might infer from additional spam examples that emails that originate from the country Belarus[24] tend to be more likely to be spam than emails from other countries. Similarly, the algorithm might learn that emails sent from parties that the reader has previously corresponded with are less likely to be spam than those from complete strangers. These additional heuristics that the algorithm learned from analyzing additional data will allow it to make better automated decisions about what is or is not spam.

As illustrated, the rule sets[25] that form the internal model are inferred by examining and detecting patterns within data. Because of this, such rule-sets tend to be built cumulatively over time as more data arrives. Machine learning algorithms typically develop heuristics incrementally by examining each new example and comparing it against prior examples to identify overall commonalities that can be generalized more broadly. For example, an algorithm may have to analyze several thousand examples of spam emails before it detects a reliable pattern such that the text "Earn Extra Cash" is a statistical indicia of likely spam.

For this reason, a machine learning algorithm may perform poorly at first when it has only had a few examples of a phenomenon (e.g., spam emails) from which to detect relevant patterns. At such an early point, its internal rule-set will likely be fairly underdeveloped. However, the ability to detect useful patterns tends to improve as the algorithm is able to examine more examples of the phenomenon at issue. Often, such an algorithm will need data with many hundreds or thousands examples of the relevant phenomenon in order to produce a useful internal model (i.e. robust set of predictive computer rules).[26]

The prior example illustrates what is meant by "learning" in the machine learning context: it is this ability to improve in performance by detecting new or better patterns from additional data. A machine

---

24. *See* Paul Ducklin, *Dirty Dozen Spam Sending Nations*, NAKED SECURITY (Oct. 17, 2013), http://nakedsecurity.sophos.com/2013/10/17/dirty-dozen-spam-sending-nations-find-where-you-finished-in-our-q3-spampionship-chart/.

25. It is important to note that these rule-sets are often actually mathematical functions or some other data structure representing the object to be modeled, rather than a series of formal, general rules. *See* KULKARNI, *supra* note 9, at 2–10.

26. CHRISTOPHER D. MANNING, INTRODUCTION TO INFORMATION RETRIEVAL 335 (2008).

learning algorithm can become more accurate at a task (like classifying email as spam) over time because its design enables it to continually refine its internal model by analyzing more examples and inferring new, useful patterns from additional data.

This capability to improve in performance over time by continually analyzing data to detect additional useful patterns is the key attribute that characterizes machine learning algorithms. Upon the basis of such an incrementally produced model, a well-performing machine learning algorithm may be able to automatically perform a task—such as classifying incoming emails as either spam or wanted emails—with a high degree of accuracy that approximates the classifications that a similarly situated human reviewer would have made.[27]

2. Detecting Patterns to Model Complex Phenomena

There are a few points to emphasize about the above example. First, machine learning often (but not exclusively) involves learning from a set of verified examples of some phenomenon. Thus, in the prior example, the algorithm was explicitly provided with a series of emails that a human predetermined to be spam, and learned the characteristics of spam by analyzing these provided examples. This approach is known as "supervised" learning, and the provided examples upon which the algorithm is being trained to recognize patterns are known as the "training set."[28] The goal of such training is to allow the algorithm to create an internal computer model of a given phenomenon that can be generalized to apply to new, never-before-seen examples of that phenomenon.

Second, such machine learning algorithms are able to automatically build accurate models of some phenomenon—here the characteristics of spam email—without being explicitly programmed.[29] Most software is developed by a manual approach in which programmers explicitly specify a series of rules for a computer to follow that will produce some desired behavior. For instance, if designing a spam filter by this manual method, a programmer might first consider the features that she believed to be characteristic of spam, and then proceed to program a computer

---

27. WILLIAM S. YERAZUNIS, THE SPAM-FILTERING ACCURACY PLATEAU AT 99.9 PERCENT ACCURACY AND HOW TO GET PAST IT (Dec. 2004), *available at* http://www.merl.com/reports/docs/TR2004-091.pdf (noting that many spam filters have achieved accuracy rates at over 99.9%).

28. FLACH, *supra* note 8, at 2.

29. Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, COMM. ACM, Oct. 2012, at 80.

with a series of corresponding rules to make automated distinctions.

However, many phenomena are so complicated and dynamic that it is difficult to model them manually.[30] The problem with a manual, bottom-up approach to modeling complex and changing phenomenon (such as spam) is that it is very difficult to specify a rule set ex-ante that would be robust and accurate enough to direct a computer to make useful, automated decisions. For instance, a programmer might not think to include a rule that an email with a Belarus origin should be considered somewhat more likely to be spam. It is often difficult to explicitly program a set of computer rules to produce useful automation when dealing with complex, changing phenomenon.

Machine learning algorithms, by contrast, are able to incrementally build complex models by automatically detecting patterns as data arrives. Such algorithms are powerful because, in a sense, these algorithms program themselves over time with the rules to accomplish a task, rather than being programmed manually with a series of pre-determined rules.[31] The rules are inferred from analyzed data and the model builds itself as additional data is analyzed. For instance, in the above example, as the algorithm encountered new examples of spam with different features, it was able to add to its internal model additional markers of spam that it was able to detect (e.g., emails originating from Belarus). Such an incremental, adaptive, and iterative process often allows for the creation of nuanced models of complex phenomena that may otherwise be too difficult for programmers to specify manually, up front.[32]

Third, what made the discussed spam filtering algorithm a machine *learning* algorithm was that it was able to *improve* its accuracy in classifying spam as it received more examples to analyze. In this sense, we are using a functional meaning of "learning." The algorithms are not learning in the cognitive sense typically associated with human learning. Rather, we can think of the algorithms as learning in the sense that they are changing their behavior to perform better in the future as they receive more data.[33] Thus, in the above example, if the spam filter

30. *Id.*

31. TOM MITCHELL, THE DISCIPLINE OF MACHINE LEARNING, REPORT NO. ML-06-CMU-108 § 1 (2006), *available at* http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf ("Machine Learning focuses on . . . how to get computers to program themselves (from experience plus some initial structure).").

