

## Thirty Years of *Artificial Intelligence and Law: The Third Decade*

Serena Villata · Michał Araszkiewicz · Kevin Ashley · Trevor Bench-Capon · L. Karl Branting · Jack G. Conrad · Adam Wyner

the date of receipt and acceptance should be inserted later

**Abstract** The first issue of *Artificial Intelligence and Law* journal was published in 1992. This paper offers some commentaries on papers drawn from the Journal's third decade. They indicate a major shift within Artificial Intelligence, both generally and in AI and Law: away from symbolic techniques to those based on Machine Learning approaches, especially those based on Natural Language texts rather than feature sets. Eight papers are discussed: two concern the management and use of documents available on the World Wide Web, and six apply machine learning techniques to a variety of legal applications.

### 1 Introduction

The most significant feature of the journal's third decade was the dramatic increase in the application of machine learning techniques to a variety of AI and Law tasks. This is not to say that more traditional aspects of AI and Law were neglected entirely. The first issue of the journal in 1992 contained papers on argumentation about cases (Skalak and Rissland, 1992), reasoning with norms (Jones and Sergot, 1992), and a methodology for representing legal knowledge (Bench-Capon and

---

Serena Villata  
Université Côte d'Azur, CNRS, Inria, I3S, France

Michał Araszkiewicz  
Jagiellonian University, Krakow, Poland

Kevin Ashley  
University of Pittsburgh, Pennsylvania, USA

Trevor Bench-Capon  
University of Liverpool, UK.

L. Karl Branting  
The MITRE Corporation, USA

Jack G. Conrad  
Thomson Reuters, TR Labs, USA

Adam Wyner  
University of Swansea, UK

Coenen, 1992)<sup>1</sup> and all of these topics continued to be pursued in the third decade. For reasoning with cases there were developments from the original formalisation of precedential constraint proposed by John Horty (Horty, 2011) and presented in this journal in Horty and Bench-Capon (2012)<sup>2</sup>. In particular Horty extended his theory to include factors with magnitude as well as Boolean factors (Horty (2019) and Horty (2021)). This work was critiqued in Rigoni (2015) and Rigoni (2018), and a comparative formalisation and analysis of the various approaches was given in Prakken (2021). Norms were discussed in a special issue (Andrighetto et al., 2013), and papers such as Mahmoud et al. (2015) and Bench-Capon and Modgil (2017). Methodologies for representation legal knowledge were proposed in Al-Abdulkarim et al. (2016) and Kowalski and Dato (2021). Some new topics also emerged: technological developments raised the possibility of “smart” contracts (e.g. Azzopardi et al. (2016) and Governatori et al. (2018)) and the ever increasing capabilities of AI systems led to discussions of the legal status of such systems, notably Bryson et al. (2017) and Solaiman (2017), as well as a special issue on the topic (Indurkha et al., 2017)). None the less the rise of machine learning in the second half of the decade under consideration was striking, and the bulk of the papers discussed in this article relate to that development.

The first two papers, however, represent a continuation of work which had become established in the previous decade, following the development of the internet, with the consequent ready availability of large quantities of legal information. The first, Francesconi (2014), commented on by Michał Araszkiewicz, provides a description logic framework for reasoning over normative provisions to facilitate the application of Semantic Web technologies to AI and Law. The second, Boella et al. (2016), commented on by Adam Wyner, describes a comprehensive suite of tools to provide for the management of legal documents on the Web. The remaining six papers discuss various aspects of the application of Machine Learning techniques to AI and Law.

Two concern the use of Machine Learning to predict the outcome of legal cases. Medvedeva et al. (2020a), commented on by Trevor Bench-Capon, provides a representative example of such approaches, conducting three experiments relating to the European Court of Human Rights. The other, Branting et al. (2021), commented on by Kevin Ashley, addresses one of the important limitations of such systems, namely the ability to explain and justify the predictions in legal terms. The paper suggest that by using machine learning to ascribe factors, use can be made of the techniques for explanation developed for traditional AI and Law systems.

The other papers discuss a variety of other legal tasks. Nguyen et al. (2018), commented on by Jack Conrad, uses neural networks to identify and label particular parts of Japanese legal documents. Abood and Feltenberger (2018) tackles the particular, but important, task of identifying patents relevant to a particular topic. The commentary, by Karl Branting, also contains an insightful discussion of some of the reasons why Machine Learning approaches became so popular at this time. The final two papers have commentaries by Serena Villata. Ruggeri et al. (2022) analyses contracts to detect clauses which may potentially be unfair, while

---

<sup>1</sup> See Sections 2, 3 and 4 of Governatori et al. (2022), elsewhere in this issue.

<sup>2</sup> See Section 10 of Sartor et al. (2022), elsewhere in this issue.

Tagarelli and Simeri (2021) supports the task of retrieving relevant law articles from the Italian Civil Code.

The papers discussed in this article provide a good flavour of how the availability of vast amounts of legal data gave rise to a number of new applications. Not only were tools needed to manage the data, but the existence of the data opened up the possibility that it could be used to train systems to perform variety of legal tasks.

## **2 A description logic framework for advanced accessing and reasoning over normative provisions (Francesconi, 2014). *Commentary by Michal Araszkiewicz***

Francesconi (2014) is a contribution to both theoretical and practical discussions concerning the application of Semantic Web technology in the domain of law, which has become one of the central topics in AI and Law since the 2000s. The paper presents a model of statutory legal knowledge (provisions and related axioms) encompassing Hohfeldian relations, using the RDF(S)<sup>3</sup>/OWL<sup>4</sup> standard developed by the World Wide Web Consortium<sup>5</sup>. The model enables the query and retrieval of particular types of provisions as well as reasoning over the stored knowledge using a Description Logic (OWL-DL<sup>6</sup>). Moreover, the paper presents a prototype architecture of a provision query system (ProMISE).

The model addresses a practical problem concerning the retrieval of all related provisions from the statutory text. A lawyer or an addressee of a regulation is interested in retrieving all provisions that may play a role in the assessment of a given fact situation or a legal relationship. Standardized knowledge representation techniques developed in the Semantic Web approach enable meaningful relations between provisions and their elements to be captured.

One approach is that of the *Provision Model* introduced earlier by Biagioli (Biagioli (1997), Biagioli (2009)). In that model, legal provisions are understood as textual entities: sentences endowed with meaning. They may be analyzed from the point of view of two profiles: a structural (or formal) profile, focusing on the organization of the legislative text (articles, paragraphs etc.) and a semantic profile, representing the specific organization of the substantial meaning of the provisions. The Provision Model enables the classification of provision types (Term Definition, Duty, Right, Power, Procedure and more specific categories) as well as provision attributes such as Bearer or Counterpart. Attributes may assume values from certain ranges. On a general level, provisions types are classified into two categories: *Rules*, encompassing both constitutive and regulative rules, and *Rules on Rules*, which comprise various types of amendments.

Importantly, the Provision Model distinguishes between *logical* relations and *technical* relations between (the elements of) legal provisions. Logical relations are domain-neutral dependencies which encompass *inter alia* the relations between fundamental legal concepts as introduced by Hohfeld (Hohfeld, 1913). The investigation of technical relations is not possible in abstraction because they can be

---

<sup>3</sup> Resource Description Framework

<sup>4</sup> Web Ontology Language

<sup>5</sup> [https://www.w3.org/2001/sw/wiki/Main\\_Page](https://www.w3.org/2001/sw/wiki/Main_Page)

<sup>6</sup> <https://www.w3.org/TR/owl-guide/>

identified only in connection with a given regulation. The technical relations obtain due to a decision of the legislator. For instance, they may follow from legal definitions: a defined term, *Definiendum*, should be interpreted consistently in all its instances on the statutory text. Another type of a technical relation may follow from ascription of a sanction in case of non-compliance of a Bearer with a related duty. The technical relations may be indicated in the statutory text through referring provisions or they may be reconstructed taking into account the content of the provisions and related attribute values.

