The benefits and dangers of using machine learning to support making legal predictions

Received: 23 September 2022 | Revised: 19 April 2023 | Accepted: 21 April 2023

DOI: 10.1002/widm.1505

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

The benefits and dangers of using machine learning to support making legal predictions

John Zeleznikow[1,2]

[1]Research Unit of Excellence Digital Society: Security and Protection of Rights, University of Granada, Granada, Spain

[2]Law and Technology Group, Law School, La Trobe University, Bundoora, Victoria, Australia

Correspondence
John Zeleznikow, Research Unit of Excellence Digital Society: Security and Protection of Rights, University of Granada, Granada, Spain.
Email: john.zeleznikow@vu.edu.au

Edited by: Witold Pedrycz, Editor-in-Chief

Abstract

Rule-based systems have been used in the legal domain since the 1970s. Save for rare exceptions, machine learning has only recently been used. But why this delay? We investigate the appropriate use of machine learning to support and make legal predictions. To do so, we need to examine the appropriate use of data in global legal domains—including in common law, civil law, and hybrid jurisdictions. The use of various forms of Artificial Intelligence, including rule-based reasoning, case-based reasoning and machine learning in law requires an understanding of jurisprudential theories. We will see that the use of machine learning is particularly appropriate for non-professionals: in particular self-represented litigants or those relying upon legal aid services. The primary use of machine learning to support decision-making in legal domains has been in criminal detection, financial domains, and sentencing. The use in these areas has led to concerns that the inappropriate use of Artificial Intelligence leads to biased decision making. This requires us to examine concerns about governance and ethics. Ethical concerns can be minimized by providing enhanced explanation, choosing appropriate data to be used, appropriately cleaning that data, and having human reviews of any decisions.

This article is categorized under:

Commercial, Legal, and Ethical Issues > Legal Issues

Commercial, Legal, and Ethical Issues > Fairness in Data Mining

KEYWORDS

civil law, common law, ethics, legal data, machine learning

1 | INTRODUCTION

Until recently, the principal use of Artificial Intelligence for providing assistance for legal decision-makers has been in areas of Statutory Interpretation, including the burgeoning rules as code movement.[1] Except in academic circles there has been little focus upon the use of machine learning in law and the appropriate use of data in law.

In 2019, the Victoria Law Foundation examined how the administrators of Victoria, Australia's civil justice system used what the Foundation viewed as administrative data.[2] They examined what data was available; the accuracy and consistency of such relevant data; how that data is currently used; and how to improve the quality of the data being used so that the data can be appropriately used to provide answers relating to various access to justice questions (McDonald et al., 2021). In this paper, we do not discuss administrative data. Instead, we focus upon that data appearing in reports of conflicts and which is specifically used to draw legal conclusions.

The first Artificial Intelligence (AI) systems were constructed in the 1960s and used rule-based reasoning. Case-based reasoning followed in the 1980s, followed by rule induction and machine learning. Slightly earlier, more numerically based techniques (but less cognitive) were developed, including game theory (Nash, 1953), operations research and statistics. In their construction of a three-step model for Online Dispute Resolution (ODR), Lodder and Zeleznikow (2005) argued that Artificial Intelligence involves the study of all forms of automated human intelligence and thus should include numerically based techniques.

Rajkomar et al. (2019) claimed that a major issue for developers when constructing a machine-learning model is using a diverse, non-contradictory and representative data set. Schmitz and Zeleznikow (2021) claim that it is possible to construct such data-sets in medicine, but is alas is a far more difficult task in law. Much more and cleaner data is available in medicine than is the case in legal domains. Thus, the use of machine learning in law is unlikely to ever rival its use in medicine.

Surden (2020) claims that "machine learning refers to a category of AI approaches in which algorithms automatically learn patterns from large amounts of data. These learned patterns can then be harnessed to automate tasks. Machine learning can only be applied if sufficient suitable data are available. In legal domains, machine learning is having an important impact on prediction, the automated examination of legal documents, and the analysis of legal contexts." In this paper, we will focus upon the appropriate use of machine learning to assist professionals to make legal predictions.

As Zeleznikow and Hunter (1994) say, in the rule-based approach, the knowledge of a specific legal domain is represented as a collection of rules of the form IF <condition(s)> THEN <action>. About a decade after Weizenbaum (1966) developed the ELIZA system which modeled the reasoning of a psychiatrist, the earliest Legal Expert Systems were developed. Such systems included TAXMAN (McCarty, 1977) and the British Nationality Act as a Logic Program (Sergot et al., 1986).

Zeleznikow and Hunter further claim that "Case-based reasoning is the process of using previous experience to analyze or solve a new problem, explain why previous experiences are or are not similar to the present problem being studied and adapting past solutions to meet the requirements of the current problem." Ashley (2004) argues that precedents play a more central role in Common law countries than they do in countries based on Civil Law. Hence it is more appropriate to use adversarial case-based reasoning in Common Law domains.

The electronic availability of case law, especially via the internet, has led the use of precedents becoming more important in countries whose legal system is based upon Civil Law. Further, Rissland et al. (2005) claim that case-based reasoning is most appropriate for modeling legal reasoning because Common law is precedent-based as its judicial standard, stare decisis,[3] mandates that similar cases should be decided similarly. A comprehensive discussion of the application of the use of Case-based Reasoning in the legal domain is provided in Ashley (1992) and Rissland et al. (2005).

While Common Law focuses on reasoning from case to case, it does not explicitly indicate how to determine "similarity." Similarity is variable—depending on one's viewpoint and the desired outcome that the user requires. Thus common law is a useful domain for exploring reasoning with cases and for studying issues about indexing and retrieval, similarity assessment, and case comparison. Rissland et al. (2003) indicate that there are many other characteristics that result in case-based reasoning being useful in Common Law legal domains: "(1) the open-textured nature of legal concepts[4]; (2) a variety of task orientations (e.g., advocacy, adjudication, advising); (3) diverse categories of knowledge (e.g., cases, statutes, regulations) and modalities of reasoning (e.g., case-based, rule-based, hybrid); (4) significant repositories of richly linked case knowledge; and (5) its use of hypotheticals in reasoning."

Fayyad et al. (1996) claim that "Machine learning is that subsection of learning in which the artificial intelligence system attempts to learn automatically. Knowledge Discovery from Databases is the non-trivial extraction of implicit, previously unknown and potentially useful information from data. and Data mining is a problem-solving methodology that finds a logical or mathematical description, eventually of a complex nature, of patterns and regularities in a set of data." Stranieri and Zeleznikow (2005) provide a detailed discussion of Knowledge Discovery from Legal Databases.

2 | HOW LEGAL DATA DIFFERS FROM OTHER DATA

Legal data are by its nature very different from other data, especially medical data. Legal data are not as precise as medical data and often need to be transformed so that they can be gainfully used when making decisions.

Stevens (1946) introduced the notion of levels of measurement or scales of measurement of data. He classified data into four types: nominal, ordinal, interval, and ratio types. A nominal scale does not have any natural order or ranking. An example might be classifying what college a student attended, for example, Harvard.

Ordinal data allows ranking. An example might be classifying the highest level of education achieved by a student, for example, a PhD is ranked higher than a Master's degree which is ranked higher than a Bachelor's degree.

An interval scale allows for order and the difference between two values is meaningful. For example, 20°C is warmer than 10°C, but it is not a measure of twice as warm.

A ratio scale has all the properties of an interval scale, with a clear definition of 0 as a starting point for calculations.[5] An example of this scale is weight: a man who weighs 120 kg is twice as heavy as a man who weighs 60 kg.

Legal data are invariably of the nominal form. Further, such data are often imprecise and based upon value judgments rather than measurable facts. As an example, let us consider how the United States Supreme Court dealt with the issue of whether racially segregated schools are allowed under the US Constitution.

The two relevant United States landmark cases that deal with this issue are Plessy v. Ferguson[6] and Brown v. Board of Education of Topeka.[7] During the US Civil War (1861–1865), the 13th and 14th amendments to the US constitution were ratified.[8]

In 1896, in the case of Plessy v. Ferguson[9] the United States Supreme Court ruled that the demands of the 14th Amendment to the US Constitution were satisfied if the states provided separate but equal facilities for all its citizens and the fact of segregation alone did not make facilities automatically unequal. In 1954, in Brown v. Board of Education of Topeka[10] the Supreme Court seemingly overturned the decision made in Plessy v. Ferguson. In Brown v. Board of Education of Topeka, in the opinion of Black (1990), the US Supreme Court declared racial segregation in public schools to be in violation of the equal protection clause of the 14th Amendment. The Supreme Court did so, not by overturning the ruling in Plesy v. Ferguson[11] but by using sociological evidence to show that racially segregated schools can never be equal and thus contravened the 14th amendment (Dworkin, 1986).

What we can observe from this example is that legal decision-making uses data in a very different manner than occurs with the use of data in more scientific positivist domains. For many such reasons, until now, the influence of machine learning for legal decision-making in common law countries has been considerably less significant than its use in other professional domains.

3 | THE USE OF DATA IN DIFFERENT LEGAL DOMAINS

3.1 | Contrasting the use of data in common law and civil law domains

As we have observed previously, machine learning is used differently in Common Law domains from the way that it is used in Civil Law domains. So, we first consider the differences in how data is used in the two domains.

Stranieri and Zeleznikow (2005) state that "Common law is the legal tradition that evolved in England from the Norman invasion of 1066 onwards. The principles behind Common Law appear in reported judgments, usually of the higher courts. Common law is usually much more detailed in its prescriptions than in civil law domains. Common law is the foundation of private law, in England, Wales, and Ireland,[12] but also in 49 U.S. states,[13] nine Canadian provinces,[14] Australia and New Zealand and in most countries that that were British colonies."