32. *Id.* ("[S]peech recognition accuracy is greater if one trains the system, than if one attempts to program it by hand.").

33. I. H. WITTEN, DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES 8 (2d ed. 2005).

algorithm became more accurate at identifying spam as it received more examples of spam and refined its internal rule-set. We can conceptualize this shift as "learning" from a functional perspective in an analogous way that we often associate human learning with improved performance on some task.

Fourth, the filtering algorithm described used statistical techniques to classify spam. Machine learning algorithms are often (although not exclusively) statistical in nature. Thus, in one sense, machine learning is not very different from the numerous statistical techniques already widely used within empirical studies in law.[34] One salient distinction is that while many existing statistical approaches involve fixed or slow-to-change statistical models, the focus in machine learning is upon computer algorithms that are expressly designed to be dynamic and capable of changing and adapting to new and different circumstances as the data environment shifts.

II. INTELLIGENT RESULTS WITHOUT INTELLIGENCE

A. Proxies and Heuristics for Intelligence

The prior example was meant to illustrate a broader point: one can sometimes accomplish tasks associated with human intelligence with non-intelligent computer algorithms. There are certain tasks that appear to require intelligence because when humans perform them, they implicate higher-order cognitive skills such as reasoning, comprehension, meta-cognition, or contextual perception of abstract concepts. However, research has shown that certain of these tasks can be automated—to some degree—through the use of non-cognitive computational techniques that employ heuristics or proxies (e.g., statistical correlations) to produce useful, "intelligent" results. By a proxy or heuristic, I refer to something that is an effective stand-in for some underlying concept, feature, or phenomenon.

To say it differently, non-cognitive computer algorithms can sometimes produce "intelligent" results in complex tasks without human-level cognition. To employ a functional view of intelligence, such automated results can be considered "intelligent" to the extent that they approximate those that would have been produced by a similarly situated person employing high-level human cognitive processes. This is an outcome-oriented view of intelligence—assessing based upon

34. *See, e.g.*, David L. Schwartz, *Practice Makes Perfect? An Empirical Study of Claim Construction Reversal Rates in Patent Cases*, 107 MICH. L. REV. 223 (2008).

whether the results that were produced were sensible and useful—rather than whether the underlying process that produced them was "cognitive" in nature.

The machine learning spam filtering example illustrated this idea. We might normally think of the identification of spam email by a person as entailing a series of advanced cognitive processes. A human user determining whether a particular email is spam may do the following: visually process the email, read, absorb, and understand the language of the email text, contextualize the meaning of the email contents, reason about whether or not the email was solicited, and based upon that assessment determine whether the email constituted unwanted spam.[35]

One might conclude that, because spam determination involves intelligence when conducted by people, the task is inherently cognitive. In terms of automation, however, most of the advanced cognitive processes just described have not been artificially matched by computer systems to any significant degree.[36] Given that identifying spam emails appears to involve cognition, and that computers have not been able to replicate advanced human level cognitive processes—such as understanding arbitrary written text at the level of a literate person—one might presume it would not be possible to automate a task as abstract as identifying spam emails.[37]

However, in the example described earlier, the machine learning algorithm was able to automate the task of spam filtering through non-cognitive processes. Through the use of pattern detection, the algorithm was able to infer effective proxy markers for spam emails: that emails with the text "Earn Extra Cash" or with an origin from Belarus were statistically more likely to be spam. On that basis, the algorithm was able to make automated classifications that were useful and "intelligent" in

---

35. *See, e.g.*, Argye E. Hillis & Alfonso Caramazza, *The Reading Process and Its Disorders*, *in* COGNITIVE NEUROPSYCHOLOGY IN CLINICAL PRACTICE 229, 229–30 (David Ira Margolin ed., 1992) ("[A] cognitive process such as reading involves a series of transformations of mental representations. . . . On this view, even very simple cognitive tasks will involve various processing mechanisms . . . .").

36. RUSSELL & NORVIG, *supra* note 11, at 3–10.

37. For detailed explanations of the limits of Natural Language Processing (NLP) as of the writing of this Article, *see* RUSSELL & NORVIG, *supra* note 11, at 860–67; Robert Dale, *Classical Approaches to Natural Language Processing*, *in* HANDBOOK OF NATURAL LANGUAGE PROCESSING 1, 1–7 (Nitin Indurkhya & Frederick J. Damerau eds., 2d ed. 2010); Richard Socher et al., *Semantic Compositionality through Recursive Matrix-Vector Spaces*, *in* CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING § 1 (2012) (noting that particular NLP approaches are limited and "do not capture . . . the important quality of natural language that allows speakers to determine the meaning of a longer expression based on the meanings of its words and the rules used to combine them").

the sense that they approximated what a human user would have done after reading and comprehending the email.

However, notably, the algorithm did not engage in the meaning or substance of the email text in a manner comparable to a similarly situated person, nor did it need to.[38] In other words, the algorithm did not need to understand abstract concepts such as "email," "earning cash," "Belarus," or "spam"—in the way that a person does—in order to make accurate automatic spam classifications. Rather, it was able to detect statistical proxies for spam emails that allowed it to produce useful, accurate results, without engaging in the underlying meaning or substance of the email's constituent words.

Thus, the machine learning spam filter example illustrated a rather profound point: it is sometimes possible to accomplish a task typically associated with cognition not through artificial simulations of human intellectual processes, but through algorithms that employ heuristics and proxies—such as statistical correlations learned through analyzing patterns in data—that ultimately arrive at the same or similar results as would have been produced by a similarly situated intelligent person employing higher order cognitive processes and training.