Francesconi (2014) introduces an important extension of the original Provision Model by taking into account the relations encompassed in the Hohfeldian squares of opposition (Hohfeld, 1913), where the first square involves *deontic* concepts (Right, Duty, No-right and Privilege) and the second *potestative* concepts (Power, Liability, Disability and Immunity). In order to represent these relations adequately, the provision attributes are specified to capture the provision type (for instance, the model will use the expression *hasDutyBearer* rather than *hasBearer*). Moreover, it is observed that the Hohfeldian relations enable implicit attributes to be inferred on the basis of those explicitly expressed in the statutory text. For instance, if a provision states explicitly that a person A has a right towards a person B, it typically warrants an inference that the person B has a duty towards a person A (Sartor, 2006). This leads to the further specification of the model, as in the classes representing Hohfeldian concepts it is possible to define two disjoint subclasses, referring to implicit and explicit versions of the concepts, for instance *ImplicitRight* and *ExplicitRight*. This results in an extension of the catalogue of attributes of provisions as it will contain not only explicit attributes (for instance *hasExplicitDutyBearer*) but also implicit attributes (*hasImplicitDutyBearer*). The relations between correlative Hohfeldian concepts enable new equivalence relations to be defined in the model; for instance, the classes *ImplicitDuty* and *ExplicitRight*, or attributes *hasExplicitRightBearer* and *hasImplicitDutyCounterpart* will belong to such relations.

The expressiveness of the model is corroborated by examples. The selected domain is European Union consumer protection law: the Directive 2002/65/EC of the European Parliament and of the Council of 23 September 2002 concerning the distance marketing of consumer financial services. An excerpt thereof has been annotated using a CEN-Metalex (Boer et al., 2010) compliant mark-up syntax and represented through the consumer protection domain ontology DALOS (Agnoloni et al., 2009). The OWL-DL description of the Provision Model and the instantiated regulation enables inference through the OWL-DL reasoner. The RDF triple store (both inferred and non-inferred models) may be queried using SPARQL<sup>7</sup>. The paper confirms the validity of results of such queries for both logical relations and technical relations in the selected domain.

Finally, the paper describes the architecture of a prototype of ProMISE (Provision Model-based Inferential Search Engine). A data set of Italian legislative documents, published on the Web using the URI and XML NormeInRete (Francesconi, 2006) standards has been created, and domain ontologies have been used to annotate the semantic content of the provisions. The DALOS multilingual European consumer protection law ontology (Agnoloni et al., 2009) was used in the testing

<sup>7</sup> SPARQL is a recursive acronym for SPARQL Protocol and RDF Query Language, an RDF query language. See <https://www.w3.org/TR/rdf-sparql-query/>

environment. The result is a set of RDF triples representing normative provisions stored in the OpenLink Virtuoso Database Server<sup>8</sup>. Since the associated reasoner has limitations, external reasoners such as Pellet<sup>9</sup> (a Java-based OWL-DL reasoner) can be used. Finally, the paper comments on a Web application providing the users with semantic web search facilities able to query the system according to the Provision Model.

The paper presents an excellent example of a research project which integrates insights from legal theory (Hohfeldian concepts and provision classifications represented in the Provision Model) with formal and computational modeling realized for a practical purpose. The proposed model enables advanced retrieval and reasoning over normative regulations (also in multilingual settings) by exploiting relations between legal provisions. The proposed approach enables the complexity of the problems to be kept within the DL computational tractability. A limitation of this and similar solutions is the reliance on the proper, validated structural and semantic mark-up of normative texts. The paper rightly indicates that the development of legislative XML mark-up standards and the availability of software tools enabling automation or facilitating the mark-up process (both structural and semantic) are important factors in the development of similar projects. Over the last decade, the LegalXML community has developed standards for structuring legal texts (for instance, based upon Akoma Ntoso<sup>10</sup>), developed a language for modeling legal rules (LegalRuleML<sup>11</sup>) and URI naming convention (for instance, ELI/ECLI (Van Opijnen, 2011)). The solutions developed in this framework have already achieved a relatively highly standardized form, but they are still subject to further development (see Sartor et al. (2011) for the broad introduction to the field; see also Athan et al. (2015), Casanovas et al. (2016), and Palmirani (2020)).

The project is also a step towards the standardization of the semantic annotation process for legislative documents. The existence of such standards may not only contribute to increased access to legal information (e.g. Nazarenko et al. (2021)), but also to improved quality of legislation, as it may effectively contribute to the detection of legislative errors, for instance contradictions in regulation (Araszkievicz et al., 2021). Francesconi (2014) has also been cited in the legal-theoretical literature on legislative techniques (e.g. Kłodawski (2021)).

### **3 Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law (Boella et al., 2016). *Commentary by Adam Wyner***

The aim of EUNOMOS (Boella et al., 2016) was to develop an online system for legal researchers, knowledge engineers, and practitioners to manage and monitor legislative information by searching, classifying, annotating, and building legal knowledge which keeps up to date with legislative changes<sup>12</sup>. To achieve the objective, the system is a legal document and knowledge management system that

<sup>8</sup> <https://virtuoso.openlinksw.com/>

<sup>9</sup> <https://github.com/stardog-union/pellet>

<sup>10</sup> <http://www.akomantoso.org/>

<sup>11</sup> [https://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=legalruleml](https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legalruleml)

<sup>12</sup> The Eunomos software was originally developed to support regulatory compliance in the context of the ICT4Law project, subsequently further extended in the context of the projects

represents legislative sources using XML and enables access to them using legal knowledge in an ontology. EUNOMOS is said to be useful to finance compliance officers, legal professionals, public administrators, voluntary sector staff, and citizens in a range of domains. Out of scope are logical representations of legal rules or information extraction of legislative text.

A range of issues are identified and approaches offered:

1. *The scope and volume of laws, which need to be drawn together into a machine-readable corpus.* Approach: a large database of laws extracted from legislative portals, converted into XML, and updated.
2. *Laws are not clearly or uniquely classified with respect to domain topics and jurisdictions, leading to fragmentation.* Approach: identify domain topics, then classify and annotate documents for the domains. Cluster documents according to similarity. The result is that users can query and view legislation across topics and jurisdictions from the same Web interface.
3. *Users may specialise, where some portions of a law may be more relevant to a user.* Approach: semi-automated, fine-grained classification of sections of articles, enabling users to view sections that are relevant to them.
4. *Heterogeneous, distributed legal sources are in tension with user demands to access the law over the internet using Open Government Data and Linked Open Data.* Approach: machine-readable law represented in accordance with open standards and open data.
5. *The law must be kept up-to-date and consolidated.* Approach: annotate versions to consolidate and update users.
6. *Laws may amend or abrogate other laws.* Approach: annotate links between laws with semantic information on their relationship.
7. *The meaning of legal terms can vary with respect to jurisdiction and over time.* Approach: domain-specific ontologies of terms and the ontological concepts can be linked with text in the source.
8. *Legal language can be vague, imprecise, or polysemous.* Approach: link language with additional information, clarifications, and interpretations, which knowledge engineers provide based on research.
9. *Laws contain a range of cross-references which must be accessed.* Approach: identify cross-references and link documents such that the content of the cross-reference either pops up or is accessed by the link.
10. *Between the source text and the machine-readable resource, there is a knowledge bottleneck; that is, how to transfer the rich content of the source into the resource.* Approach: each issue is treated as a module in an overall process. Each module uses tools relevant to the task and data, which are largely semi-automatic, interactive support tools, so that a legal knowledge engineer uses the tool to filter or structure the data as it passes into the machine-readable representation. The aim is that the interfaces do not require specialist skill, e.g., text classification, rule application, NLP, and so on. Some of the tasks include adding legislation to the database, checking XML parses, adding cross-references, classifying references, checking document domain attribution, adding terms and concepts to the ontology, commenting on the text, linking the ontology to the source text, and adding explanations or interpretations.