Tetley (1999) views "Civil law is that legal tradition that originated in Roman law and then slowly developed in much of Continental Europe. It can be viewed as dividing into two streams: the codified Roman law (which is primarily seen in the French Napoleonic Civil Code of 1804 and is followed in most of Continental Europe, Louisiana and Québec); and uncodified Roman law (which can be seen in Scotland {which also integrates UK Common law} and South Africa)." Tetley claims that Civil law is highly systematized and structured[15] and relies on declarations of broad, general principles, often ignoring the details.

Statutory law, or law found in legislation made by relevant parliaments, is common to both civil and common law jurisdictions. In civil law jurisdictions, the important legal principles are explicitly stated in the code or legislation. In

common law jurisdictions, most of the accepted rules are found in the jurisprudence and statutes complement these rules.

David and Brierly (1985) observe that both common law and civil law legal traditions share similar social objectives (individualism, liberalism and personal rights). Given the differences between the Common Law and Civil Law codes, it seems sensible to ask whether there is any practical value to be gained from using case-based models of legal reasoning in a civil law context? In Ashley (2004), Kevin Ashley considered this issue and claimed that judges in civil law jurisdictions do indeed reason with legal cases, but differently from their common law counterparts.

Stranieri and Zeleznikow (2005) argue that while case-based reasoning is very well suited for modeling reasoning in common law, Knowledge Discovery from Databases (KDD) is appropriate for use in both common law and civil law domains. They justified this claim by stating that in legal domains, KDD can be used to understand how judges exercise discretion. They argue that even in civil law countries, where there is an emphasis on codifying the law, there still exist what Black (1990) defines as discretionary acts.[16] They further claim that use of KDD in civil law countries can help support consistency and transparency in decision-making.

We thus need to investigate how different AI strategies are used to conduct legal interpretation in each of Common Law and Civil Law domains. As we shall see rule-based reasoning has proved very use for statutory interpretation, while case-based reasoning is very useful for modeling analogical reasoning in Common Law domains. But can Machine Learning be useful and if so, for what legal domains?

3.2 | Modeling legal discretion

In his book on jurisprudence, Fletcher (1996) makes some significant points on using rules to model legal domains. He claims "that the two great vices of contemporary jurisprudence are to:

1. Believe either that rules dictate results as do computer algorithms (as do positivists or formalists) or
2. That rules have no bearing at all on the adjudication of disputes (as do legal realists or skeptics)."

He argues that all legal problems lie between these two extremes.

Kannai et al. (2007) claim that to model discretionary reasoning requires the developer to carefully analyze the relevant exercise of discretion and to then use a modeling technique appropriate for the task. Galigan (1986) claims that administrative discretionary powers are allocated to authorities, who exercise their powers to achieve their goals efficiently and effectively. Administrative discretion is exercised only when authorized by law. There has been extensive research in the development of AI to model administrative discretion, including the seminal work of Sergot et al. (1986) which attempted to interpret provisions of the British Nationality Act of 1981.

In attempting to model discretionary legal reasoning, Kannai et al. (2007) developed an eight-octant approach for classifying legal tasks. The following summarizes their approach.

To conduct such modeling, they introduced the notion of three axes—(a) bounded-unbounded axis; (b) well-defined-undefined axis; and (c) bounded-continuous axis.

a. A legal task is bounded if all issues involved in resolving the task are known (Schild et al. 1999). The concept of a bounded task is well understood in computer science.
b. A task is well-defined if all underlying predicates are measurable, and it is known how the predicates interrelate. For instance, most social welfare benefits are well-defined (because there are well known rules to determine entitlement) while child welfare issues are undefined (because there are few or indeed no rules to help determine the task). As an example, currently, to receive an aged person's pension in Australia, a man must be at least 67 years of age and have income below a certain threshold (as measured by the applicant's income tax return for the preceding financial year).

   A task which is bounded but undefined cannot be modeled by rules. The underlying issues are known, but there is no knowledge about how the issues should be combined. In common law domains, case law can be used when the legislation fails to provide norms for deciding new cases. Landmark cases provide beacons for which future cases can be decided.[17] Since argumentation is by analogy, case-based reasoning is very appropriate for modeling legal domains in the bounded-undefined quadrant as long as sufficient landmark cases exist. Commonplace cases do not establish any norms.[18] In such domains, Knowledge Discovery from Databases can be used to model the domain. A

concrete example of a Bounded Undefined domain is the domain of criminal sentencing in England in the 1980s and 1990s. The domain involves sentencing guidelines that explicitly state what issues are relevant, without providing rules about how the issues should be combined to reach a decision. These sentencing guidelines could in principle form a (partial) database for a case-based system. Another approach to modeling Bounded Undefined was developed by Stranieri et al. (1999), who used machine learning to model such reasoning.

c. In some fields of law there are only two possible outcomes. One such example is the determination of applications for refugee status in Australia. The granting of refugee status, decision is an "all or nothing decision." Either applicants are granted refugee status, or, after a hearing, they are either returned to their country of origin or a third country. It is rare for countries to allow fixed-term refugee status. Thus, the decision of refugee status, can be a life-or-death matter. This can be seen when an application for refugee status is refused, and the applicant is returned to his country of origin where he is subsequently executed.

When comparing and contrasting child welfare and Refugee Law decision-making as discretionary tasks, Kannai et al. (2007) noted the differences in the way discretion is exercised between tasks in which there are only binary choices, and others in which there are multitudes of choices. For example, Stranieri et al. (1999) noted that Australian Family Court judges have essentially 13 possible outcomes for the distribution of marital property following divorce. As a result, when making such a decision, judges have more flexibility than do refugee review tribunal members and the magnitude of an "error" in Family Court property distribution is less than the magnitude of an error in deciding refugee status.

Theoretically, when making decisions about the residency of children following divorce in Australia, Family Court judges can choose from infinitely many solutions. However, in practice, each parent spends between zero and 365 (or 366) nights a year with their child. This leads to the introduction of the notion of a binary-variable axis. A decision is binary if there are only two possible outcomes (such as the decision whether or not to grant refugee status, or whether to execute an appropriately convicted criminal). As one moves further along the axis, the decision-maker has progressively more choices. At the other end of the axis lies a task such as child welfare, in which there are seemingly infinitely many options available to the decision-maker.

While the availability of a continuum of decisions gives the decision-maker much discretion, the presence of binary decisions forces the decision-maker to essentially make what we can view as "harder decisions." The option the judge chooses when faced with binary decisions (such as in the cases of capital punishment or the determination of refugee status) has dire consequences for the recipient of the decision.

The ever-growing movement toward using various forms of Alternative Dispute Resolution (ADR) rather than engaging in conventional litigation[19] is a further example of the movement from binary to continuous decision-making. When negotiating a decision, disputants are in general free to choose from a variety of options. In litigation, disputants are constrained by the decisions of the judicial decision-maker.

The ability to allow disputants flexibility in resolving their affairs is one of the many evolving legal trends.

This examination of decision-making in legal domains indicates that machine learning is most appropriate for domains in the bounded, undefined, continuous octant. It can also be used in the bounded, undefined, binary octant. It is more appropriate to use Ruled-based systems in the bounded and defined octants. Further, to use machine learning requires an abundance of commonplace cases.

4 | ARTIFICIAL INTELLIGENCE REASONING IN LAW

As was discussed in Section 1, the earliest use of AI in Law, was the development of Rule-based expert systems. At that time statutes were modeled as rules. One of these earliest systems, TAXMAN (McCarty, 1977) was a logic based deductive reasoner that examined the taxation of corporate reorganizations. Sergot et al. (1986) used logic programming to check if an individual was entitled to British Citizenship under the legislation of the *British Nationality Act* 1981. Zeleznikow and Hunter (1994)[20] argue that while the system is an interesting application of logic, it is jurisprudentially flawed because it believes that law is straightforward and unambiguous.

In the early 1980s, the Rand Corporation used AI to develop numerous systems which supported settlement. These systems provided advice about risk assessment in damages claims. Lift Dispatching System (LDS) (Waterman & Peterson, 1981) assisted legal experts in settling product liability cases. SAL, the system for asbestos litigation (Waterman et al., 1986) helped insurance claims adjusters evaluate claims related to asbestos exposure. Schlobohm and

Waterman (1987) developed EPS (Estate Planning System) which was a prototype expert system that performed testamentary estate planning by interacting directly with clients or paralegal professionals.

Only recently, rule-based expert systems have been used commercially to enforce compliance. One example of a compliance system is the legislation regarding driving infringements in Victoria Australia, which was mentioned in Section 1. Further examples of compliance systems deal with taxation law and social security benefits.

It is vital to acknowledge that compliance systems must be used appropriately. In the Australian Robodebt debacle, thousands of Australians were informed that they owed money to Centrelink (Dowling, 2022). The compliance check which delivered this advice was inappropriately determined by a computer system, rather than by humans (Whiteford, 2021). In a joint submission to the Australian Royal Commission enquiry to the Robodebt scheme,[21] the Australian Society for Computers and Law and the The Allens Hub for Technology, Law and Innovation at the University of New South Wales stated "governments must address poorly designed and poorly implemented automation. There needs to be recognition and understanding that automated systems are most appropriate for dealing with prescriptive rules and, in most government decisions, there is a need for the exercise of discretion. There is a great deal of value in being able to automate the easy, routine parts of the decision to free up human arbiters to deal with the discretionary aspects of a decision-making process. But it is important to recognize that the exercise of human discretion is critical in a just society and must not be eliminated. It's also important to recognize that computers are not magic. If the policy or law is bad, all that automated decision-making will do will enable the faster application of bad laws or bad policy, and at a larger scale. A transparent, traceable decision will still deliver an unjust decision if it is the implementation of an unjust law."[22]

The Rules as Code movement develops applications to automatically apply coded norms to check whether the business processes of an enterprise comply with the relevant rules and help users to access legal information.[23] While Rules as Code may provide efficiency benefits, one can ask whether Rules as Code is constitutional, or does it appropriate, undermine or limit the role of courts to interpret the law? How authoritative is the drafter/coder's view of the meaning of the law? If a Rules as Code tool provides incorrect information—for example it advises the users erroneously that they are ineligible for a welfare payment—how will the mistake be identified and who will be liable for such mistakes?