1. Approximating Intelligence by Proxy

More generally, the example is illustrative of a broader strategy that has proven to be successful in automating a number of complex tasks: detecting proxies, patterns, or heuristics that reliably produce useful outcomes in complex tasks that, in humans, normally require intelligence.[39] For a certain subset of tasks, it may be possible to detect proxies or heuristics that closely track the underlying phenomenon without actually engaging in the full range of abstraction underlying that phenomenon, as in the way the machine learning algorithm was able to identify spam emails without having to fully understand substance and context of the email text. As will be discussed in Part III this is the principle that may allow the automation of certain abstract tasks within law that, when conducted by attorneys, require higher order cognition.

It is important to emphasize that such a proxy-based approach can have significant limitations. First, this strategy may only be appropriate for certain tasks for which approximations are suitable. By contrast, many complicated problems—particularly those that routinely confront attorneys—may not be amenable to such a heuristic-based technique.

38. SEGARAN, *supra* note 23, at 4.

39. *Id.* at 1–3.

For example, an attorney counseling a corporate client on a potential merger is a task of such scale, complexity, and nuance, with so many considerations, that a simple proxy approach would be inappropriate.

Second, a proxy-based strategy can often have significant accuracy limitations. Because proxies are stand-ins for some other underlying phenomenon, they necessarily are under- and over-inclusive relative to the phenomenon they are representing, and inevitably produce false positives and negatives. By employing proxies to analyze or classify text with substantive meaning for an abstract task, for example, such algorithms may produce more false positives or negatives than a similarly situated person employing cognitive processes, domain knowledge, and expertise. Thus, for example, automated spam-filters can often do a reasonably accurate job of classifying spam, but often make errors in substantively complex cases that would be trivial for a person to detect.[40] However, once the limitations are properly understood, for certain common purposes (e.g., classifying emails) where the efficiency of automation is more important than precision, such approximations may be sufficient.

2. Developments in AI Research

The strategy just described parallels changes among computer science artificial intelligence research over the last several decades. In the earliest era of AI research—from the 1950s through the 1980s—many researchers focused upon attempting to replicate computer-based versions of human cognitive processes.[41] Behind this focus was a belief that because humans employ many of the advanced brain processes to tackle complex and abstract problems, the way to have computers display artificial intelligence was to create artificial versions of brain functionality.[42]

However, more recently, researchers have achieved success in automating complex tasks by focusing not upon the intelligence of the automated processes themselves, but upon the results that automated processes produce.[43] Under this alternative view, if a computer system is able to produce outputs that people would consider to be accurate, appropriate, helpful, and useful, such results can be considered "intelligent"—even if they did not come about through artificial versions

40. YERAZUNIS, *supra* note 27, at 1–5.

41. *See, e.g.*, RUSSELL & NORVIG, *supra* note 11, at 3–10.

42. *Id.*

43. *See* Surden, *supra* note 5, at 685–86.

of human cognitive processes.

In general, this has been the approach followed by many successful AI systems of the past several years. These systems have used machine learning and other techniques to develop combinations of statistical models, heuristics, and sensors that would not be considered cognitive in nature (in that they do not replicate human-level cognition) but that produce results that are useful and accurate enough for the task required.[44] As described, these proxy-based approaches sometimes lack accuracy or have other limitations as compared to humans for certain complex or abstract tasks. But the key insight is that for many tasks, algorithmic approaches like machine learning may sometimes produce useful, automated approaches that are "good enough" for particular tasks.

A good example of this principle comes from the task of language translation. For many years, the translation of foreign languages was thought to be a task deeply connected with higher-order human cognitive processes.[45] Human translators of foreign languages call upon deep knowledge of languages, and abstract understanding of concepts, to translate foreign language documents. Many early AI projects sought to replicate in computers various language rules believed to reside within the human brain.[46] However, these early, bottom-up, rules-based language translation systems produced poor results on actual translations.[47]

More recent research projects have taken a different approach, using statistical machine learning and access to large amounts of data to produce surprisingly good translation results without attempting to replicate human-linguistic processes.[48] "Google Translate," for example, works in part by leveraging huge corpuses of documents that experts previously translated from one language to another. The United Nations (UN) has for instance, over the years, employed professional translators to carefully translate millions of UN documents into multiple languages, and this body of translated documents has become available in electronic

---

44. *See* RUSSELL & NORVIG, *supra* note 11, at 3–10.

45. *See* EKATERINA OVCHINNIKOVA, INTEGRATION OF WORLD KNOWLEDGE FOR NATURAL LANGUAGE UNDERSTANDING 215–20 (2012).

46. Mathias Winther Madsen, The Limits of Machine Translation 5–15 (Dec. 23, 2009) (unpublished Master thesis, University of Copenhagen), *available at* http://www.math.ku.dk/~m01mwm/The%20Limits%20of%20Machine%20Translation%20%28Dec.%2023,%202009%29.pdf.

47. *Id.*

48. *See* ENCYCLOPEDIA OF MACHINE LEARNING 912–13 (Claude Sammut & Geoffrey I. Webb eds., 2011).

form.[49] While these documents were originally created for other purposes, researchers have been able to harness this existing corpus of data to improve automated translation. Using statistical correlations and a huge body of carefully translated data, automated algorithms are able to create sophisticated statistical models about the likely meaning of phrases, and are able to produce automated translations that are quite good.[50] Importantly, the algorithms that produce the automated translations do not have any deep conception of the words that they are translating, nor are they programmed to understand the meaning and context of the language in the way a human translator might. Rather, these algorithms are able to use statistical proxies extracted from large amounts of previously translated documents to produce useful translations without actually engaging in the deeper substance of the language.

While this automated translation often falls short of expert human translations in terms of accuracy and nuance in many contexts, and may not be sufficient for tasks requiring high degrees of accuracy (e.g., translating legal contracts), the interesting point is that for many other purposes, the level of accuracy achieved by automated translation may be perfectly sufficient (e.g., getting a rough idea of the contents of a foreign web page).[51] Such automation has allowed for approximate but useful translations in many contexts where no translation was previously available at all.