---

ITxLaw and EUCases, and then used in the context of the projects ProLeMAS, BO-ECLI, and MIREL.

The core system of EUNOMOS has three layers:

- a database of Italian national laws that are processed for similarity (Cosine Similarity with TF-IDF) and for domain topic (SVM) then converted into XML using the ITTIG XML parser to convert the NormaInRete XML format;
- an ontology of legal terms from the Legal Taxonomy Syllabus, which are augmented with lexical semantic relations and contextual or interpretive information, then linked to the law (using the TULE parser (Lesmo, 2009)); and
- a user interface for querying the database with respect to the ontology, keywords, and other metadata. Each layer has a range of subcomponents. A web interface is discussed, and workflows for users and knowledge engineers are outlined. The EUNOMOS system is the basis of the Menslegis commercial service for compliance, which was distributed by Nomotika, a spinoff from the University of Turin<sup>13</sup>.

Future work aimed to consolidate legal texts across versions, enable multilingual search, extend the ontology, and extract information. It is observed that the data resource requires considerable maintenance work to keep up to date: providing the workflow is one matter, but keeping the flow going is another.

### 3.1 Discussion

As can be gathered from the catalogue of issues and approaches, EUNOMOS has a global, ambitious aim, covering a heterogeneous range of familiar issues and technologies which bear on one another. An integrated workbench would be a very useful contribution; as an outline of the issues, approaches, and technologies available at the time, the paper serves a useful purpose.

However, the practical and scientific contributions are rather limited. It is useful to have issues and approaches sketched in one place, even if they individually would not appear to be novel, being of long standing in AI and Law. On the practical side, the data and software have been acquired by a company and are unavailable for any further academic research and development. On the scientific side, the chief aims were to reuse standard approaches (e.g., TF-IDF, SVM, and ontologies) and integrate them with developments from prior projects (e.g., Legal Taxonomy Syllabus, NormInRete, and the TULE parser). As the prior art of the components was not new, the source code of this integration could have been a good contribution were the source code of the system available. Issues specifically about integration are not discussed, which could have yielded insights. There are several evaluations reported for some of the components, but it is not clear if this is new work; and in any case, independent peer review of such components and evaluations would have greater weight. As the system often makes use of Italian oriented tools applied to Italian legislation, the claims about generality are hard to support. And finally, despite evidence of an integrated platform and identification of user communities, there is no report about user experiences or user results.

Nonetheless, the article has had some influence as evidenced by some 93 citations to it on Google Scholar.<sup>14</sup> The citing articles are rather heterogeneous,

<sup>13</sup> <https://en.unito.it/research/knowledge-transfer/spin-companies>

<sup>14</sup> Last accessed on 16 May 2022.

picking up on one theme or another or applying to some specific domain. This observation is itself somewhat interesting, as EUNOMOS promotes an inclusive, across the board approach to ingesting, processing, structuring, and serving legal information to clients. One has the sense that such an integrated platform remains a challenge, even if the commercial interests are there. Moreover, it would appear that it is not the technologies themselves that are the source of friction, but rather knowledge acquisition and maintenance, as indicated in (10) above; which appear to still be very much at issue.

#### **4 Automated patent landscaping (Abood and Feltenberger, 2018).** *Commentary by L. Karl Branting*<sup>15</sup>

*Automated Patent Landscaping* by Aaron Abood and Dave Feltenberger (Abood and Feltenberger, 2018) exemplifies many of the important recent trends in AI and Law research. This section first summarizes these trends and then discusses Abood and Feltenberger’s work in relation to the trends.

##### 4.1 The Rise of Data-Centric Approaches

In its early decades, research in AI and Law largely focused on formal models of legal argumentation based on manually constructed representations of case facts and legal rules and norms (Branting, 2017). This was consistent with practice in AI as a whole during this period, which was dominated by research in inference, planning, parsing, and other symbolic and logic-based methods. This overall approach was consonant with the then-influential functionalist stance in philosophy, under which certain types of symbol manipulation were held to be sufficient for general intelligence (Fodor, 1975) and perhaps even consciousness itself (Hofstadter, 1979).

Rapid growth in large-scale data analytic capability in the 21st century dramatically shifted the emphasis of AI research from symbolic computation toward empirical, corpus-based techniques, which typically emphasize statistical and other machine learning techniques. A proliferation of community-wide datasets, predictive tasks with straightforward evaluation criteria, and leader boards created institutional rewards for incremental improvements in predictive accuracy.

The AI and Law community was slow to adopt this corpus-based approach<sup>16</sup> for several reasons. Attorneys’ ability to predict outcomes of disputes typically increases with legal experience, but it is much less central to legal expertise than articulating the most cogent argument for a given outcome. Human attorneys’

<sup>15</sup> Acknowledgement: Approved for Public Release; Distribution Unlimited. Public Release Case Number 22-0686. The author’s affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE’s concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author. ©2022 The MITRE Corporation. ALL RIGHTS RESERVED.

<sup>16</sup> For example, by 1996 statistical methods had gone “from being virtually unknown in computational linguistics to being a fundamental given” (Abney, 1996); almost every paper in the 2002 Human Language Technology conference (Marcus, 2002) involved corpus analysis techniques. By contrast, the proportion of papers reporting on corpus-based work at the International Conferences on AI and Law increased only slightly during this period, from about 18% in ICAIL 1997 to about 21% in ICAIL 2003.

training typically emphasizes the ability to generate, understand, and evaluate conflicting textual arguments under various factual scenarios. While corpus-based methods for texts have a lengthy history, particularly for author identification and stylometry (Holmes and Kardos, 2003), the shallow statistical methods commonly used in these applications were ill-suited for the argument-centered analysis characteristic of much legal problem solving. In addition, parsing methods developed for non-legal text corpora were inadequate for the unique characteristics of legal text, such as syntactic complexity and nested enumerations (Morgenstern, 2014). Finally, the lack of accessibility of large legal corpora drastically impaired the ability of researchers to apply corpus-based methods to legal decisions.<sup>17</sup> Thus, the prevailing view of what constituted legal expertise, limitations on analytical tools, and relative lack of availability of suitable legal corpora all impeded progress in corpus-based or linguistic approaches to legal text analysis.

The last decade, however, has seen a dramatic increase in interest in applying text analysis techniques to legal problem solving. Three factors have contributed to this increase.

First, new Human Language Technology (HLT) techniques, some based on technologies originally developed for vision, have improved the ability to analyze the legally relevant aspects of texts. These techniques include semantic vector spaces, which permit the detection of synonymy in words, sentences, and larger text spans; more robust parsing and semantic role-labeling capabilities; graph analysis of statutory or case citation networks; argumentation analysis; and explainable AI. Dramatic improvements in Deep Learning have significantly enhanced performance in these and other legal text analysis tasks (Chalkidis and Kamps, 2019).<sup>18</sup>

Second, there has been a growth in interest in legal problem-solving applications beyond decision support and argumentation. Attorneys perform a wide range of activities beyond those taught in first-year law school classes, and the AI and Law community has become increasingly aware of the abundant opportunities for AI to assist in many of these activities. Finally, there has been a dramatic increase in the availability of legal decision corpora.

One indication of the growth of technical capabilities and proliferation of applications is the rapid expansion of legal technology in government and the private sector. A recent study showed that almost half of administrative agencies in the US have experimented with AI or machine learning, primarily in enforcement, regulatory research, analysis, monitoring, internal management, public engagement, and adjudication (Engstrom et al., 2010). As of this writing, the Stanford CodeX Legal Tech Index lists over 1,800 companies that provide legal services, such as contract analysis, document automation, eDiscovery, analytics, and compliance<sup>19</sup>.