According to Zeleznikow and Hunter (1994)[24] case-based reasoning systems were developed as a reaction to the limitations of rule-based systems. Case-based systems have many advantages for modeling law compared with rule-based systems—they can arrive at conclusions based on a number of cases, rather than using the whole body of possibly contradictory and complex rules; they can interpret open-textured concepts by using analogies; and in contrast to the limitations of rule-based systems, the more information that is stored in a case-based reasoner, the higher the potential accuracy of the reasoner.

Kevin Ashley (1992) identified five distinct case-based reasoning paradigms—statistically oriented, model-based, planning or design oriented, exemplar-based and precedent-based. In the paradigm most relevant to legal domains, the precedent-based paradigm; cases are precedents employed in arguments using analogy to justify the conclusion, as well as giving competing precedents. Two examples of such early and significant legal case-based reasoners were GREBE (Branting, 1991) and Hypo (Ashley, 1991).

HYPO was an interpretive Case based reasoning system, which performed the total sequence of case-based reasoning steps from analysis to argumentation. It analyzed a new case; retrieved relevant cases from its case base; sorted the cases according to how on-point they were; selected best cases for each side of the issue; generated "3-ply" point–counterpoint–rebuttal style arguments; and explored strengths and weaknesses of each side's arguments using hypotheticals Ashley has built many case-based legal tutoring systems (Ashley et al., 2002). Ashley (2017) discussed how CATO harnessed HYPO's model of legal argument to teach law students how to make arguments with cases.

Rissland and Skalak (1991) developed the hybrid system CABARET, which integrated Rule based reasoning and Case-based reasoning. It used a dynamic agenda-based control architecture to interleave Rule-based reasoning and Case-based reasoning, so that these two reasoning modes could complement and supplement each other.

The GREBE system pioneered the adaptation and reuse of case justifications for argument creation. GREBE's model of argument structure included both rule applications and exemplar-based explanations that related portions of case facts to the conclusions they justified. GREBE was a hybrid Rule based/Case based program that reasoned with both rules and cases. When rules were insufficient, GREBE created case analogies. It used a heuristic measure of argument strength to rank the arguments for and against a given conclusion. Zeleznikow et al. (1994) created the IKBALS III system. It combined Case based reasoning with rules induced rules from decision trees. The system augmented rule-based explanations with case examples. As will be seen in Section 5, Stranieri and Zeleznikow (2005) in the Split-Up project

combined Rule based reasoning with structured connectionist networks to predict judicial allocations of marital property in divorce cases.

5 | MACHINE LEARNING IN LAW

DoCarmo et al. (2021) argue that computational systems, including machine learning, artificial intelligence, and big data analytics, are essential aspects of modern social life. They claim that the existence of such systems is changing legal practice. The authors present a variety of cases in three domains—algorithmic governance in law, legal jurisdictions and agency. They study law in computation and discover how new technological systems' integration with legal processes pushes the distinction between "law on the books" and "law in action" into new domains. The academic use of Machine learning in law is more than 25 years old, and many of the limitations of using machine learning in law were identified at that time.

Kevin Ashley in Ashley (2019) claimed that in its earliest form, machine learning extracted legal knowledge by automatically inducing rules from decision trees or generated statistical models from data. The principal data used was judicially decided cases. The rules that were learned or the resulting models were then used to predict the results of new cases. Learned models and their relevant features, do not necessarily correspond to legal knowledge recognizable by human experts. Hence, machine learning programs cannot easily explain their predictions in language or terms acceptable to lawyers.

Surden (2014) suggests that there is a subset of legal tasks often manually performed by lawyers, which are potentially partially automatable via techniques such as machine learning; provided that we understand and account for limitations of machine learning. Surden claims that these tasks may be partially automatable, because often the goal of such automation of tasks is not to replace an attorney, but instead, to support a lawyer, as for example in filtering likely irrelevant data to help make an attorney more efficient.

Automation for litigation discovery document review is now common in litigation practice. It is the most signicant current example of the use of machine learning in law. For this task, machine learning algorithms are not used to replace crucial attorney tasks such as of determining whether specific ambiguous documents are relevant under uncertain laws or whether the documents will have significant strategic value in the proposed litigation. Often, the algorithms may be able to reliably filter out large swathes of documents that are likely to be irrelevant. As a result, the attorney does not have to expend limited cognitive resources analyzing the documents. Additionally, the algorithms can highlight certain potentially relevant documents for increased attorney attention.

Kleinberg et al. (2018) investigated whether the use of machine learning can improve human decision making. He did so by examining the domain of decision-making with regards to bail. Each year Judges in USA make millions of decisions on jail-or-release decisions that hinge upon a prediction of what a defendant would do if released. The concreteness of the prediction task combined with the volume of data available makes bail decision-making a promising application for machine-learning.

Kleinberg et al believe that making comparisons between the operation of the machine learning algorithm they developed and the manner in which the judges reason, is very complicated for the following reasons:

1. 'The available data were generated by prior decisions made by judges. Kleinberg et al observed crime outcomes for those defendants who were released, but not for those who judges detained. Thus, it is difficult to evaluate counterfactual decision rules based on algorithmic predictions.
2. Judges may have a broader set of preferences than the variable predicted by the algorithm; for instance, judges may care specifically about violent crimes or racial inequities. Kleinberg et al dealt with these problems by using different econometric strategies, such as quasi-random assignment of cases to judges.

They claim that "even taking these concerns into account, the results of the machine learning algorithm indicate potentially large welfare gains: one policy simulation showed crime reduction up to 24.7% with no change in rates of jailing, or rate reductions of jailing up to 41.9% with no increase in crime rates. All categories of crime, including violent crimes, show reductions in rates; these gains were achieved while simultaneously reducing racial disparities. These results suggest that while machine learning can be valuable, realizing this value is not automatic. It requires integrating machine leaning tools into an economic framework: being clear about the link between predictions and decisions; specifying the scope of payoff functions; and constructing unbiased decision counterfactuals."

Chen (2019) argues that predictive judicial analytics can increase fairness in law. This is despite much empirical work observing that there are significant inconsistencies in judicial behavior. By predicting judicial decisions-with more or less accuracy depending on judicial attributes or case characteristics-machine learning offers an approach to detecting when judges are most likely to allow extraneous biases to influence their decision making. He claims that low predictive accuracy may identify cases of judicial "indifference" where case characteristics (interacting with judicial attributes) do not strongly dispose a judge in favor of a particular outcome. In such cases, biases may hold greater sway, querying the fairness of the legal system.

Zeleznikow et al. (1994) went beyond developing first generation, production rule legal expert systems by integrating traditional rule-based reasoning and case-based reasoning with intelligent information retrieval. Rather than relying upon a centralized blackboard architecture, as had previously been the norm, the researchers used cooperating agents. Their resulting IKBALS system used a specialized induction algorithm to induce rules from cases. These rules were then used as indices during the case-based retrieval process.

In the same Donald Berman Laboratory for Information Technology and Law at La Trobe University, the Split-Up system (Stranieri et al., 1999), showed how the use of machine leaning, in the form of neural networks, with the assistance of rule-based reasoning, can provide automated advice upon the distribution of marital property following separation in Australia. The system could be used by those without detailed legal knowledge, thus supporting self-represented litigants and access to justice (Zeleznikow, 2002 and Schmitz & Zeleznikow, 2021).

5.1 | The split-up system: Using machine learning to make predictions

In the early 1990s, Stranieri and Zeleznikow at La Trobe University, wished to demonstrate that machine learning could be gainfully used to predict the outcomes of legal conflicts. Because of the assistance of domain experts at Victoria Legal Aid and with the support of a Family Court of Australia judge, Tony Graham, they chose to model property distribution in the domain of Australian Family Law. In their first paper, Stranieri and Zeleznikow (1992), they introduced the SPLIT-UP prototype, which reasoned with statutes and expert legal knowledge.

While the first Split-Up prototype provided useful advice, it did not meet the developers' goals. Stranieri and Zeleznikow wished to use neural networks in conjunction with other appropriate AI tools to provide significant advice. To do so they needed to develop both jurisprudential and computer science techniques to deal with:

1. What cases to use—landmark[25] or commonplace[26]?
2. What hybrid AI tools to use—rules or neural networks?
3. How to provide explanations—they decided to use argument trees.

In order to discover how Australian Family Court judges weight different factors, Stranieri et al. (1999), used as source material, written judgments handed down by judicial decision makers in commonplace cases. Stranieri's group had access to 400 family law cases stored within the Melbourne registry of the Family Court of Australia. As the focus of Split Up was solely upon property distribution and many of the 400 cases involved custody or child welfare issues in addition to property, not all cases could be used. However, expert opinion indicated that property proceedings are strongly influenced by child welfare matters.

Eventually, only 103 cases solely involved property.[27] Three raters extracted data from these cases by reading the text of the judgment and recording values of 94 template variables.[28] Inter-rater agreement tests were performed informally. Any variable that seemed ambiguous or unclear was highlighted so that a consensus could be reached between the raters.

To represent knowledge in the domain, in particular to decide when to use neural networks and when to use rules and also to provide explanations, Stranieri et al used Toulmin's (1958) theory of argumentation. In the Split Up system, data from commonplace Australian divorce case judgments was submitted to a connectionist algorithm. The algorithm learned to weight factors in the same way as judges had done in past cases, so that the outcome of future cases could be predicted.

Toulmin (1958) examined arguments from a variety of domains and concluded that all arguments, regardless of the domain, have a structure which consists of six basic invariants: "claim, data, modality, rebuttal, warrant and backing. Every argument makes an assertion based on some data. The claim is the assertion of the argument. The mechanism which is used to justify the claim is known as the warrant. The backing supports the warrant. In a legal argument the

backing is generally a reference to a statute, precedent case or even a commentary. The rebuttal component specifies an exception or any condition that negates the claim."