In sum, the translation example illustrates a larger strategy that has proven successful in recent AI automation: applying machine learning analysis to large bodies of existing data in order to extract subtle but useful patterns that can be employed to automate certain complex tasks. Such pattern detection over large amounts of data can be used to create complex, nuanced computer models that can be brought to bear on problems that were previously intractable under earlier manual approaches to automation.

---

49. *See Find Out How Our Translations Are Created*, GOOGLE, http://translate.google.com/about (last visited Feb. 24, 2014).

50. *See id.*

51. *See* Madsen, *supra* note 46, at 10 (citing *Google Translate FAQ*, GOOGLE, http://www.google.com/-intl/en/help/faq_translation.html (last visited Mar. 25, 2009)).

III. MACHINE LEARNING AND LAW

A. Machine Learning Applied to Law

Because machine learning has been successfully employed in a number of complex areas previously thought to be exclusively in the domain of human intelligence, this question is posed: to what extent might these techniques be applied within the practice of law?[52] We have seen that machine learning algorithms are often able to build useful computer models of complex phenomena frequently by detecting patterns and inferring rules from data. More generally, we have seen that machine learning techniques have often been able to produce "intelligent" results in complex, abstract tasks, often not by engaging directly with the underlying conceptual substance of the information, but indirectly, by detecting proxies and patterns in data that lead to useful results. Using these principles, this Part suggests that there are a subset of legal tasks often performed manually today by attorneys, which are potentially partially automatable given techniques such as machine learning, provided the limitations are understood and accounted for.

I emphasize that these tasks may be *partially* automatable, because often the goal of such automation is not to replace an attorney, but rather, to act as a complement, for example in filtering likely irrelevant data to help make an attorney more efficient. Such a dynamic is discussed below in the case of automation in litigation discovery document review. There, the machine learning algorithms are not used to replace (nor are they currently capable of replacing) crucial attorney tasks such as of determining whether certain ambiguous documents are relevant under uncertain law, or will have significant strategic value in litigation. Rather, in many cases, the algorithms may be able to reliably filter out large swathes of documents that are likely to be irrelevant so that the attorney does not have to waste limited cognitive resources analyzing them. Additionally, these algorithms can highlight certain potentially relevant documents for increased attorney attention. In this sense, the algorithm does not replace the attorney but rather automates certain typical "easy-cases" so that the attorney's cognitive efforts and time can be conserved for those tasks likely to actually require higher-order legal skills.

There are particular tasks for which machine learning algorithms are

52. This is not to say that other AI techniques will not have an impact on the law. As I have written elsewhere, logic-based AI is impacting legal domains such as contracting. *See generally* Surden, *supra* note 5.

better suited than others. By generalizing about the type of tasks that machine learning algorithms perform particularly well, we can extrapolate about where such algorithms may be able to impact legal practice.

B. Predictive Models

1. Legal Predictions

Machine learning algorithms have been successfully used to generate predictive models of certain phenomena. Some of these predictive capabilities might be useful within the practice of law.[53]

The ability to make informed and useful predictions about potential legal outcomes and liability is one of the primary skills of lawyering.[54] Lawyers are routinely called upon to make predictions in a variety of legal settings. In a typical scenario, a client may provide the lawyer with a legal problem involving a complex set of facts and goals.[55] A lawyer might employ a combination of judgment, experience, and knowledge of the law to make reasoned predictions about the likelihood of outcomes on particular legal issues or on overall issue of liability, often in contexts of considerable legal and factual uncertainty.[56] On the basis of these predictions and other factors, the lawyer might counsel the client about recommended courses of action.

The ability to generally assess the likelihood of legal outcomes and relative levels of risk of liability in environments of considerable legal and factual uncertainty is one of the primary value-added functions of a good lawyer. As a general matter, attorneys produce such estimations by employing professional judgment, knowledge, experience, training, reasoning and utilizing other cognitive skills and intuitions.[57] However, as Daniel Katz has written, such prediction of likely legal outcomes may be increasingly subject to automated, computer-based analysis.[58] As

---

53. STEPHEN MARSLAND, MACHINE LEARNING: AN ALGORITHMIC PERSPECTIVE 103 (2011).

54. *See, e.g.*, Tanina Rostain, *Ethics Lost: Limitations of Current Approaches to Lawyer Regulation*, 71 S. CAL. L. REV. 1273, 1281–82 (1998); Brian Z. Tamanaha, *Understanding Legal Realism*, 87 TEX. L. REV. 731, 749–52 (2009).

55. *See, e.g.*, PAUL BREST & LINDA HAMILTON KRIEGER, PROBLEM SOLVING, DECISION MAKING AND PROFESSIONAL JUDGMENT 29–30 (2010).

56. *Id.*

57. *See, e.g.*, Patrick E. Longan, *The Shot Clock Comes to Trial: Time Limits for Federal Civil Trials*, 35 ARIZ. L. REV. 663, 687 (1993) ("Lawyers with trial experience and the consequent ability to predict outcomes more accurately can charge more.").

58. Katz, *supra* note 10, at 912.

Katz notes, there is existing data that can be harnessed to better predict outcomes in legal contexts.[59] Katz suggests that the combination of human intelligence and computer-based analytics will likely prove superior to that of human analysis alone, for a variety of legal prediction tasks.[60]

This Part will sketch a simple overview of what such an approach to legal prediction, involving machine learning, might look like. In general, such a method would involve using machine learning algorithms to automatically detect patterns in data concerning past legal scenarios that could then be extrapolated to predict outcomes in future legal scenarios. Through this process, an algorithm may be able to detect useful proxies or indicia of outcomes, and general probability ranges.

One relevant technique to apply to such a process is the "supervised learning" method discussed previously.[61] As mentioned, supervised learning involves inferring associations from data that has been previously categorized by humans.[62] Where might such a data set come from? Law firms often encounter cases of the same general type and might create such an analyzable data set concerning past cases from which associations could potentially be inferred. On the basis of information from past clients and combining other relevant information such as published case decisions, firms could use machine learning algorithms to build predictive models of topics such as the likelihood of overall liability. If such automated predictive models outperform standard lawyer predictions by even a few percentage points, they could be a valuable addition to the standard legal counseling approach. Thus, by analyzing multiple examples of past client data, a machine learning algorithm might be able to identify associations between different types of case information and the likelihood of particular outcomes.