As recently as a decade ago, workshops on legal text analysis were a rarity, with the first ICAIL workshop on the topic being held only in 2011 (Wyner and Branting, 2011) and the next not until 2015 (Branting, 2015). Today there is a proliferation of forums for presenting work in this field, including ASAIL (Automated

---

<sup>17</sup> In the United States even today “a range of technical and financial obstacles blocks large-scale access to public court records” Pah et al. (2020).

<sup>18</sup> Papers on neural networks were published in the *Journal* throughout its history, including prior to the development of Deep Learning, e.g., Crombag (1993), Stranieri et al. (1999), and eight others in the 1990s. However, these papers generally applied neural networks to symbolic representations rather than text.

<sup>19</sup> <https://techindex.law.stanford.edu/>. Last accessed February 27, 2022.

Semantic Analysis in Law)<sup>20</sup>, NLLP (Natural Legal Language Processing)<sup>21</sup>, IEEE Applications of Artificial Intelligence in the Legal Industry<sup>22</sup>, and many others, and papers on this topic are increasingly viewed as mainstream work in AI and Law forums, such as this Journal. The original version of Abood and Feltenberger (2018) was presented at a workshop on *Legal Text, Document, and Corpus Analytics* at the University of San Diego in 2016 (Branting, 2016) and was expanded into a submission to a special issue of the Journal (Conrad and Branting, 2018).

#### 4.2 Discussion of Abood and Feltenberger (2018)

“Automated Patent Landscaping” (hereinafter “Abood and Feltenberger”) exemplifies many of these trends. Rather than proposing a normative model, explicating an argumentation framework, or depending on a manually curated representation, Abood and Feltenberger present a tractable solution to a significant real-world legal problem that integrates human expertise with recent advances in HLT and machine learning.

Identification of patents relevant to a given topic—together with the ownership and litigation status of those patents—is a task of importance to many commercial, government, and academic stakeholders. Patent landscapes are used in industry to inform strategic decisions on investments; research and development; to gain insight into competitors’ activities; and to estimate the advisability of developing new products in a given market (Trippe, 2015). Similarly, public policy makers may consult a patent landscape to inform high-level policy matters, in fields such as health, agriculture and the environment. Patent landscaping is an expensive, challenging, and time-consuming activity for human experts. In view of the importance and difficulty of patent landscaping, automated assistance for this activity could have significant practical benefits.

Patent data consists of text (title, abstract, detailed description, and claims) together with various forms of metadata, including class codes, citations, and family relationships. Class codes are organized into complex hierarchies of topic labels; multiple codes are typically applicable to a single patent. Patents typically contain citations to prior published patents or patent applications that are intended to help establish that the citing patent is novel and non-obvious. Patent families consist of different publications of a single patent together with publications of other patents having a priority relationship, e.g., a patent filed in multiple countries or “multiple applications in the same country that share a detailed description, but have different claims.”

Abood and Feltenberger uses a semi-automated approach that starts with a human-curated seed set that is then expanded through family citation and class code links into an over-inclusive set. This expanded set is pruned using a machine learning model trained to distinguish between seed set (positive) instances and randomly sampled instances from outside of the expansion set (negative instances).

The accuracy of Abood and Feltenberger’s process depends on an initial seed set that is representative of the sub-topics that should characterize the desired

<sup>20</sup> <https://sites.google.com/view/asail/asail-home?authuser=0> Last accessed 14/03.2022.

<sup>21</sup> <https://nllpw.org/> Last accessed 14/03.2022.

<sup>22</sup> <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=145044&copyownerid=127864> Last accessed 14/03.2022.

landscape. Selecting such a seed set requires human expertise, but it is far simpler than requiring a user to identify keywords or class codes sufficient to identify all and only the patents in the desired landscape. Indeed, an appropriate seed set can relieve the user from having to understand the “technical nuance” of a patent topic. Thus, Abood and Feltenbergerer illustrates extending or enhancing human expertise rather than attempting to substitute for it.

The seed set is expanded based on the two forms of structured patent metadata mentioned above: family citations and class code (restricted to those that are “highly relevant,” meaning roughly highly discriminant of the seed set). Depending on the particular technical area and the aims of the user, the expansion can be either one level (“narrow”) or two levels (“broad”).

The expansion step is, in general, high recall but low precision, so a pruning step is needed to refine the results. A key insight in Abood and Feltenbergerer is that the instances that are in neither the expansion set nor among the original seeds can be used as negative training instances. These instances are termed the “anti-seeds” in Abood and Feltenbergerer. A model to estimate relevance of instances in the expansion set can therefore be trained on seeds and a sampled subset of anti-seeds.

A key observation of Abood and Feltenbergerer is that “the specific machine learning methodology is orthogonal to the broader technique of deriving the training data in a semi-supervised way.” The procedure for developing a seed/anti-seed training set is independent of the particular technique used to train the model used for pruning. The paper itself compared three types of predictive model. The first was an LSTM (Long Short Term Memory) neural network (Hochreiter and Schmidhuber, 1997) taking as input class codes, citations, and a word2vec (Goldberg and Levy, 2014) embedding of text from patent abstracts. The second approach was an ensemble of shallow neural networks with SVD (singular value decomposition) embeddings (de Lathauwer et al., 2000), and the third was Perceptron (Rosenblatt, 1958) with random feature projection (Blum, 2005). In 10-fold cross validation, the LSTM approach was observed to have the highest accuracy. However, Abood and Feltenbergerer’s methodology can incorporate improvements in classifier technology as they occur.

The authors observe that the paradigm they propose – creating a candidate set by following metadata links from seeds, then pruning the candidates using a model trained from seeds and anti-seeds – could be applied to other domains in which documents have both text and linking metadata, such as “scholarly articles and legal opinions” Indeed, this is a broadly applicable approach. For example, legislative and regulatory text are also characterized by both text and metadata amenable to network analysis (e.g., citations). There is active current research that exploits this hybrid character of decisions and statutes by integrating text similarity metrics with metadata-based network analysis in a predictive framework, e.g., Dadgostari et al. (2021), Leibon et al. (2018), Adusumilli et al. (2022), and Sadeghian et al. (2018).

Notwithstanding the applicability of this approach to legal data other than patents, citations to Abood and Feltenbergerer have been almost exclusively from patent analysis researchers, e.g., Choi et al. (2022). However, this work exemplifies four significant distinguishing characteristics of much recent research in AI and law that go far beyond patent analysis.

The first is emphasis on assisting rather than replacing human expertise in legal problem solving. The second is casting this assistance in a prediction framework that is amenable to machine-learning techniques. The lack of explanatory capability that makes machine-learning arguably inappropriate for discretionary legal judgements has far less relevance for data analysis tasks designed to facilitate, rather than substitute for, expert human judgments, provided that the predictive models' accuracy can be evaluated. The third is inventive construction of training sets. Machine learning performance is typically very sensitive to training set size, but manual development of large training sets is unfeasible for many tasks. As a result, practical applications of machine learning often depend on devising or discovering naturally occurring potential training sets. Finally, Abood and Feltenberger (2018) illustrates that even prototype implementations need not be limited to a handful of examples but can be implemented at scale (roughly 10 million patents in the case of Abood and Feltenberger).

These four characteristics, which were relatively uncommon in the AI and Law research community at the time of Abood and Feltenberger's original publication, now represent an increasingly influential strand of AI and Law research<sup>23</sup>.

### **5 Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts (Nguyen et al., 2018). *Commentary by Jack G. Conrad***

In *Recurrent Neural Network-based Models for Recognizing Requisite and Effectuation Parts in Legal Texts*, Nguyen et al. train an RNN to identify and label two central components within Japanese legal documents: a "requisite part" and an "effectuation part" (Nguyen et al., 2018). What is noteworthy about this paper is that it is one of the first substantial neural network (a.k.a. deep learning) papers published by the *Artificial Intelligence and Law Journal* since the renaissance of neural approaches in the decade after 2010. Indeed, this was one of the reasons why the authors were invited to contribute an extended version of this work for the journal's special issue on Legal Data Analytics (Conrad and Branting, 2018). Because the authors were aware of the seminal role of their publication, they performed an extraordinary job of providing their audience with a thorough in-depth appreciation of how the layers of Bi-LSTM (Bi-directional Long Short-Term Memory) models are operationalized. They dedicate an entire section of their paper to this objective, and the superior quality of their diagrams is worth acknowledging as well. Not only do the authors depict the layers representing the neural architecture, but they also give their readers a more concrete understanding of how the different components interleave and operate collectively.