The Split Up system explains its reasoning behind inferring an argument's assertion by presenting the relevant data, the warrant and the backing components of the argument to the user, whenever she desires this knowledge. The approach generates explanations for conclusions which are reached quite independently of inferencing methods used to reach those conclusions.

5.2 | Machine learning used to model US supreme court cases

Many researchers have used Machine Learning to understand and model US Supreme Court decisions. Kaufman et al. (2019) argue that increasing the predictive accuracy of forecasting models of US Supreme Court decisions, allows researchers to better understand significant policy rulings. Previous attempts to develop predictive models of Supreme Court behavior have found success using either (1) text data taken from oral argument proceedings, or (2) quantitative legal data. Kaufman et al incorporated both data sets using an AdaBoost decision tree regressor.[29] They claimed that this approach substantially outperformed existing predictive models of Supreme Court outcomes which use exclusively one data source or rely on simpler modeling strategies.

Katz et al. (2014) developed and evaluated a supervised machine learning program to predict if a single US Supreme Court Justice or the Full Court will affirm or reverse a lower court's judgment. Using the randomized tree method initially proposed by Geurts et al. (2006), Katz et al. (2014) predicted 60 years of decisions by the Supreme Court of the United States (from the years 1953 to 2013). They state that the model correctly identified 69.7% of the Court's overall affirm/reverse decisions and correctly forecasted 70.9% of the votes of individual justices. The model used 7700 cases and more than 68,000 justice votes. Katz et al. claimed that their model was the first robust, generalized, and fully predictive model of US Supreme Court voting behavior.

Lex Machina is a private analytics company founded in 2010 aiming to predict the cost and outcome of intellectual property litigation. Lex Machina is based in Silicon Valley and is part of LexisNexis company.[30] According to Surdeanu et al. (2011), Lex Machina used a different (as compared with the Split Up system and the work of Katz et al described above) supervised machine learning approach to make predictions. Lex Machina did not consider the substantive merits of cases, but focused on the litigation participants and their behavior, the lawsuit parties, their attorneys and law firms, the judges assigned to a case, the districts where the complaints were filed, judicial and district "bias" (computed as the ratio of cases won by the plaintiff from the set of past cases assigned to the corresponding judge or district) and the outcomes of the cases.

Surdeanu et al. (2011) claimed that Lex Machina employed logistic regression, a statistical machine learning model, to predict the outcomes of intellectual property claims based on all Intellectual Property lawsuits in a 10+ year period. They argued that the system had an accuracy of 64% The most significant factors leading to accuracy were the judge's identity, followed by the plaintiff's law firm, the defendant's identify, the district where the claim was filed, the defendant's law firm, and the defendant's attorney.

Surdeanu et al. (2011) stated that Lex Machina's participant-and-behavior features can be extracted automatically from the texts of cases. For most features, it required identification of named entities (firms, courts, people) and checking the names against directories or lists of names. Extracting the outcomes of cases was more complex. Three Intellectual Property experts annotated sentences stating the outcomes for cases used in a training set, and a machine learning model was able to automatically learn to extract the outcomes. Surden et al emphasized that the model is "agnostic to the merits of the case". Given enough data, participant-and behavior features alone were a substitute for information about a case's merits. Lacking substantive information about the case, however, Lex Machina was unable to explain its predictions in terms that legal professionals would recognize as a legal explanation or appropriate argumentation.

In a series of papers, Bentley University professor Noah Giansiracusa used machine learning to model the decision-making of US Supreme Court Justices. In the first paper, Giansiracusa and Ricciardi (2019), he modeled such decision-making by using three different spatial voting preference models: an instance of the widely used single-peaked preferences, and two models in which vote outcomes have a strength in addition to a location. He introduced each model from a formal axiomatic perspective, discussed the practical motivation for each model in terms of judicial behavior, proved mathematical relationships among the voting coalitions compatible with each model, and then studied the two-dimensional setting by presenting computational tools for working with the models and by exploring these models with judicial voting data from the Supreme Court.

In Giansiracusa (2021), he imported methods from evolutionary biology to illuminate what he claimed is the intricate and often overlooked branching structure of US Supreme Court justices' voting behavior. He used phylogenetic tree estimation based on voting disagreement rates,[31] to extend ideal point estimation to the non-Euclidean setting of hyperbolic metrics. After introducing this framework, comparing the framework to one- and two-dimensional multidimensional scaling, and arguing that the framework flexibly captures important higher-dimensional voting behavior, Giansiracusa presented a handful of potential ways to apply the tool. The emphasis throughout this research is on interpreting these judicial trees and extracting qualitative insights from the trees.

In Giansiracusa (2022), Noah Ginsiracusa argued that ideal point estimation provides empirical legal scholars with spatial representations of the Supreme Court justices. These estimations help elucidate ideological inclinations and voting behavior. The estimation is primarily performed in one dimension, where politics dominates, though later work details a second dimension capturing differing attitudes on the authority of various legal actors.

Giansiracusa introduced and explored a network-theoretic tree-based method for visualizing the relationships between the justices. Based on Justices' voting records, the method allows scholars to study what he claims is the intricate branching structure of the Court. He showed how his tool can be used to uncover times in the Court's history where the balance on the bench fractured in unusual and interesting ways. By defining several tree-based measures and charting their evolution over time, it became clear that over the past 50 years, the Supreme Court has become increasingly one-dimensional and bipolar, dividing along political lines.

5.3 | Further examples of using machine learning for legal prediction

Medvedeva et al. (2020) used data obtained from the European Court of Human Rights to investigate how natural language processing tools can be used to analyze texts of the court proceedings. Their goal was to automatically predict (future) judicial decisions. They achieved an average accuracy of 75% in predicting the violation of nine articles of the European Convention on Human Rights. This shows how machine learning approaches can assist decision-making in legal domains. The research showed that making predictions for future cases based upon the results of past cases negatively impacts upon performance (the average accuracy range was from 58 to 68%). Medvedeva et al. (2020) also demonstrated that they could achieve a relatively high classification performance (average accuracy of 65%) when predicting outcomes based only on the surnames of the judges who try the relevant case.

In further work, Medvedeva et al. (2023) examined analyzing court decisions using computational techniques. They examined differences between forecasting decisions, categorizing judgments according to the verdict and identifying the outcome based on the text of the judgment. They illustrated that each task examined is strongly dependent on the type of data used, in line with discussions above.

Rai (2018) argued that machine learning provides a useful domain for the examination of patents. She claimed that the use of machine learning to support patent examination does not raise the same significant concerns about individual rights and discrimination that are prevalent in other areas of administrative and judicial process.

Perhaps, the most important challenge for using machine learning to support legal decision-making relates to explaining the derived decisions. The U.S. Patent and Trademark Office stresses the significance of transparency for the general public as a necessary condition for achieving adequate explanations.[32] However, full transparency diminishes the likelihood of the provision of private sector expertise and is also susceptible to gaming.

Alarie et al. (2016) introduced the Blue J Legal project which used machine learning technologies (specifically neural networks) to provide predictions in uncertain areas of Canadian tax law such as the vexed question of whether a worker is an independent contractor or an employee. Alarie et al. foresee a world where information about legal rights and responsibilities is more affordable; where the informational asymmetries that lead to wasteful expenditure on litigation is reduced, and where regulators use such tools to create a more effective and efficient administration of government. Schmitz and Zeleznikow (2021) describe such a world in which Online Dispute Resolution tools support Self Represented Litigants.

5.4 | Using machine learning in law enforcement and sentencing

There are numerous domains related to law enforcement, where machine learning decision support is highly valued, but explanation is not required. In such domains, advice is useful, but decisions to initiate action are taken by humans and not software. Examples include crime data mining and sentencing decision support systems.

McCue and Parker (2003) claim that data mining tools are gainfully used by many law enforcement organizations. Analytical tasks considered include around-the-clock crime analysis, behavioral analysis of violent crime, officer safety, risk and threat assessment, risk-based deployment and tactical crime analysis. According to Oatley (2021) data mining and decision support systems are significant in assisting human inference in crime analytics. Crime Analytics technologies are used for clustering crimes, finding links between crime and profiling offenders, generating suspects, geographical information systems, identifying criminal networks, matching crimes and predicting criminal activity, and social network analysis,

Oatley et al. (2006) examined the challenges police face when dealing with the detection and prevention of burglaries. Police focus upon the use of "soft" forensic evidence to examine the modus operandi and the temporal and geographical features of the crime, rather than "hard" evidence such as DNA or fingerprint evidence. Oatley et al focused upon the needs of crime systems users and studied the different types of data police collect and reasons why police store data in particular ways.

In older work, Gottfredson and Gottfredson (1987) claimed that forecasting has always been an integral part of the United States criminal justice system. Cunningham and Reidy (2002) argued that judges, law enforcement and correctional personnel, have for some time used projections of relative and absolute risk to help inform their decisions. Donohue (2018) claims that the sentencing of criminals is one of the most difficult responsibilities of judging because judges face multiple and conflicting instructions from both the legislature and the wider community. The sentence must:

1. exact proportional retribution for the wrong committed,
2. deter the defendant from offending,
3. discourage others from offending,
4. be of sufficient length to protect society from reoffenders, and
5. be of a suitable length and type to rehabilitate the defendant for re-entry into society.

Schild and Zeleznikow (2008) reviewed statistical sentencing information systems. In such systems, the judge initially selects a particular offense from a predefined list of offenses, and then determines a number of offender and offense characteristics. Factors related to the victim of the crime can also be part of the system classification. After specifying relevant facts of the case at hand, the user can request a histogram providing types of sentence ranges for all relevant. If desired, the systems can easily use machine learning to support decision-making.

Starr (2014) claimed that as of 2013 almost every US state had adopted some type of risk-based assessment tools to provide support in sentencing. The primary concern about using these tools revolved around the use of computerized algorithms, which provide risk scores based on the result of questions that are either answered by defendants or pulled from criminal records. There has been much concern that such tools ultimately penalize racial minorities by overpredicting the likelihood of recidivism in these groups. The most widely used tool is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), a software tool owned by Northpointe, Inc., which has been used by a number of jurisdictions, including Broward County, FL, the State of New York, the State of Wisconsin, and the State of California, among others.