For example, imagine that a law firm that represents plaintiffs in employment law cases records key data about past client scenarios into a database. Such data might include the nature of the incident, the type of company where the incident occurred, the nature of the claim. The firm could also keep track of the different aspects of the case, including the outcome of the case, whether it settled, how much it settled for, the judge involved, the laws involved, and whether it went to trial, etc. This data set of past case information that the firm has encountered over the

59. *Id.*

60. *Id.*

61. *See* FLACH, *supra* note 8, at 16–18.

62. *Id.*

years, combined with other data such as published case decisions or private sources of data about case outcomes, would be the "training set." And similar to the spam filter example, the machine learning algorithm could be trained to study the past examples to learn the salient features that are most indicative of future outcomes. Over time, after examining sufficient examples of past client cases, a machine learning algorithm could potentially build a predictive model determining the weights of the factors that are most predictive of particular outcomes.

For example, (to oversimplify) we could envision an algorithm learning that in workplace discrimination cases in which there is a racial epithet expressed in writing in an email, there is an early defendant settlement probability of 98 percent versus a 60 percent baseline. An attorney, upon encountering these same facts, might have a similar professional intuition that early settlement is likely given these powerful facts. However, to see the information supported by data may prove a helpful guide in providing professional advice.

More usefully, such an algorithm may identify a complex mix of factors in the data associated with particular outcomes that may be hard or impossible for an attorney to detect using typical legal analysis methods. For instance, imagine that the algorithm reveals that in cases in which there are multiple hostile emails sent to an employee, if the emails are sent within a three week time period, such cases tend to be 15 percent more likely to result in liability as compared to cases in which similar hostile emails are spread out over a longer one-year period. Such a nuance in timeframe may be hard for an attorney to casually detect across cases, but can be easily revealed through data pattern analysis. As such an algorithm received more and more exemplars from the training set, it could potentially refine its internal model, finding more such useful patterns that could improve the attorney's ability to make reasoned predictions.

In sum, entities concerned with legal outcomes could, in principle, leverage data from past client scenarios and other relevant public and private data to build machine learning predictive models about future likely outcomes on particular legal issues that could complement legal counseling. In essence, this would be formalizing statistically to some extent what lawyers often do intuitively today.[63] Lawyers who see

63. This is reminiscent of the quote from great mathematician Pierre-Simon Laplace who said several hundred years ago, "The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enable us to appreciate with exactness that which accurate minds feels with a sort of instinct for which ofttimes they are unable to account." H. C. TIJMS, UNDERSTANDING PROBABILITY 3–4 (3d ed. 2012) (quoting LaPlace).

similar cases often over time develop an internal, intuitive understanding of the likely outcomes in particular cases once they factor in particular salient facts. Attorneys combine their judgment, training, reasoning, analysis, intuition, and cognition under the facts to make approximate legal predictions for their clients. To some extent, machine learning algorithms could perform a similar but complementary role, only more formally based upon analyzed data.

2. Limitations to Machine Learning Legal Predictive Models

There may be some limitations to predictive models that should be noted. Generally speaking, the goal of using machine learning is to analyze past data to develop rules that are generalizable going forward. In other words, the heuristics that an algorithm detects by analyzing *past* examples should be useful enough that they produce accurate results in *future*, never-before-seen scenarios. In the prior discussion for instance, the goal would be to analyze the data from past client scenarios, associate variables (e.g., hostile emails) with particular outcomes (e.g., increased settlement probability) in order to devise a set of heuristics that are sufficiently general that they would be predictive in cases with facts somewhat different from those in the training set. Such a learned model is thus only useful to the extent that the heuristics inferred from past cases can be extrapolated to predict novel cases.

There are some well-known problems with this type of generalization. First, a model will only be useful to the extent that the class of future cases have pertinent features in common with the prior analyzed cases in the training set.[64] In the event that future cases present unique or unusual facts compared to the past, such future distinct cases may be less predictable. In such a context, machine learning techniques may not be well suited to the job of prediction. For example, not every law firm will have a stream of cases that are sufficiently similar to one another such that past case data that has been catalogued contain elements that will be useful to predicting future outcomes. The degree of relatedness between future and past cases within a data-set is one important dimension to consider regarding the extent that machine learning predictive models will be helpful. Additionally, machine learning algorithms often require

64. There are other well-known problems with induction. Induction relies upon analyzing examples from the past to generalize about the future. However, under the so-called "Black Swan" problem, there may be never-before-seen, but salient scenarios that may arise in the future. In such an instance, a model trained upon past data may be insufficiently robust to handle rare or unforeseen future scenarios. *See, e.g.*, NASSIM NICHOLAS TALEB, THE BLACK SWAN: THE IMPACT OF THE HIGHLY IMPROBABLE 1–10 (2d ed. 2010).

a relatively large sample of past examples before robust generalizations can be inferred. To the extent that the number of examples (e.g., past case data) are too few, such an algorithm may not be able to detect patterns that are reliable predictors.

Another common problem involves overgeneralization. This is essentially the same problem known elsewhere in statistics as overfitting.[65] The general idea is that it is undesirable for a machine learning algorithm to detect patterns in the training data that are so finely tuned to the idiosyncrasies or biases in the training set such that they are not predictive of future, novel scenarios. For example, returning to the spam filter example, imagine the emails that were used as a training set happen to be systematically biased in some way: they all were sent from a data server located in Belarus. A machine learning algorithm may incorrectly infer from this biased training data that spam emails only originate from Belarus, and might incorrectly ignore spam emails from other countries. Such an inference would be accurate based upon the particular training data used, but as applied in the wider world, would produce inaccurate results because the training data was non-representative of spam emails generally.