Since AI and Law papers from other regions of the world may not often refer to key components of legal sentences in terms of *requisite* and *effectuation* parts, it may be helpful to define them here. A requisite part of a sentence refers to a portion that is necessary for the achievement of a particular end, for example, "A person with limited resources, who uses fraudulent means to convince others that he is a person with substantial resources." By contrast, an effectuation part of a

<sup>23</sup> For example, roughly two thirds of the papers in ICAIL 2021 contained a data-analytic or corpus-based component.

sentence refers to the accomplishment of that end (or providing a practical means of achieving that end), for example, following the sentence above, “His/Her act may not be rescinded and may be disciplined.”

The task the authors pursue in this work is a form of semantic parsing for legal texts. The capability can be useful for a variety of applications. It can be leveraged to improve the quality of a legal retrieval system, allowing users to focus on relevant parts of the document instead of the text as a whole (Son et al., 2017). In a legal summarization system, identifying requisite and effectuation parts may help extract specific essential information (Ji et al., 2020). In addition, the capability can assist in improving the quality of a question answering (QA) system by exploiting the cause and effect relationship of legal texts or by exposing a model to sufficiently granular quantities of meaningful training examples (Tran et al., 2014). Indeed, the diversity of approaches investigated by Nguyen et al. anticipates the extensive largely QA-focused deep language models pursued in the research awarded the ICAIL 2021 best paper award<sup>24</sup> (Zheng et al., 2021).

The strengths of the paper, as described below, clearly illustrate why the work merits inclusion in this special commemorative issue. As one would expect of a paper of this scope, Nguyen et al. perform their experiments on both a Japanese and an English language data set: the Japanese data collection is the Japanese National Pension Law Requisite-Effectuation Recognition dataset [JPL-RRE]; the English language data collection is the Japanese Civil Code dataset [JCC-RRE].

The authors compare their approach to a non-neural baseline consisting of a sequence of Conditional Random Fields (CRFs) (Lafferty et al., 2001), but they also investigate three distinct variations of their Bi-LSTM architecture to run against the JPL-RRE (Japanese) dataset and five to run against the JCC-RRE (English) dataset. They initially harness sequences of singular Bi-LSTMs or Bi-LSTMs with external features from CRFs, and progress to using multi-layered Bi-LSTMs, finally performing experiments on one that uses a single multi-connected layer and another that uses two multi-connected layers. The latter offers the authors the advantages of avoiding duplicate data, performing iterative training cycles, and ultimately saving time and expense when training their models.

Going beyond what many papers on this topic accomplish, the authors present in meaningful detail the algorithms used by the underlying models, rather than simply citing some of the key papers from the formative Bi-LSTM, BERT, or other language model literature.

Collectively, Nguyen et al. propose several variations of neural network techniques for recognizing requisite and effectuation parts in legal text. First, they introduce a modified Bi-LSTM-CRF that permits them to use external features to recognize non-overlapping requisite-effectuation (RE) parts. Then, they propose the sequence of Bi-LSTM-CRF models and two types of multi-layer models to recognize overlapping requisite-effectuation (RE) parts, which include the Multi-layer-Bi-LSTM-CRF and the Multi-layer-BiLSTM-MLP-CRF (the latter, multi-layer perceptron variant eliminates redundant components and can reduce training time and redundant parameters). The authors’ techniques significantly outperform previous approaches and deliver state-of-the-art results on the JPL-

---

<sup>24</sup> The 18th International Conference on Artificial Intelligence and Law, <https://icaill.lawalgorithm.com.br>: see <http://www.iaail.org/?q=page/icaill-best-paper-awards-winners> for the Award Recipients

RRE collection (F1 of 88.81%  $\Rightarrow$  93.27%, +5.0%). Regarding the JCC-RRE data collection, their techniques outperform CRFs, a strong algorithmic approach for the sequence labeling task (F1 of 73.7%  $\Rightarrow$  78.24%, +6.2%). For the recognition of overlapping RE parts, the multi-layer models are desirable because they represent unified models which expedite the training and testing processes, but deliver competitive results compared to the sequence of Bi-LSTM-CRF models. Given the two types of multi-layer models, the Multi-Bi-LSTM-MLP-CRF solves limitations of the Multi-Bi-LSTM-CRF because it eliminates redundant components, thus offering a smaller model and faster training and testing times with no sacrifice in performance.

Another strength of Nguyen et al.’s work is its anticipation of the growing importance of *explainability* for models that historically have been considered opaque or “black boxes” (Barredo Arrieta et al., 2020; Stranieri et al., 1999). The authors pursue two means of addressing the lack of transparency and explainability typical of such neural models. First, they harness a more traditional NLP technique – Conditional Random Fields – together with their model to facilitate transparency. And second, they perform a series of error analyses where the incorrect tagging by the hybrid model allows their readers to gain insights into why these trained deep learning models make the kinds of decisions (and sometimes errors) they do.

Some members of the AI and Law community, including reviewers, have expressed concern over the publication of deep learning papers of this kind, due at least in part to the aforementioned transparency issues. In response, Nguyen et al. anticipated such reservations, and have addressed them directly through their hybrid deep learning and traditional NLP models that permit a view “under the hood” and error analyses which reveal both the strengths and deficiencies of the classification assignments involved. They have even proposed new training procedures that could close such performance gaps. By demonstrating such concern for quality and dedication to transparency, the authors have not only allayed the concerns of their critics, but delivered superior outcomes as well.

In pursuing these additional measures, the authors present state-of-the-art techniques and models to the research community, while serving as an inspiration to the next generation of AI and Law researchers. Nguyen et al. are nothing short of trail blazers who have established an incentive to adopt such deep language models, apply them to challenges of the legal domain, and reach still higher levels of performance.

## **6 Using machine learning to predict decisions of the European Court of Human Rights (Medvedeva et al., 2020a). *Commentary by Trevor Bench-Capon***

Since the early days of AI and Law, the idea that the ultimate goal might be to replace judges with computers has been regularly discussed (e.g. D’Amato (1977), van den Herik (1991), Verheij (2021)). In practice this has not been a common goal in AI and Law research: rather the emphasis has been that laid out in Buchanan and Headrick (1970), which is a candidate for the first true AI and Law paper, which suggested that “the computer modeling of legal reasoning would be a fruitful area for research”, so as to “understand legal reasoning and legal argument formation better”.

Thus the focus of early systems was predominately on the reasoning justifying an outcome rather than the predicted outcome itself. Case based systems in the style of HYPO (Rissland and Ashley, 1987) and CATO Aleven and Ashley (1995) produce the arguments that could be put forward for both sides but any evaluation was to be left to the user. Rule based systems following Sergot et al. (1986) did suggest an outcome, but the emphasis was on the explanation of how the recommendation was reached. It was also suggested (Bench-Capon and Sergot, 1988) that rule based systems could also produce the reasons to find for both sides and allow the user to decide which reasons should be preferred. IBP (Brüninghaus and Ashley (2003) and Ashley and Brüninghaus (2009)<sup>25</sup>) did use CATO style techniques to predict outcomes, but the prediction was accompanied by a detailed justification. Thus symbolic systems tended to focus on the rationale for a decision rather than the decision itself. The idea was that the systems should support rather than replace lawyers, and that a simple prediction unsupported by any reasons justifying it offered no real support. This point has recently been convincingly argued in Bex and Prakken (2021).