Despite Kirkpatrick (2017) claiming that the COMPAS tool is viewed as a success by many jurisdictions, there has been controversy and litigation re COMPAS. New York State issued a 2012 report[33] highlighting the effectiveness of the recidivism scale, noting, "the Recidivism Scale worked effectively and achieved satisfactory predictive accuracy," with an accuracy rate of 0.71 AUC (area under curve) value (the optimal AUC value is 1.0, which would indicate no false positives/all true positives were identified)". The report noted actual and expected rates for any re-arrest were closely aligned across scores, and that the tool was more effective with higher risk cases (53.8% re-arrest rate for those deemed high-risk by the tool, versus 16.9% for those deemed low risk by the tool).

Kirkpatrick claims that in recent years, there has been significant criticism from many in academia and a scathing investigative analysis from ProPublica[34] which charged that the COMPAS algorithm and questions used to inform the algorithm were biased, since they relied on factors that could correlate with race. Critics say factors such as poverty, postal codes, and employment status can be used as proxies for race, as some are more highly correlated with minorities.

Loomis challenged the use of the risk assessment portion of the COMPAS report based on three reasons: (1) the COMPAS risk assessment violates his right to be sentenced based upon accurate information since the proprietary nature of COMPAS prevents him from assessing its accuracy; (2) it violates his right to an individualized sentence; and (3) it improperly uses gendered assessments in sentencing. Loomis' appeal was denied by the Wisconsin Supreme Court

(Yan & Zeleznikow, 2022). TheSupreme Court ruled that judges can consider such risk scores during sentencing, but warnings must be attached to the scores to flag the tool's "limitations and cautions." Moreover, the court specified that a computerized risk score cannot be the "determinative factor" in deciding whether someone is incarcerated or granted probation. The court raised concerns about how many of its risk factors could be correlated with race (Washington, 2018).

Despite criticism of the use of COMPAS, there are significant potential benefits when using data-driven risk assessments in criminal sentencing. For example, risk assessments have been endorsed as a mechanism to enable courts to reduce or waive prison sentences for offenders who are unlikely to reoffend. The Brookings Institutes argues that multiple US states have recently enacted laws requiring the use of risk assessment instruments.[35]

Villasenor and Foggo (2019) claims that along with many benefits, the growing use of algorithm-based risk assessment tools raises important concerns about due process. Due process is a core constitutional right provided through both the US 5th and 14th Amendments, both of which protect people from being deprived of "life, liberty, or property, without due process of law.[36]" A key subcategory of due process is procedural due process (Goodman, 2022). He argues "The aim of procedural due progress is to ensure fairness in legal proceedings when life, liberty, or property are at risk. When algorithm-based risk assessment tools are used in criminal proceedings, due process issues can arise because they do not provide any opportunity for meaningful cross-examination, knowledge of opposing evidence, or the true reasoning behind a decision." Villasenor and Foggo discuss an offender's right to information regarding the algorithm used to compute risk scores, and an offender's right to know exactly what those scores are.

Berk and Hyatt (2015) describe the random forests algorithm,[37] a machine learning procedure that is very useful as a forecasting tool in criminal justice settings. Random forests is essentially a flexible regression procedure for outcome variables that are usually represented by two or more categories. There is no model in random forests, so its predictive portraits are assembled inductively. The focus is upon forecasting accuracy rather than providing an explanation.

Berk and Hyatt argue that actuarial risk assessment tools such as random forests are compatible with a system of punishment based on the just deserts and limited retributivist ideologies. Predictions of risk, derived from machine learning procedures, can better help judges decide when to sentence toward the top (or the bottom) of the range of recommended sentences, or perhaps even stay outside of those ranges.

Frase (2005) claims that under a just deserts philosophy, balanced sentencing jurisprudence requires that judges consider blameworthiness and proportionality as well as how the defendant is likely to act in future circumstances. Principles of uniformity and proportionality are used to set a range for a given sentence. Then "other principles provide the necessary fine-tuning of the sentence imposed in a particular case, includ[ing] not only traditional crime-control purposes such as deterrence, incapacitation, and rehabilitation, but also a concept known as parsimony—a preference for the least severe alternative that will achieve the purposes of the sentence."

In Section 6, we shall investigate issues of explanations and fairness arising from the use of machine learning in law.

6 | PROBLEMS RELATED TO THE USE OF MACHINE LEARNING IN MAKING LEGAL PREDICTIONS

Machine Learning can be very useful for prediction when a large amount of suitable data are available. Steging et al. (2021) claim that the justification of an algorithm's outcomes is important in many domains and is particularly so in the case of law. They argue that machine learning systems can sometimes make the right decisions for the wrong reasons: despite high accuracies, not all of the conditions that define the domain of the training data are learned.

Mehrabi et al. (2021) identify two potential sources of unfairness that can be observed in outcomes derived from the use of machine learning algorithms:

1. Outcomes that occur due to biases in the data; and
2. Outcomes that occur because of actual biases in the algorithms, even if the data being used is not itself biased.

When using Machine Learning to make or support legal decisions many ethical and governance decisions are required to be made. As early as 1999, Stranieri et al. (1999), when developing their research prototype Split Up suggested how to address ethical decisions involved in the use of machine learning:

1. How to choose data—In Split Up they chose unreported commonplace cases from the Melbourne Registry of the Family Court of Australia. In 1995, neural networks were slow, expensive in computing cost and took much hard disc space—today developers can use many more cases, if they can be found.
2. How to clean and transform data—In Split Up they converted 103 free text judgment into a database. PhD students (not lawyers) conducted the cleaning and transformation of data. Stranieri et al rejected cases that stopped their neural networks from learning. Family Court of Australia judges later told the researchers that the cases they rejected were indeed by a rogue judge whose decisions were often in contradiction with other judges.
3. How to provide explanations—in Machine Learning (except for decision trees which essentially learn rules) decisions are made from black boxes with no readily available explanations. In Split Up the researchers rationalized an explanation of the answer—once they were confident of the answer, they used Toulmin' (1958) theory of argumentation to provide explanations, modeling the way Family Court of Australia judges did. Much legal theory says judges make a decision which they justify instead of rationally working their way up a tree of arguments, that is, top-down rather than bottom-up reasoning.
4. Bias in data—in Rule Based Reasoning and Case Based Reasoning, due to transparency, bias is evident. In Machine Learning, cases in the training set can lead to bias. In Split Up Stranieri et al eliminated cases that lead to no results or results that are unfair. This required human intervention.
5. Evaluation—When using a Machine Learning black box, developers want to feel fairly confident that the results are valid—The Split Up researchers used the evaluation theory of Reich and Barai (1999)

In the late 1990s, no-one in any legal community seriously considered using Machine Learning to support decision making. Hence issues of choice and transformation of data, bias in data, evaluation and explanation were ignored. But recently, there has been a growing interest in the legal domain. Thus, issues that have long been ignored are finally being investigated.

Atkinson et al. (2020) claim that until very recently the use of machine learning in the domain of AI and Law to make and predict decisions was very limited, mainly due to the fact that the explanation facilities were unsatisfactory. But as Schmitz and Zeleznikow (2021) indicate, legal decision support systems can be very useful for Self-Represented litigants. Meanwhile, many, if not most, people in need of legal redress cannot afford lawyers. Accordingly, they assert their claims in court or defend themselves without the aid of a lawyer. These people are defined as pro se or self-represented litigants (Landsman, 2009).

6.1 | Machine learning and explanation

Explanation has long been a feature of systems developed in the domain of Artificial Intelligence and law. Argumentation schemes which explained their reasoning were used in many early intelligent legal decision support systems. Many of them used the theory of Stephen Toulmin (Toulmin, 1958). These systems did not use machine learning.

In this section, we examine how explanation has been provided for three systems that rely on machine learning for reaching conclusions: Split Up, the work of Branting et al. (2021) and the work of Collonette et al (2023). Sadly, empirical work justifying the explanations arising from the systems is limited, but an ongoing feature of current research.

Atkinson et al. (2020) argue that in a judicial resolution of a legal dispute, all parties have the right to an explanation of why their case was unsuccessful. Given that such an explanation is an integral part of the decision, the losers may either accept the decision or consider if there are grounds to proceed with an appeal against the decision. Without an explanation, the required judicial transparency is missing. Recently, there has been significant research on providing explanations for the outcomes offered by machine learning systems. The general machine learning literature is full of publications discussing deficiency of the state-of-the-art explainability methods and ways to resolve them. Such examples include Lahav et al. (2018), Fryer et al. (2021), Kovalerchuk et al. (2021) and Watson (2022).

So, how are these deficiencies reflected in the legal domain? What are the advantages and disadvantages of existing explainability methods?

So let us consider the provision of explanations in intelligent legal decision support systems.

When using Rule-Based Systems to support legal decision-making, a trace of the rules being used can provide an explanation for the decision being made. Such was the case for the decision-making of The British Nationality Act as a Logic Program developed by Sergot et al. (1986). The provision of the appropriate rules on which the decision is made act as an explanation for that decision.

For Case-Based Legal Decision Support Systems, the most relevant precedent cases obtained during the case retrieval process can serve as the explanation. As Ashley (1992) states "Prior case explanations serve as patterns for explaining a new case."

But how do we provide explanations when using machine learning? Bench-Capon et al. (2009) claim that legal reasoning has many distinctive features including that many of its concepts are imprecise; precedent cases play an important role; and all conclusions are defeasible, subject often to formal appeal. Thus, the provision of argumentation is vital.

Stranieri et al. (1999) provided explanation for a legal decision support system which reached its conclusion using machine learning through the use of rationalization—once the machine learning system provides an answer, a separate argumentation system provides an explanation. The quality of the explanation can be measured by how closely the solution aligns with the argumentation made by the system. Stranieri et al. (1999) admit that obstacles to their approach include difficulties in generating explanations once conclusions have been inferred and difficulties associated with integrating two vastly different paradigms.