Similarly, in the legal prediction context, the past case data upon which a machine learning algorithm is trained may be systematically biased in a way that leads to inaccurate results in future legal cases. The concern, in other words, would be relying upon an algorithm that is too attuned to the idiosyncrasies of the past case data that is being used to train a legal prediction algorithm. The algorithm may be able to detect patterns and infer rules from this training set data (e.g., examining an individual law firm's past cases), but the rules inferred may not be useful for predictive purposes, if the data from which the patterns were detected were biased in some way and not actually reflective enough of the diversity of future cases likely to appear in the real world.

A final issue worth mentioning involves capturing information in data. In general, machine learning algorithms are only as good as the data that they are given to analyze. These algorithms build internal statistical models based upon the data provided. However, in many instances in legal prediction there may be subtle factors that are highly relevant to legal prediction and that attorneys routinely employ in their professional assessments, but which may be difficult to capture in formal, analyzable data.

For example, imagine that there is an administrative board that

65. *See* RUSSELL & NORVIG, *supra* note 11, at 705.

adjudicates disciplinary cases and there has recently been a change in the board's personnel. An experienced attorney who has worked in a particular area for many years may be familiar with the board personnel and the types of cases that these individuals are and are not sympathetic to. Thus, such an attorney may make a recommendation as to a course of action to a client based upon a nuanced understanding of the board personnel and their particular inclinations. This might be the kind of information that would be available to an experienced attorney, and which is often used in legal counseling, but might be difficult to consistently and accurately capture in a data model. Consequently, a data model that does not include such hard-to-capture but predictive information may in fact produce inferior predictive results to an attorney.

Similarly, there are certain legal issues whose outcomes may turn on analyzing abstractions—such as understanding the overall public policy of a law and how it applies to a set of facts—for which there may not be any suitable data proxy. Thus, in general, if there are certain types of salient information that are both difficult to quantify in data, and whose assessment requires nuanced analysis, such important considerations may be beyond the reach of current machine learning predictive techniques.

C. Finding Hidden Relationships in Data

Machine learning techniques are also useful for discovering hidden relationships in existing data that may otherwise be difficult to detect. Using the earlier example, attorneys could potentially use machine learning to highlight useful unknown information that exists within their current data but which is obscured due to complexity. For example, consider a law firm that tracks client and outcome data in tort cases over the span of several years. A machine learning algorithm might detect subtle but important correlations that might go unnoticed through typical attorney analysis of case information. Imagine, for instance, that the algorithm detects that the probability of an early settlement is meaningfully higher when the defendant sued in a personal injury case is a hospital as compared to other types of defendants. This is the type of relationship that a machine learning algorithm might detect, and which may be relevant to legal practice, but might be subtle enough that it might escape notice absent data analysis.

In general, the mining of the law firm's existing data may give attorneys new information about important factors affecting outcomes (such as the category of the defendant as a hospital) that may otherwise escape traditional professional analysis. This represents a departure from

the normal mode of legal assessment of information. Attorneys typically rely upon internal intuition and previous experience to determine the factors that tend to be relevant to particular outcomes in particular instances. Machine learning as a technique—since it excels at ferreting out correlations—may help to supplement the attorney intuitions and highlight salient factors that might otherwise escape notice. The discovery of such embedded information, combined with traditional attorney analysis, could potentially impact and improve the actual advice given to clients.

1. Judicial Decisions and Data Relationships

There are some other potentially profound applications of machine learning models that can reveal non-obvious relationships, particularly in the analysis of legal opinions. A basis of the United States common law system is that judges are generally required to explain their decisions. Judges often issue major legal judgments in written opinions and orders.[66] In such a written document, judges typically explain why they decided the way that they did by referencing the law, facts, public policy, and other considerations upon which the outcome was based.[67]

Implicit in such a system of written opinions is the following premise: that the judge actually reached the outcome that she did for the reasons stated in the opinion. In other words, the justifications that a judge explicitly expresses in a written opinion should generally correspond to that judge's actual motivations for reaching a given outcome. Correspondingly, written legal decisions should not commonly and primarily occur for reasons other than those that were expressly stated and articulated to the public. At least one reason why legal opinions that do not reflect actual judicial motivations are undesirable is that there are thought to be certain motivations that are thought to be improper, illegal, or unseemly. For example, legal decisions based upon racial animus are illegal, and legal outcomes driven by pure partisanship over substance may be perceived as unseemly or improper. Moreover, it is desirable that stated judicial rationales correspond with actual rationales, because in a common law system, societal actors (and lawyers) rely upon legal opinions, and the stated justifications for these decisions, to make predictions about future legal outcomes and to understand and comply with the law.

---

66. Jonathan R. Macey, *Promoting Public-Regarding Legislation Through Statutory Interpretation: An Interest Group Model*, 86 COLUM. L. REV. 223, 253–54 (1986).

67. *Id.*

Since machine learning algorithms can be very good at detecting hard to observe relationships between data, it may be possible to detect obscured associations between certain variables in legal cases and particular legal outcomes. It would be a profound result if machine learning brought forth evidence suggesting that judges were commonly basing their decisions upon considerations other than their stated rationales. Dynamically analyzed data could call into question whether certain legal outcomes were driven by factors different from those that were expressed in the language of an opinion.

An earlier research project illustrated a related point. In that project, Theodore Ruger, Andrew Martin, and collaborators built a statistical model of Supreme Court outcomes based upon various factors including the political orientation of the lower opinion (i.e. liberal or conservative) and the circuit of origin of the appeal.[68] Not only did the statistical model outperform several experts in terms of predicting Supreme Court outcomes, it also highlighted relationships in the underlying data that may not have been fully understood previously.[69]

For example, the Supreme Court hears appellate cases originating from many different appellate circuits. Many experts had deemed the circuit of origin (e.g., Ninth or Sixth Circuits) of such a lower opinion as less important than other factors (e.g., the substantive law of the case) in relating to particular outcomes. However, the analysis of the data showed a stronger correlation between the circuit of origin and the outcome than most experts had expected based upon their intuition and judgment.[70] Although this earlier project did not involve machine learning algorithms in particular, it did involve some similar statistical techniques that might be used in a machine learning approach.