However, those interested in applying Machine Learning (ML) techniques to legal cases do tend to speak of prediction. Early examples are Groendijk and Oskamp (1993) and Pannu (1995) which used neural networks and genetic algorithms respectively. Even with ML, however, there was recognition of the need to justify the prediction: The Split-Up system (Stranieri et al., 1999)<sup>26</sup> placed great emphasis on its argumentation based explanation and Bench-Capon (1993) attempted to extract a set of rules from a neural network model. Interestingly the latter found that the rules applied by the system were not, despite the very high accuracy achieved, the correct set of rules governing the domain. This result was recently replicated for more contemporary ML techniques in Steging et al. (2021).

The use of ML techniques in AI and Law fell into abeyance for a while but, as discussed in Section 4, developments in ML techniques and the much increased availability of large amounts of data has led to an upsurge of interest in such techniques. Much of this has been directed towards legal tasks other than decision support and argumentation, but predicting case outcomes has received a good degree of attention. A very well publicised example was Aletras et al. (2016) which attracted a great deal of press attention<sup>27</sup>. Aletras et al. (2016) used a Support Vector Machine (Vapnik, 1999) to classify decision of the European Court of Human Rights (ECHR) and achieved an accuracy of 79%. The ECHR is a particularly attractive domain given the ready public availability of its decisions<sup>28</sup>, and it has been the subject of a number of subsequent investigations including Medvedeva et al. (2020a), Chalkidis et al. (2019), Medvedeva et al. (2020b), Kaur and Bozic (2019) and Medvedeva et al. (2022). In this section we will discuss Medvedeva et al. (2020a), which was the first paper on ML prediction of legal cases to appear in the journal. Like Aletras et al. (2016), Medvedeva et al. (2020a) used Support Vector Machine Linear Classifier to classify ECHR cases into violations and non-

---

<sup>25</sup> See Sartor et al. (2022), section 7, elsewhere in this issue.

<sup>26</sup> See section 10 of Governatori et al. (2022), elsewhere in this issue.

<sup>27</sup> See for example the *Guardian* article <https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-college-london-computer-scientists> which very much took the “computers can replace judges” line.

<sup>28</sup> <https://hudoc.echr.coe.int/eng#%20>

violations, and similar to Aletras et al. (2016) achieved an accuracy of 75%<sup>29</sup>. The paper describes three experiments, the first of which is the classification task performed in Aletras et al. (2016).

A key issue is what information to use for classification. The decisions that are available are, of course, written after the decision has been made, and in fact include the outcome of the case. Since any meaningful prediction would be done before the decision was made, it should only use material available before the case is heard, and so some of the reported decision must be excluded. A decision of the ECHR comprises a number of sections:

- *Introduction*, consisting of the title, date, Chamber, judges etc.;
- *Procedure*, describing the handling of the claim;
- *Facts*, consisting of two parts:
  - *Circumstances*, information on the applicant and events and circumstances that led to the claim;
  - *Relevant Law*, containing relevant provisions from legal documents other than the ECHR;
- *Law*, containing legal arguments of the Court;
- *Judgement*, the outcome of the case;
- *Dissenting/Concurring Opinions*, containing the additional opinions of judges.

Clearly not all of this can be used, as some sections, e.g. judgement, would make the classification task rather easy. Therefore only the procedure and facts sections were used. However, even here there is a problem: the facts are drafted by the decision maker after the decision is known and so may reflect the thinking that led to the decision itself. For many ECHR cases there is a statement of facts written before the hearing, the *Communicated Case*, sent to the potentially violating Government for a response. However, using the communicated cases rather than the facts from the final decisions led to a substantial (around 10%) degradation in accuracy in a series of experiments reported in Medvedeva et al. (2021), leading to the conclusion that “the experiments conducted in this paper show that performance seems to be substantially lower when forecasting future judgements compared to classifying decisions which were already made by the court”.

The second experiment in Medvedeva et al. (2020a) looked at a second problem for predicting cases. For ML the more cases the better. But case law evolves: as time progresses new factors may be introduced, different preferences emerge and landmark rulings will revise the significance of previous findings (e.g. Levi (1948) and ?). The experiment showed that “training on one period and predicting for another is harder than for a random selection of cases” and that the older the cases used, the greater the drop in performance. The experiments confirm that there is some concept change happening, but it does not indicate which cases may still be relied on: the landmark cases may occur at any time, and landmark cases for different aspects of the law will occur at different times. The problem that the current data may not be a reliable guide to the future is identified, but no solution seems possible without an analysis of the cases, which is what the data-centric approaches hoped to avoid.

---

<sup>29</sup> For comparison the earlier system of Ashley and Brüninghaus (2009), discussed elsewhere in this issue in Section 7 of Sartor et al. (2022), achieved 70% accuracy by combining ML and symbolic methods, .

The third experiment used only the names of the judges to predict the outcome. One would suppose that this should not be a good basis for prediction, but in fact the results achieved accuracies of between 60% and 70%, very similar to those achieved for communicated cases. This again suggests the need for an explanation of the prediction, so that the reasons can be made clear, so that the rationale of the decision can be checked (Bex and Prakken, 2021).

The topic of explanation is not discussed in Medvedeva et al. (2020a), but was discussed in Aletras et al. (2016) which reported a very similar experiment. There the explanation took the form of listing the most predictive topics, represented by the 20 most frequent words, listed in order of their SVM weight. One such list, for violation of Article 6 is “court, applicant, article, judgment, case, law, proceeding, application, government, convention, time, article convention, January, human, lodged, domestic, February, September, relevant, represented”. Given that many of the words seem likely to appear in every case, and the presence of several names of months, one wonders whether an acceptable rationale is being applied, or whether this is another example where reasonable accuracy is possible, even though the model does not reflect the law (Steging et al., 2021).

The work reported in Medvedeva et al. (2020a), and subsequent papers by this group represent a careful exploration of the possibility of classifying and predicting legal decisions using machine learning techniques unsupported by domain analysis. The experiments raise some reservations about the approach: using only information available before the hearing, and testing on cases later than the training set - both essential for any practical use of prediction - both lower performance. There is also a suspicion that the predictions are not based on sound law, indicated by the results obtained by only using the names of the judges and by the lack of explanation. The difficulties of prediction are well illustrated by the group’s JURI SAYS website<sup>30</sup> which charts the performance of the system month by month. Currently accuracy stands at only 58.3% for the last year and only 35.5% for the last month (although there was a score of 76% in February 2022). Part of the problem may simply be the number of violations - as pointed out in Verheij (2021), simply always predicting violation would outperform many ML systems (this would give an accuracy of 73.3% over the last 30 cases on 28/06/22).

The significant improvements in ML techniques, especially those that use natural language rather than feature vectors as input, coupled with the ready availability of large collections of legal decisions, make the idea of training a system to predict legal cases an attractive idea. The idea is worth exploring, and Medvedeva et al. (2020a) and subsequent papers provide an excellent exploration. The results of the exploration is not, however, entirely encouraging. Classifying legal cases is a very different problem from identifying which mushrooms are poisonous. The availability of input before the hearing and the constant evolution of case law are two difficulties. The lack of explanation also poses a significant problem, since legal decisions require an explicit justification.

This is not to say that ML techniques have no place in AI and Law: there are a number of other legal tasks to which it can be effectively applied, as illustrated in other sections of this paper: the identification of relevant patents (Section 4), segmentation of legal texts (Section 5), detecting potential unfair clauses in contracts (Section 8) and retrieving relevant law articles (Section 9). As well as these

---

<sup>30</sup> <https://www.jurisays.com/>, last accessed 28th June 2022.

specific tasks, there are many other possibilities for providing useful support for lawyers. However, it seems to me that predicting - or worse, deciding - legal cases is not a task which holds out much hope of practical success. Even with advances in ML, the role of AI is to support lawyers, not to replace them.