Predictions from the Spit-Up system were compared with those from a group of lawyers with favorable results. Eight specialist family law solicitors were asked to analyze three cases. The three cases were devised to test diverse marriage scenarios. For two cases there was compatibility between Split Up predictions and those of the eight lawyers. Furthermore, explanations for the prediction given by Split Up were similar to those given by the lawyers. The third case was however more controversial. This case involved a marriage where domestic duties were performed by paid staff and not by either party to the marriage. Split Up and four of the lawyers interpreted this situation as one where both parties had contributed to the home in equal measure. The remaining lawyers regarded this situation as improbable and, despite evidence to the contrary, assigned the majority of the home-maker role to the spouse who had not engaged in paid employment. Thus, for this case there was disagreement on interpreting the facts leading to contrary decisions.

One machine learning approach for providing explanations is to extract rules. Techniques for rule discovery include inductive logic programming and data mining for association rules. For these techniques, explanations are provided by a trace of the appropriate rules.

Steging et al. (2021) investigated using state-of-the-art explainable AI techniques to show which features impact upon the decision-making process. Their results show that even high accuracy and good relevant feature detection are no guarantee for a sound rationale.

Recently, there has been an emphasis on providing appropriate explanation for machine learning software operating in the domain of law. Atkinson et al. (2020) provided a comprehensive review of the variety of techniques for explanation that have been developed in AI and Law. As well as considering explanation by examples, explanations using rules and explanations using hybrid systems, they presented an in-depth discussion of argumentation and examined the provision of interactive explanations through dialogues. Argumentation was used to provide explanation in the Split Up system, but more generally has been used as a tool for building systems in Artificial Intelligence and Law (Walton, 2005).

Collonette et al (2023) used Article 6 of the European Convention on Human Rights to design, implement and evaluate explainable decision-support tools for deciding legal cases. They argue that Argumentation based explanations, both those based on precedent cases and those based on argumentation schemes have been able to provide effective explanations.

Kovalerchuk et al. (2021) argue that a detailed analysis of many of such claims shows that they are not well supported by actual evidence being quasi-explanations. Collonette et al (2023) justified their conclusion that the claim is supported by empirical studies of the response of lawyers with regards to outputs of the system. Their implementation used their purposely designed ANGELIC Methodology. (Al-Abdulkarim et al., 2019). The evaluation of their system had two different aspects: a. evaluating a total of 30 cases and examining whether the program produces the correct output—it did so for 29 of the 30 cases; and b. determining the admissibility of a case—which lawyers rated positively.

Branting et al. (2021) investigated the issue of how best to integrate legal knowledge and machine learning so that the resulting system can both predict and explain its results. They trained a machine learning program to identify text excerpts in case decisions that correspond to relevant legal concepts in the governing rules. Given a small sample of decisions annotated with legally relevant factual features, their program predicted outcomes of textually described cases, and identified features that help to explain the predictions, for example, by indicating the elements of the legal rule that have been satisfied or are still missing.

Branting et al. (2021) claim that "while computational techniques for explainable legal problem solving have existed for many years, broad adoption of these techniques has been impeded by their requirement for manual case

representation. The rise of large-scale text analytics and machine learning promised a way to finesse this obstacle, but the limited explanatory capability to these approaches has limited their adoption in law."

The research described two approaches to explainable legal decision prediction that operate on textual inputs. Each system was prototyped on a collection of 16,024 World Intellectual Property Organization (WIPO) domain name dispute cases.

The first approach, which uses an attention network for prediction and attention weights to highlight salient case text, was shown to be capable of predicting decisions, but attention-weight-based text highlighting did not demonstrably improve human decision speed or accuracy in an evaluation with 61 human subjects. No benefit was observed from highlighting relevant text and conflicting results on the relative benefits of providing outcome relevant case comparison features.

The second approach, termed semi-supervised case annotation for legal explanations (SCALE), exploited structural and semantic regularities in case corpora to identify textual patterns that have both predictable relationships with case decisions and sufficient explanatory value. No empirical data was provided to demonstrate the accuracy of the prediction. The authors hypothesized that users will be more accurate in predicting case outcomes if presented with features corresponding to factual findings in prior cases and that this will be particularly beneficial when comparing a new case to precedents. This poses an important empirical question: what is the threshold of predictive accuracy that a SCALE system must achieve? Branting et al.'s objective was a proof-of-concept demonstrating the feasibility of this approach for bootstrapping a small set of annotated instances into an entire labeled corpus usable for explainable prediction. This is a future goal. They proposed that future research will improve the accuracy of the overall process.

Sadly, the authors state that the use of these features for justification and explanation is beyond the scope of their paper.

6.2 | Machine Learning, Governance and Ethics

Raji (2019) claims that machine-learning systems possess inherent characteristics that warrant the regulation and study of its fairness criteria. The very traits that lead to the use of machine learning systems—that they are scalable, reliable, and persistent are also the reasons to remain skeptical and actively evaluate the fairness of the system's prediction outcomes.

Smith (2020) claims "data mining places data before theory by searching for statistical patterns without being constrained by prespecified hypotheses. Machine learning systems often rely on data-mining algorithms to construct models with little or no human guidance."

An abundance of patterns is inevitable in large data sets, and computer algorithms have no effective way of assessing whether the patterns they unearth are truly useful or just meaningless coincidences. While data mining sometimes discovers useful relationships, the data deluge has caused the number of possible patterns that can be discovered relative to the number that are genuinely useful, to grow exponentially. Thus, machine learning can be very useful for decision support, but not for making decisions.

Miller (2019) focussed on three areas in legal domains in which machine learning techniques are commonly used: adjudication in law, profiling and predictive policing, and a machine's compliance with legally enshrined moral principles. He concluded that while machine learning techniques have considerable actual and potential benefits, they also have, limitations, and both actual and potential ethical downsides. Miller argued that in liberal democracies, the use of machine learning techniques for profiling potential criminals can be inconsistent with the individual rights of citizens. Such rights include not being subject to unwarranted interference by the state.

Kirkpatrick (2017) claims that when using machine learning to make decisions about risk assessment and predictive policing, the problem is not the quality of the algorithm, but rather the use of biased data. Such data can yield unfair results. Hence, appropriately choosing and cleaning data to be used in the machine learning process greatly influences outcomes.

Kirkpatrick argues that there are two primary requirements to be examined when using machine learning in sentencing convicted criminals:

1. The use of risk-assessment algorithms, which weigh a variety of factors related to recidivism, or the likelihood that an individual will commit another crime and ultimately be re-incarcerated.
2. The use of predictive policing, using data analytics and algorithms to better pinpoint where and when a crime might occur, so police resources can be more efficiently deployed.

Both issues are fraught with many logistical, moral and political challenges. Late in 2021, the US Justice Department stated that the algorithmic tool it developed, for assessing the risk that a prisoner would be a recidivist, sadly often produced uneven results. The algorithm, Pattern, overpredicted the risk that many Black, Hispanic and Asian people would commit new crimes or violate probation rules after leaving prison. At the same time, the algorithm also underpredicted the risk for some inmates of color returning to violent crime.[38]

Machine learning models depend upon data. Existing biases, whether they be societal or structural are often inherently and invisibly embedded in data sets which are used to train machine learning algorithms. When machine learning systems use such data sets, biases which are very difficult to detect, may occur. The end result may be that certain groups suffer discrimination (Miceli et al., 2022). It should be the goal of AI and Law researchers to minimize the manner in which machine learning uses "compromised data" to make automated decisions that are discriminatory.

Surden (2020) argues that when assessing the fairness of AI-aided decision-making, one must always compare it to the baseline: what legal processes existed before the technology was introduced, and what biases are displayed in the current system? Prior to the introduction of AI-aided technology, bail and sentencing decisions were made by judges based upon evidence, and also upon a judge's personal beliefs, discretion, intuition, and experience. Judges are subject to a variety of conscious and unconscious biases. Their decisions may in themselves be biased in undesirable ways.

Surden asks if the biases the systems exhibit are necessarily worse than those in current legal structures? It could be the case that AI systems actually foster more equal treatment under the law compared with existing legal processes. In some cases, applying AI models to legal data can enhance the value of equal treatment by exposing unknown but existing biases in the current system, that may have been overwise overlooked.

Machine learning systems are good at identifying patterns, and some of these patterns might reflect existing structural injustices that can be brought to the fore to be corrected once observed. Others suggest that data-based AI systems can add more consistency to bail, sentencing, and other legal decisions as compared with the current system, involving thousands of different human judges, all with different backgrounds, experiences, and conscious and unconscious biases, applying considerable discretion and subjectivity.[39]

7 | CONCLUSION

The article commenced by noting that prior to the last decade, the major practical use of AI to support Legal Decision Making has been in areas of Statutory Interpretation. Save for academic circles, until recently there has been minimal focus upon the use of machine learning in law. Such research requires an appropriate examination of the use of data in law.

We thus commenced by examining the different use of data in both common law and civil law domains. We provided examples of how rule-based reasoning and case-based reasoning were first used to support decision-making in legal domains. The appropriate use of AI techniques to support legal decision making depends upon jurisprudential issues of open texture and boundedness of legal predicates and whether decision-making is binary or continuous.

An examination of how to use machine learning for legal decision making relies on arguments made in Stranieri et al. (1999) that there are legal tasks for which the use of statistics will be very helpful despite it being very difficult to replicate the cognitive processes conducted by lawyers. This use of machine learning is particularly appropriate for non-professionals such as self-represented litigants or those relying upon legal aid services.

We next discussed a number of systems that provided legal advice arising from the use of machine learning algorithms. One such example was our Split Up system. Split Up uses rules and neural networks to proffer advice to disputants about the potential distribution of marital property in Australian Family Law. We also examined a number of systems that advise about decision-making in courts—by:

1. The US Supreme Court,
2. The European Court of Human Rights,
3. The U.S. Patent and Trademark Office and
4. Canadian tax law.