That project illustrates a basic point: that statistically analyzing decisions might bring to light correlations that could undermine basic assumptions within the legal system. If, for example, data analysis highlights that the opinions are highly correlated with a factor unrelated to the reasons articulated in the written opinions, it might lessen the legitimacy of stated opinions.[71] It also demonstrates the more general

---

68. Andrew D. Martin et al., *Competing Approaches to Predicting Supreme Court Decision Making*, 2 PERSP. ON POL. 761, 761–68 (2004); *see also* Theodore W. Ruger et al., *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decision-Making*, 104 COLUM. L. REV. 1150, 1151–59 (2004).

69. Martin, *supra* note 68, at 761–68.

70. *Id.*

71. To be clear, this is not to suggest that correlation implies causation. It is perfectly consistent for Supreme Court decisions to be correlated with a non-substantive factor (e.g. circuit of origin) and still be based upon substantive determination. Thus, for example, if one circuit court was

point that statistical heuristics can be predictive and informative in a domain as abstract and full of uncertainty as law, even when computers do not actually engage with the underlying legal substance (e.g., underlying meaning and goals of the laws, doctrines, or policies) that is typically the primary focus of attorneys.

D. Document Classification and Clustering

The practice of law is intertwined with the production, analysis, and organization of text documents. These include written legal opinions, discovery documents, contracts, briefs, and many other types of written legal papers. Outside of law, machine learning algorithms have proven useful in automatically organizing, grouping, and analyzing documents for a number of tasks.[72] This Subpart will explore two machine learning methods that may be relevant to the automated analysis and organization of legal documents: 1) document classification; and 2) document clustering.

1. Automated Document Classification

In a document classification task, the goal of a machine learning algorithm is to automatically sort a given document into a particular, pre-defined category.[73] Often such classification is based upon the document's text and other document features.[74]

The earlier spam filtering example illustrated the idea of such an automated document classification. We can think of the machine learning algorithm described as attempting to classify a given incoming email document into one of two categories: unwanted spam or wanted email. The algorithm was able to make such automatic classifications based upon the various indicia of spam emails that it had automatically detected from past examples of spam (e.g., text included "Earn Extra Cash" or country of origin was Belarus). Moreover, the algorithm was able to "learn"—refine its internal model of the characteristics of spam emails as it examined more examples of spam—and improve in its classification ability over time as its internal model and rule-set of spam

---

consistently making errors in its interpretation of the law, one outcome (reversed) might be highly correlated with a particular circuit, but that outcome would not necessarily mean that the decision was being made based upon considering the circuit of origin.

72. SEGARAN, *supra* note 23, at 6–9.

73. *See, e.g.*, Kevin D. Ashley & Stefanie Brüninghaus, *Automatically Classifying Case Texts and Predicting Outcomes*, 17 ARTIFICIAL INTELLIGENCE & L. 125, 125–65 (2009).

74. *Id.*

became more sophisticated. Thus, we consider such a task to be "classification" because a human user, examining an email, is essentially performing the same classification task—deciding whether a particular incoming email is or is not in the category "spam."

Within law, there are numerous similar tasks that can be thought of as document classification problems. For these, machine learning algorithms may be useful, and in some cases have already been deployed.

2. Classification of Litigation Docket Documents

Since about 2002, documents associated with lawsuits have been typically contained in online, electronically accessible websites such as the Federal "PACER" court records system.[75] Such core documents associated with a lawsuit might include the complaint, multiple party motions and briefs, and the orders and judgments issued by the court. In a complicated court case, there may be several hundred documents associated with the case. However, obscured within such collections of hundreds litigation docket documents, there may be a few especially important documents—such as the active, amended complaint—that might be crucial to access, but difficult to locate manually. Electronic court dockets can become very lengthy, up to several hundred entries long. A particular important document—such as the active, amended complaint—may be located, for example, at entry 146 out of 300. Finding such an important document within a larger collection of less important docket entries often can be difficult.

The task of finding and organizing core case documents can be thought of as a document classification task. Analogous to the spam filtering example, a machine learning algorithm may be trained to learn the telltale characteristics that indicate that a particular document is a complaint rather than, say, a party motion. Such an algorithm could be trained to automate classifications of the documents based upon features such as the document text and other meta information such as the descriptive comments from the clerk of the court. Thus, key electronic court documents could be automatically identified as "complaints," "motions," or "orders," by machine learning algorithms, and parties could more easily to locate important docket documents thanks to such

75. *See* Administrative Office of the U.S. Courts, *25 Years Later, PACER, Electronic Filing Continue to Change Courts*, THE THIRD BRANCH NEWS (Dec. 9, 2013), http://news.uscourts.gov/25-years-later-pacer-electronic-filing-continue-change-courts; Amanda Conley et al., *Sustaining Privacy and Open Justice in the Transition to Online Court Records: A Multidisciplinary Inquiry*, 71 MD. L. REV. 772 (2012).

automated classification.

Projects such as the Stanford Intellectual Property Litigation Clearinghouse have employed similar machine learning techniques in order to automate the organization of very lengthy and complex case dockets, and to ease the finding of crucial court documents.[76] More broadly, machine learning algorithms are capable of providing intelligent classification of documents to aid in overall organization.

3. E-Discovery and Document Classification

Similarly, certain aspects of litigation discovery can be thought of as a document classification problem. In litigation discovery, each party is often presented with a voluminous trove of documents, including emails, memos, and other internal documents that may be relevant to the law and the facts at hand. A crucial task is sorting through such discovery documents in order to find those few that are actually relevant to some issue at hand. Thus, for example, in a case involving securities fraud, certain crucial emails demonstrating the intent to defraud may be extremely crucial to proving an element of the law. The major problem is that in modern litigation, the number of documents presented during discovery can be enormous, ranging from the tens of thousands to the millions.