## 7 Scalable and explainable legal prediction (Branting et al., 2021).

*Commentary by Kevin Ashley*

A major shift in the history of the field of Artificial Intelligence and Law has been the move away from knowledge-based approaches towards machine learning and text analysis which has occurred during the Journal's third decade. The former involves manually representing legal rules and concepts to enable computational models to reason about legal problems, predict outcomes, and explain predictions. By contrast, machine learning induces features automatically from legal case texts with which to predict outcomes of textually described, previously unseen, legal problems (e.g. Medvedeva et al. (2020a), discussed in section 6)<sup>31</sup>. The extent to which a machine learning model can explain its prediction in terms meaningful to legal professionals is still, however, unclear. The pressing research question is how best to integrate legal knowledge and machine learning so that a system can both predict and explain<sup>32</sup>.

In Branting et al. (2021) Branting and his colleagues have reported a significant step toward answering that question with their work on semi-supervised case annotation for legal explanations (SCALE). They have 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 could predict outcomes of textually described cases, and identify the features that could help to explain the predictions, for example, by indicating the elements of the legal rule that have been satisfied or are still missing.

This is a major improvement over an alternative method for "explaining" legal predictions reported in section 3 of Branting et al. (2021), attention-based highlighting. There, Branting's team employed a Hierarchical Attention Network (HAN), a kind of neural network architecture (Yang et al., 2016), to predict outcomes of textually described cases. HANs assign network attention weights to portions of the input text measuring the extent to which the portions influenced the network's outcome. While, intuitively, these highlighted portions might explain the prediction, the team's experiments, involving novice and expert users in solving a legal problem, indicated that the participants "had difficulty understanding the connection between the highlighted text and the issue that they were supposed to decide" (Branting et al. (2021), p. 221).

---

<sup>31</sup> Machine learning has also been widely used for legal tasks other than outcome prediction. For example information retrieval (see section 8), document management (see Section 4), text recognition (see Section 5) and document analysis (see Section 9.) The influence of Machine Learning has been pervasive across the whole field of AI and Law.

<sup>32</sup> Explanation is also important in other applications of Machine learning (see Sections 9 and 8).

Consequently, Branting’s team turned to the five-step approach of SCALE. They manually annotated the findings sections of a representative sampling<sup>33</sup> of their collection of more than 16,000 World Intellectual Property Organization (WIPO) domain name dispute cases. The labelled sentences are then mapped automatically to all similar sentences in the Findings sections of a selection of unannotated cases. Two machine learning models are trained on this enlarged set of annotated cases: one to predict the tags that apply to the Fact and Contention sections of new cases and the other their outcomes on the basis of these tags. The “human-understandable tags generated by the first model can be used for case-based reasoning or other argumentation techniques or for predicting case outcomes”, and ultimately to assist in explaining the predictions.

The tags or labels are key. Text annotation involves marking-up texts of case decisions to identify instances of semantic tags or types of information, the concepts of interest in the texts. In SCALE, those labels represent the types of findings, legal issues, factors, and attributes that arise in the WIPO domain name dispute cases. For example, one element of the Rules for Uniform Domain Name Dispute Resolution Policy (UDRP rules) is a showing of bad faith. Under Rule 4(b)(iv) a finding of “using the domain name ... for commercial gain ... by creating a likelihood of confusion with the complainant’s mark” “shall be evidence of the registration and use of a domain name in bad faith.” Figure 6 of the article illustrates a sentence in a case findings section, “Such use constitutes bad faith under paragraph 4(b)(iv) of the Policy”. which has been annotated as a LEGAL.FINDING-BadFaith-Confusion4CommGain. Text spans are also annotated in terms of attributes capturing citations to UDRP rules and whether or not the span supports the issue, that is, its polarity.

Intuitively, one can appreciate the predictive significance of such a finding and, when highlighted, its utility in helping to explain a prediction by linking to the relevant legal rule. Training machine learning models to identify legal concepts such as findings, issues, and factors, would seem to be key in assisting human users to make the connection between the highlighted portions of text and the legal task they are meant to perform. To the extent that an approach like SCALE can identify findings, issues, and factors in other types of legal cases, it would help to connect modern text analytic techniques with knowledge-based computational models of case-based legal reasoning. If, in addition to predicting outcomes of textually described cases, text analytics could also identify the applicable factors, then case-based models that account for legal rules, underlying values, and case factors could assist machine learning in explaining and testing those predictions. See, for example, Grabmair (2016), Chorley and Bench-Capon (2005b) and Chorley and Bench-Capon (2005a).

Of course, the question for the future is whether the SCALE technique *can* be applied successfully to other legal domains. Branting acknowledges that, “WIPO cases, ... have a high degree of stylistic consistency in the language used in Findings sections” (Branting et al. (2021), p. 232). Beyond “stylistic” consistency, it is an open question how factually diverse the WIPO UDRP arbitration cases really are. Do they involve more than a relatively small number of oft-repeated issues in

---

<sup>33</sup> The sample comprised only 25 representative decisions (0.156% of the entire corpus of 16,024 WIPO decisions). That such impressive results can be achieved with such a small sample is encouraging with respect to the feasibility of the approach.

oft-repeated factual contexts? For example, it has been noted that in the “body of [UDRP] precedent ... many disputes involve similar issues and facts.” (Kelley (2002), fn 120) That might suggest the UDRP cases are less factually diverse than litigated, non-arbitration cases from other legal domains. At least two machine learning models have been trained with varying degrees of success to identify trade secret misappropriation factors in case summaries (Ashley and Brüninghaus (2009)) or full texts of cases (Falakmasir and Ashley (2017)). Using a current language model like Legal-BERT, pretrained on legal vocabulary from a large case law corpus, should lead to performance gains in learning to identify factors. See Zheng et al. (2021).

Time and research will tell whether Branting is correct in suggesting that an approach like SCALE can identify findings, issues, and factors in cases beyond the WIPO domain name domain. If true, it provides the means for linking two main thrusts in the history of AI and Law research, knowledge-based and text analytical, as reflected in thirty years of the *Journal*.

## **8 Unsupervised law article mining based on deep pre-trained language representation models with application to the Italian civil code (Tagarelli and Simeri, 2021). *Commentary by Serena Villata.***

This paper presents an interesting empirical approach to AI and Law. More precisely, the focus of the paper is the construction of a BERT-based solution for legal text in Italian. The authors present a novel deep learning framework named LamBERTa, trained on Italian civil-law codes. Two particular features make this work an exemplary contribution in the field of AI and Law.

First, the paper addresses the task of creating a deep pre-trained language model for legal text, following the example of the popular BERT learning framework (Devlin et al., 2019). BERT (Bidirectional Encoder Representations from Transformers) has been designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. Roughly, BERT learns contextual embeddings for words in the training text. In this paper, the authors define a model called LamBERTa (Law article mining based on BERT architecture) which fine-tuned an Italian pre-trained BERT on the Italian civil code text. The overall Natural Language Processing (NLP) task is Text Classification with the goal of addressing automatic law article retrieval based on civil-law-based corpora, which given a natural language query, will predict the most relevant article(s) from the Italian Civil Code.

Second, the paper addresses the task of creating a pre-trained language model for legal text in Italian, which is a language low in resources and lacks useful training materials such as annotated data. The proposed LamBERTa framework is completely specified using the Italian Civil Code (ICC) as the target legal corpus. This means that the paper provides a valuable contribution to the community of AI and Law not only from the perspective of the legal text application but also concerning the language targeted by the application.

Furthermore, several technical challenges arise considering the above features. First, the task being tackled is very challenging, as it is characterized not only by a high number of classes, but also by the need to build suitable training sets given the lack of test query benchmarks for the Italian legal article retrieval task.

Second, given that these neural models are basically black-boxes whose results are hard to interpret, the authors investigate the explainability of the LamBERTa models to understand how they form complex relationships between the words.