However, we have concluded that the major use of machine learning to support legal decision-making in the provision of judicial services, has been in the domain of criminal detection and sentencing. The use of machine learning in sentencing has led to major concerns that the inappropriate use of AI leads to biases in decision making. This leads to

concerns about governance and ethics of AI systems which provide legal advice. Some of these ethical concerns can be minimized by providing enhanced explanation, choosing appropriate data to be used and having human review of all decisions. Sadly, there is limited empirical research indicating the ability of machine learning systems to provide useful explanations.

ACKNOWLEDGMENT

Open access publishing facilitated by La Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

The author has declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

John Zeleznikow https://orcid.org/0000-0002-8786-2644

RELATED WIREs ARTICLE

Themes in data mining, big data, and crime analytics

ENDNOTES

[1] For early work on Law as Code see Lessig (2003).

[2] The authors viewed administrative data as that data used in case management, rather than in the reports of cases.

[3] *Stare decisis* is the rule that the holding by a court in a previous case is binding on the same court or an inferior court in a similar case. It is one of the cornerstones of the Common Law System and provides the mechanism for case law to be as important as legislation (Zeleznikow & Hunter, 1994).

[4] Open textured legal predicates contain questions that cannot be structured in the form of production rules or logical propositions and which require some legal knowledge on the part of the user in order to answer (Zeleznikow & Hunter, 1994).

[5] For example, the Celsius scale of measuring temperature is an interval scale as the coldest possible temperature is −273°C. But the Kelvin scale is an interval scale as 0 K is the coldest possible temperature.

[6] 163 US 537 (1896).

[7] 347 U.S. (1954).

[8] The 13th Amendment abolished slavery and involuntary servitude while 14th Amendment gave former slaves citizenship rights and provided equal protection of the laws for all persons.

[9] 163 U.S. 537 (1896).

[10] 347 U.S. 483 (1954).

[11] The US Supreme Court cannot over-rule its own decisions.

[12] But not Scotland.

[13] Not Louisiana.

[14] Not Quebec.

[15] And thus, is suitable for being modeled by rule-based systems.

[16] Black (1990) defines discretion as "a power or right conferred upon decision-makers to act according to the dictates of their own judgment and conscience, uncontrolled by the judgment or conscience of others."

[17] A landmark case is one which alters our perception about knowledge in the domain—landmark cases are comparable to rules (Stranieri & Zeleznikow, 2005).

[18] A commonplace case is one that does not provide any lessons by itself, but together with numerous like cases can be used to derive conclusions (Stranieri & Zeleznikow, 2005).

[19] See for example Galanter (2004).

[20] At p125.

[21] https://robodebt.royalcommission.gov.au/about last viewed 8 February 2023.

[22] The author of this article was a co-author of the submission.

[23] See WONG, M.W.H.M., 2020. Rules as code-Seven levels of digitisation. At https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=5051&context=sol_research last viewed 15 January 2022.

[24] At p182.

[25] 'A landmark case is one which alters our perception about knowledge in the domain—landmark cases Landmark cases are comparable to rules. Landmark cases are the basis of analogical reasoning' (Stranieri & Zeleznikow, 2005).

[26] 'A commonplace case is one that does not provide any lessons by itself, but together with numerous like cases can be used to derive conclusions. Commonplace cases are to be found in the training sets of neural networks and rule induction systems' (Stranieri & Zeleznikow, 2005).

[27] For some of these 103, there were children of the marriage. But these children were either all adults or there was no dispute about the welfare of minor children.

[28] These variables were suggested by domain experts.

[29] An AdaBoost regressor is a meta-estimator that begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction. (Solomatine & Shrestha, 2004).

[30] See https://lexmachina.com/about/ last viewed 20 September 2022.

[31] A phylogenetic tree (also phylogeny or evolutionary tree) is a branching diagram or a tree showing the evolutionary relationships among various biological species or other entities based upon similarities and differences in their physical or genetic characteristics (Felsenstein, 2004).

[32] https://www.uspto.gov/about-us/news-updates/benefits-transparency-across-intellectual-property-system last viewed 22 February 2023.

[33] https://www.criminaljustice.ny.gov/crimnet/ojsa/opca/compas_probation_report_2012.pdf last viewed 22 February 2023.

[34] https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm last viewed 26 January 2022. The authors of the article chose to examine the COMPAS algorithm because it is one of the most popular scores used nationwide and is increasingly being used in pretrial and sentencing, the so-called "front-end" of the criminal justice system. They chose Broward County because it is a large jurisdiction using the COMPAS tool in pretrial release decisions and Florida has strong open-records laws.

[35] https://www.brookings.edu/blog/techtank/2019/03/21/algorithms-and-sentencing-what-does-due-process-require/ last viewed February 22 2023.

[36] https://www.law.cornell.edu/wex/due_process last viewed February 23 2023.

[37] Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned (Ho, 1998).

[38] https://www.npr.org/2022/01/26/1075509175/justice-department-algorithm-first-step-act last viewed 22 September 2022.

[39] At p 730.

FURTHER READING

Collenette, J., Atkinson, K., & Bench-Capon, T. (2023). Explainable AI tools for legal reasoning about cases: A study on the European court of human rights. *Artificial Intelligence*, *317*, 103861.

Rissland, E. L. (1989). Dimension-based analysis of hypotheticals from supreme court oral argument. In *Proceedings Second International Conference on AI and Law (ICAIL-89), Vancouver, BC* (pp. 111–120). Association for Computing Machinery.

REFERENCES

Al-Abdulkarim, L., Atkinson, K., Bench-Capon, T., Whittle, S., Williams, R., & Wolfenden, C. (2019). Noise induced hearing loss: Building an application using the ANGELIC methodology. *Argument & Computation*, *10*(1), 5–22.

Alarie, B., Niblett, A., & Yoon, A. H. (2016). Using machine learning to predict outcomes in tax law. *Canadian Business Law Journal Canada*, *58*(3), 231.

Ashley, K. (1991). Reasoning with cases and hypotheticals in HYPO. *International Journal of Man-Machine Studies*, *34*(6), 753–796.

Ashley, K. (1992). Case-based reasoning and its implications for legal expert systems. *Artificial Intelligence and Law*, *1*, 113–208.
Ashley, K. (2004). Case-based models of legal reasoning in a civil law context. In International congress of comparative cultures and legal systems of the instituto de investigaciones jurídicas. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.5214&rep=rep1&type=pdf
Ashley, K. (2017). *Artificial intelligence and legal analytics: new tools for law practice in the digital age*. Cambridge University Press.
Ashley, K. (2019). A brief history of the changing roles of case prediction in AI and law. *Law Context: A Socio-Legal Journal*, *36*, 93–112.
Ashley, K., Desai, R., & Levine, J. (2002). Teaching case-based argumentation concepts using dialectic arguments vs. didactic explanations. In *International conference on intelligent tutoring systems* (pp. 585–595). Springer.
Atkinson, K., Bench-Capon, T., & Bollegala, D. (2020). Explanation in AI and law: Past, present and future. *Artificial Intelligence*, *289*, 103387. https://doi.org/10.1016/j.artint.2020.103387
Bench-Capon, T., Prakken, H., & Sartor, G. (2009). *Argumentation in legal reasoning* (pp. 363–382). Springer.
Berk, R., & Hyatt, J. (2015). Machine learning forecasts of risk to inform sentencing decisions. *Federal Sentencing Reporter*, *27*(4), 222–228.
Black, H. C. (1990). *Black's law dictionary*. West Publishing Company.
Branting, L. K. (1991). Building explanations from rules and structured cases. *International Journal of Man-Machine Studies*, *34*(6), 797–837.
Branting, L. K., Pfeifer, C., Brown, B., Ferro, L., Aberdeen, J., Weiss, B., Pfaff, M., & Liao, B. (2021). Scalable and explainable legal prediction. *Artificial Intelligence and Law*, *29*, 213–238.
Chen, D. L. (2019). Machine learning and the rule of law. *Revista Forumul Judecatorilor (Judiciary Forum Review)*, *1*, 19–25.
Cunningham, M. D., & Reidy, T. J. (2002). Violence risk assessment at federal capital sentencing individualization, generalization, relevance, and scientific standards. *Criminal Justice and Behavior*, *29*, 512–537.
David, R., & Brierly, J. (1985). *Major legal systems in the world today*. Stevens and Sons.
DoCarmo, T., Rea, S., Conaway, E., Emery, J., & Raval, N. (2021). The law in computation: What machine learning, artificial intelligence, and big data mean for law and society scholarship. *Law & Policy.*, *43*(2), 170–199.
Donohue, M. E. (2018). A replacement for Justitia's scales: Machine learning's role in sentencing. *Harvard Journal of Law & Technology*, *32*, 657.
Dowling, M. E. (2022). Foreign interference and digital democracy: is digital era governance putting Australia at risk? *Australian Journal of Political Science*, *57*, 113–128. https://doi.org/10.1080/10361146.2021.2023093:1-16
Dworkin, R. (1986). *Law's empire*. Duckworth.
Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for Extracting Useful knowledge from volumes of data. *Communications of ACM*, *39*(11), 27–41.
Felsenstein, J. (2004). *Inferring phylogenies*. Sinauer Associates.
Fletcher, G. (1996). *Basic concepts of legal thought*. Oxford University Press.
Frase, R. (2005). Punishment purposes. *Stanford Law Review*, *58*, 67–84.
Fryer, D., Strümke, I., & Nguyen, H. (2021). Shapley values for feature selection: The good, the bad, and the axioms. *IEEE Access*, *9*, 144352–144360.
Galanter, M. (2004). The vanishing trial: An examination of trials and related matters in federal and state courts. *Journal of Empirical Legal Studies*, *1*(3), 459–570.
Galigan, Denis. 1986. *Discretionary powers: A legal study of official discretion*. London: Oxford Clarendon Press.
Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*, 3–42.
Giansiracusa, N. (2021). An evolutionary view of the US supreme court. *Mathematical and Computational Applications*, *26*(2), 37.
Giansiracusa, N. (2022). Branching on the bench: quantifying division in the supreme court with trees. *Constitutional Political Economy*, *1-23*, 36–58. https://doi.org/10.1007/s10602-022-09360-2
Giansiracusa, N., & Ricciardi, C. (2019). Computational geometry and the US supreme court. *Mathematical Social Sciences*, *98*, 1–9.
Goodman, C. C. (2022). AI, can you hear me? Promoting procedural due process in government use of artificial intelligence technologies. *Richmond Journal of Law and Technology*, *28*(4), 700–744.
Gottfredson, M. R., & Gottfredson, D. M. (1987). *Decision making in criminal justice: Toward the rational exercise of discretion* (Vol. 3). Springer Science & Business Media.
Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, *20*(8), 832–844.
Kannai, R., Schild, U., & Zeleznikow, J. (2007). Modeling the evolution of legal discretion: an Artificial Intelligence Approach. *Ratio Juris*, *20*(4), 530–558.
Katz, D., Bommarito, M., II, & Blackman, J. (2014). Predicting the Behavior of the Supreme Court of the United States: A General Approach. ARXIV.ORG, at 6 (2014). https://arxivorg/pdf/1407.6333.pdf, https://perma.cc/JXX-WQBY
Kaufman, A. R., Kraft, P., & Sen, M. (2019). Improving supreme court forecasting using boosted decision trees. *Political Analysis*, *27*(3), 381–387.
Kirkpatrick, K. (2017). It's not the algorithm, it's the data. *Communications of the ACM*, *60*(2), 21–23.
Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293.
Kovalerchuk, B., Ahmad, M. A., & Teredesai, A. (2021). Survey of explainable machine learning with visual and granular methods beyond quasi-explanations. *Interpretable artificial intelligence: A perspective of granular computing*, *937*, 217–267.