Only an extremely small fraction of these documents are likely to be relevant to the issue or case at hand. In some sense, the task is akin to finding the proverbial needle (e.g., smoking-gun email) in the haystack (e.g., trove of millions of discovery documents). This task can be thought of as a classification task, as the goal is to classify each of the documents into a few categories based upon relevance, such as (for simplicity's sake), highly relevant, possibly relevant, likely irrelevant, highly irrelevant.

Previously, much of this discovery was conducted manually by junior associates who pored over and read emails and used their judgment to classify emails and other documents as either likely relevant or non-relevant.[77] In essence, this is similar to the classification task described above. The major difference is that the classification of an email as spam or not spam is often a dichotomous, binary classification—an email either is or is not spam. By contrast, the classification of a given

---

76. *Stanford IP Litigation Clearinghouse*, STAN. L. SCH., http://www.law.stanford.edu/organizations/programs-and-centers/stanford-ip-litigation-clearinghouse (last visited Jan. 27, 2014).

77. *See, e.g.*, John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N. Y. TIMES (Mar. 4, 2011), http://www.nytimes.com/2011/03/05/science/05legal.html.

litigation discovery document as either relevant or non-relevant often exists upon a continuum of judgment. Some documents may be somewhat relevant, others highly relevant, and some not relevant at all. It is in this latter category that automation has proven highly useful.[78]

Today, certain aspects of litigation discovery are being automated in part, often by machine learning algorithms. Similar to the categorization tasks discussed before, in some cases, algorithms can roughly categorize documents by likelihood of relevance (often referred to as "predictive coding" or "technology assisted review"). In particular, they may be able to filter out documents that are likely irrelevant based upon dates or the parties involved. For example, such an algorithm may infer that emails that predated the core incident in the lawsuit by two years are highly likely to be irrelevant. There are, however, limitations to what these automated techniques can do. As discussed, the algorithms are not well suited to, or intended to, apply legal judgment in nuanced, uncertain areas. Rather, in many cases, the algorithms perform the role of filtering down the size of the document stack that is ultimately in need of lawyerly review. Once flagged, many of the documents still require attorney attention in order to conduct legal analysis as to relevance or privilege.

4. Clustering and Grouping of Related Documents

In a previous example, the machine learning algorithm described was used to classify documents into well-understood, predefined categories, such as "complaints," "motions," or "orders." In some cases, however, documents may have features in common, but the uniting characteristics of the documents may be unknown or non-obvious. In such an instance where there are hidden or unknown commonalities among items such as documents, a machine learning approach known as "clustering" may be useful.[79]

In clustering, a machine learning algorithm attempts to automatically group items that are similar in some way on the basis of some common characteristic that the algorithm has detected. In other words, the algorithm attempts to automatically detect hidden or non-obvious relationships between documents that would not otherwise be easily discoverable, and group such related documents together.

---

78. *See, e.g.*, Vincent Syracuse et al., *E-Discovery: Effects of Automated Technologies on Electronic Document Preservation and Review Obligations*, INSIDE COUNSEL (Dec. 18, 2012), http://m.insidecounsel.com/2012/12/18/e-discovery-effects-of-automated-technologies-on-e.

79. *See* RUI XU & DON WUNSCH, CLUSTERING 2–6 (2008).

In this way, such a machine algorithm might be used to discover that seemingly unconnected legal documents are actually related to one another in essential or useful ways. For example, imagine that there are two legal opinions in two fundamentally different areas of law: family law and trademark law. Imagine further that the two opinions share some subtle underlying commonality, such a lengthy discussion of best-practice strategies in administrative law. Such a connection between these two cases may go undetected by attorneys, since practitioners of family law may be unlikely to read trademark law opinions, and vice-versa. However, a clustering algorithm may be able to automatically find such an association and group the documents through this non-obvious relationship, by detecting a pattern among a large set of data—all opinions.

Consider another example in which automated document clustering and grouping might have uses within law. In patent law, patent examiners and patent attorneys spend a great deal of effort trying to find published documents describing inventions that are similar to a given patent.[80] Patent law has a requirement, for example, that the patent office not issue a patent on a patent application if the claimed invention is not new.[81] The way that one determines that an applied-for invention is not new is by finding "prior art" documents, which are documents that describe the invention but predate the patent application. Such prior art typically consists of earlier published scientific journal articles, patents, or patent applications that indicate than the invention had been created previously.

Given the huge volumes of published patents and scientific journals, it is a difficult task to find those particular prior art documents in the wider world that would prove that an invention was invented earlier. The task of finding such a document is essentially a problem involving automatically determining a relationship between the patent application and the earlier prior art document. Machine learning document clustering may potentially be used to help make the search for related prior art documents more automated and efficient by grouping documents that are related to the patent application at hand. More generally, automated document clustering might be useful in other areas of the law in which finding relevant documents among large collections is crucial.

---

80. JANICE M. MUELLER, PATENT LAW 30–40 (4th ed. 2012).

81. 35 U.S.C. § 102(a) (2006 & Supp. V 2011).

CONCLUSION

This Article focused upon a computer science approach known as machine learning and its potential impact upon legal practice. There has been a general view that because current AI technology cannot match the abstract analysis and higher-order cognitive abilities routinely displayed by trained attorneys, current AI techniques may have little impact upon law, barring significant technological advances. However, this Article has argued that outside of law, AI techniques—particularly machine learning—have been successfully applied to problems that had been traditionally thought to require human cognition.

This Article suggested that similarly, there are a number of tasks within the law for which the statistical assessments within the ambit of current machine learning techniques are likely to be impactful despite the inability to technologically replicate the higher-order cognition traditionally called upon by attorneys. The general insight is that statistical and other heuristic-based automated assessments of data can sometimes produce automated results in complex tasks that, while potentially less accurate than results produced by human cognitive processes, can actually be sufficiently accurate for certain purposes that do not demand extremely high levels of precision and accuracy.