Lahav, O., Mastronarde, N., & van der Schaar, M. (2018). What is interpretable? using machine learning to design interpretable decision-support systems. *arXiv preprint*, arXiv:1811.10799.

Landsman, S. (2009). The growing challenge of Pro Se litigation, 13 Lewis & Clark L. Rev. 439.

Lessig, L. (2003). Law regulating code regulating law. *Loyola University Chicago Law Journal*, *35*, 1.

Lodder, A., & Zeleznikow, J. (2005). Developing an online dispute resolution environment: dialogue tools and negotiation systems in a three step model. *The Harvard Negotiation Law Review*, *10*, 287–338.

McCarty, L. T. (1977). Reflections on taxman: An experiment in artificial intelligence and legal reasoning. *Harvard Law Review*, *90*, 837–893.

McCue, C., & Parker, A. (2003). Connecting the dots: Data mining and predictive analytics in law enforcement and intelligence analysis. *The Police Chief*, *70*(10), 115–122.

McDonald, H. M., Kennedy, C., Hagland, T., & Haultain, L. (2021). Smarter data: the use and utility of administrative data in Victorian courts and tribunals. https://apo.org.au/node/316170

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, *28*(2), 237–266.

Medvedeva, M., Wieling, M., & Vols, M. (2023). Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law*, *31*, 195–212.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35.

Miceli, M., Posada, J., & Yang, T. (2022). Studying up machine learning data: Why talk about bias when we mean power? Proceedings of the ACM on Human-Computer Interaction, 6(GROUP), pp. 1-14.

Miller, S. (2019). Machine Learning, Ethics and Law. *Australasian Journal of Information Systems*, *23*, 1–13.

Nash, J. (1953). Two-person cooperative games. *Econometrica: Journal of the Econometric Society*, *21*, 128–140.

Oatley, G. C. (2021). Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(2), e1432.

Oatley, G. C., Ewart, B. W., & Zeleznikow, J. (2006). Decision support systems for police: lessons from the application of data mining techniques to 'soft' forensic evidence. *Artificial Intelligence and Law*, *14*, 35–100.

Rai, A. K. (2018). Machine learning at the patent office: Lessons for patents and administrative law. *Iowa Law Review*, *104*, 2617.

Raji, D. (2019). That's not fair! *XRDS: Crossroads, The ACM Magazine for Students*, *25*(3), 44–48.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *The New England Journal of Medicine*, *380*, 1347–1358.

Reich, Y., & Barai, S. V. (1999). Evaluating machine learning models for engineering problems. *Artificial Intelligence in Engineering*, *13*(3), 257–272.

Rissland, E. L., Ashley, K., & Branting, L. K. (2005). Case-based reasoning and law. *The Knowledge Engineering Review*, *20*(3), 293–298.

Rissland, E. L., Ashley, K. D., & Loui, R. P. (2003). AI and Law: a fruitful synergy. *Artificial Intelligence*, *150*(1–2), 1–15.

Rissland, E. L., & Skalak, D. B. (1991). CABARET: statutory interpretation in a hybrid architecture. *International Journal of Man-Machine Studies I*, *34*(6), 839–887.

Schild, U., Stranieri, A., & Zeleznikow, J. (1999). Techniques for reasoning in discretionary legal domains. *IASTED International Conference on Law and Technology (LawTech1999)* (pp. 51–63). ACTA Press, Anaheim, California.

Schild, U. J., & Zeleznikow, J. (2008). Comparing sentencing decision support systems for judges and lawyers. *Journal of Decision Systems*, *17*(4), 523–552.

Schlobohm, D. A., & Waterman, D. A. (1987). Explanation for an expert system that performs estate planning. In *Proceedings of the first international conference on artificial intelligence and law* (pp. 18–27). Association for Computing Machinery.

Schmitz, A., & Zeleznikow, J. (2021). Intelligent Legal Tech to Empower Self-Represented Litigants. *Columbia Science and Technology Law Review*, *23*, 142–190.

Sergot, M. J., Sadri, F., Kowalski, R. A., Kriwaczek, F., Hammond, P., & Terese Cory, H. (1986). The British Nationality Act as a Logic Program. *Communications of the ACM*, *29*, 370–386.

Smith, G. (2020). Data mining fool's gold. *Journal of Information Technology*, *35*(3), 182–194.

Solomatine, D. P., & Shrestha, D. L. (2004). AdaBoost. RT: a boosting algorithm for regression problems. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)* (Vol. 2, pp. 1163–1168). IEEE.

Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, *66*, 803.

Steging, C., Renooij, S., & Verheij, B. (2021). Rationale discovery and explainable AI. In *Legal Knowledge and Information Systems* (pp. 225–234). IOS Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.

Stranieri, A., & Zeleznikow, J. (1992). SPLIT-UP: Expert system to determine spousal property distribution on litigation in the Family Court of Australia. In *Proceedings of Artificial Intelligence Conference (Australia)* (Vol. 92, pp. 51–56). World Scientific Publishers.

Stranieri, A., & Zeleznikow, J. (2005). *Knowledge Discovery from Legal Databases, Springer Law and Philosophy Library* (Vol. 69). Springer Law and Philosophy Library.

Stranieri, A., Zeleznikow, J., Gawler, M., & Lewis, B. (1999). A hybrid rule–neural approach for the automation of legal reasoning in the discretionary domain of family law in Australia. *Artificial intelligence and Law*, *7*(2), 153–183.

Surdeanu, M., Nallapati, R., Gregory, G., Walker, J., & Manning, C. D. (2011). Risk analysis for intellectual property litigation. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law* (pp. 116–120). Association for Computing Machinery.

Surden, H. (2014). Machine Learning and Law, 89 Wash. L. REV. 87.
Surden, H. (2020). Ethics of AI in Law: Basic Questions. In M. D. Dubber (Ed.), *The Oxford Handbook of Ethics of AI*. Frank Pasquale and Sunit Das.
Tetley, W. (1999). Mixed jurisdictions: Common Law v. Civil Law (codified and uncodified). *Louisiana Law Review*, *60*, 677–905.
Toulmin, S. (1958). *The uses of arguments*. Cambridge University Press.
Villasenor, J., & Foggo, V. (2019). *Algorithms and sentencing: What does due process require?* Brookings TechTank.
Walton, D. (2005). *Argumentation methods for artificial intelligence in law*. Springer Science & Business Media.
Washington, A. L. (2018). How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate. *Colorado Technology Law Journal*, *17*, 131.
Waterman, D., Paul, J., & Peterson, M. (1986). Expert systems for legal decision making. *Expert Systems*, *3*(4), 212–226.
Waterman, D., & Peterson, M. (1981). Models of legal decision-making. The RAND Corporation. R-2717-ICJ, Santa Monica.
Watson, D. S. (2022). Conceptual challenges for interpretable machine learning. *Synthese*, *200*(2), 65.
Weizenbaum, J. (1966). ELIZA: a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.
Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco–Or was it something worse? *Australian Journal of Public Administration.*, *80*(2), 340–360.
Yan, H., & Zeleznikow, J. (2022). The appropriate use of AI in law: Investigating the liability of artificial intelligence in legal decision-making. *ANU Journal of Law and Technology*, *3*(2), 8–38.
Zeleznikow, J. (2002). Using web-based legal decision support systems to improve access to justice. *Information and Communications Technology Law*, *11*(1), 15–33.
Zeleznikow, J., & Hunter, D. (1994). *Building intelligent legal information systems: Knowledge Representation and reasoning in law (No. 13)*. Kluwer Law and Taxation Publishers.
Zeleznikow, J., Vossos, G., & Hunter, D. (1994). The IKBALS project: Multimodal reasoning in legal knowledge based systems. *Artificial Intelligence and Law*, *2*(3), 169–203.

How to cite this article: Zeleznikow, J. (2023). The benefits and dangers of using machine learning to support making legal predictions. *WIREs Data Mining and Knowledge Discovery*, *13*(4), e1505. https://doi.org/10.1002/widm.1505