Overall, LamBERTa leverages the Transformer paradigm allowing for processing all textual tokens simultaneously by forming direct connections between individual elements through a mechanism known as *attention* (e.g. Bahdanau et al. (2015) and Kim et al. (2017)). Often proposed as a mechanism to add transparency to neural models, attention assigns weights to the input features (i.e., portions of the text) based on their importance with respect to some task. In the case of LamBERTa, the attention mechanism enables the model to understand how the words relate to each other in the context of the sentence, returning composite representations that the model can employ to make the classification. To train the LamBERTa models, the authors rely on the key idea of combining portions of each article from the Italian Civil Code to generate the training units for the model. Different unsupervised schemes for data labeling the ICC articles have been tested to create the training sets for the LamBERTa models. The combinations are needed since each article is usually comprised of only a few sentences, whereas the method needs a relatively large number of training units (32 are used here).

To evaluate the effectiveness of the LamBERTa models for the law article retrieval task on the ICC, the authors conducted a comparative analysis with state-of-the-art text classifiers based on deep learning architectures:

- a bidirectional Long Short Term Memory (LSTM) model as sequence encoder (BiLSTM, Liu et al. (2016)),
- a convolutional neural model with multiple filter widths for text encoding and classification (TextCNN, Kim (2014)),
- a bidirectional LSTM with a pooling layer on the last sequence output ((TextRCNN), Lai et al. (2015)),
- a Seq2Seq (Du and Huang (2018) and Bahdanau et al. (2015)) model with attention mechanism (Seq2Seq-A), and
- a transformer model for text classification adapted from Vaswani et al. (2017).

The experimental setting designed by the authors is grounded on two main requirements: the robustness of the evaluation of the competing approaches, and a fair comparison between the competing methods and LamBERTa. To ensure a robust evaluation, the authors carried out an extensive parameter-tuning phase for each of the competing methods, by varying all main parameters (e.g., drop out probability, maximum sentence length, batch size, number of epochs) within recommended range values. To ensure a fair comparison between the competing methods and LamBERTa, the authors noted that each of the competing models (with the exception of Transformer), require word vector initialization. To satisfy this need, the models were firstly provided with Italian Wikipedia pre-trained Glove embeddings, and secondly, each model was fine-tuned over the individual ICC texts following the same data labeling schemes as the LamBERTa models.

Experimental results demonstrate the superiority of LamBERTa models against the selected competitors for ICC case law retrieval construed as a classification task. In addition, these results show much promise for the development of AI tools to ease the work of jurists, which is one of the main focuses of the AI and Law research area.

## 9 Detecting and explaining unfairness in consumer contracts through memory networks (Ruggeri et al., 2022). *Commentary by Serena Villata.*

This paper presents an important contribution to the field of AI and Law by proposing a novel Natural Language Processing (NLP) approach to the identification of unfair clauses in a class of consumer contracts. More precisely, terms of service (ToS) are consumer contracts governing the relation between providers and users. In this context, clauses in ToS which cause a significant imbalance in the parties' rights and obligations are deemed unfair according to Consumer Law. Despite actions taken by the enforcers to control the quality of ToS, online providers still tend to employ unfair clauses in these documents. This became a major issue with the General Data Protection Regulation (GDPR) regulation<sup>34</sup>. Given the number of ToS published on the Web, it is not realistic to think of using manual checking to assess the fairness of these documents. For this reason, in line with the recent developments in AI and NLP, the authors propose to address this issue automatically, relying on a neural network architecture.

However, the contribution of this paper is not limited to the novel approach to the identification and classification of unfair clauses in ToS documents. The second main outcome of the paper is the definition of an approach which is transparent with respect to the predictions made by the machine learning system. This issue is of particular importance given the nature of the information this system is dealing with, i.e., legal text. Legal knowledge is, in general, difficult to understand for consumers, who cannot be expected to pinpoint unfair or unlawful conduct in ToS documents. However, they can only rely on a automated system if this system can "explain" the rationale behind the advice, i.e. indicate not only that a certain clause of a ToS *is* unfair, but *why* it is unfair. This is the main motivation for the contribution presented in this paper.

This paper is based on a project devoted to the AI and Law field, i.e., the CLAUDETTE project<sup>35</sup> which aimed at the empowerment of the consumer via AI, by investigating ways to automate reading and legal assessment of online consumer contracts and privacy policies with NLP techniques. The final goal was to evaluate the compliance of these contracts and privacy policies with EU consumer and data protection law. The results presented in this paper are the latest achievements obtained in the framework of this project.

Two main contributions are presented in this paper: first, the creation of an annotated dataset for the identification and classification of unfair clauses in ToS documents, and second, the architecture designed to automatically address this task.

Concerning the creation of an annotated dataset, this is an important contribution of the paper as the lack of existing annotated linguistic legal resources is a key issue in the AI and Law field as it slows the development of automatic approaches to ease and support jurists' activity. The dataset is composed of 100 ToS documents. These documents are standard terms available on provider's websites for review by potential and current consumers. The ToS collected in the dataset were downloaded and analyzed over a period of eighteen months by four legal experts. Potentially unfair clauses were tagged using the guidelines carefully defined

---

<sup>34</sup> See <https://gdpr-info.eu/>

<sup>35</sup> Contissa et al. (2018). The project website is at <http://claudette.eui.eu/>

by the authors. In particular, the annotation task aimed at identifying in the ToS the following clauses:

- (i) liability exclusions and limitations,
- (ii) the provider’s right to unilaterally remove consumer content from the service,
- (iii) the provider’s right to unilaterally terminate the contract,
- (iv) the provider’s right to unilaterally modify the contract and/or the service, and
- (v) arbitration on disputes arising from the contract.

In the resulting annotated resources, out of the 21063 sentences in the corpus, 2346 sentences were labeled as containing a potentially or clearly unfair clause. Arbitration clauses are the least common (they appear in only 43 documents), whilst all other categories appear in at least 83 out of the 100 documents.

Concerning the architecture to automatically identify and classify the ToS documents along with the set of five unfair clauses identified in the corpus, and keeping an eye on the transparency of the proposed approach, the authors decided to rely on a promising approach associating explanations to the outcomes of neural classifiers. This model is called Memory-Augmented Neural Networks (MANNs) (Sukhbaatar et al., 2015), and it combines the learning strategies developed in the machine learning literature for inference with a memory component that can be read and written to. The explanations can then be given in terms of rationales, i.e., ad-hoc justifications written by the annotating legal experts to motivate their conclusion to label a given clause as “unfair”. To do so, the authors train a MANN classifier to identify unfair clauses by using as facts the rationales behind the unfairness labels, and then a possible explanation of an unfairness prediction is generated based on the list of rationales used by the MANN. The huge advantage of generating this kind of explanation is that it answers a user’s need from a dialectical and communicative viewpoint, allowing her to understand the explanation she is provided with. The obtained results show that even a simple MANN architecture is sufficient to show improved performance over traditional knowledge-agnostic models, including a current state-of-the-art SVM solution (Lippi et al., 2019), previously developed in the CLAUDETTE project. These results confirm the importance of transparent machine learning approaches to information extraction and classification applied to the legal domain, and in particular, to legal documents, providing a first but important step towards the development of further, possibly multilingual, legal NLP systems.

## References

- Steven Abney. Statistical methods and linguistics. In Judith Klavans and Philip Resnik, editors, *The balancing act: Combining symbolic and statistical approaches to language*, pages 1–23. MIT Press, 1996.
- Aaron Abood and Dave Feltenberger. Automated patent landscaping. *Artificial Intelligence and Law*, 26(2):103–125, 2018.
- Keerthi Adusumilli, Bradford Brown, Joey Harrison, Matthew Koehler, Jason Kutarnia, Shaun Michel, Max Olivier, Craig Pfeifer, Zoryanna Slater, William Thompson, et al. The structure and dynamics of modern United States Federal case law. *Frontiers in Physics*, page 755, 2022.

- Tommaso Agnoloni, Meritxell Fernández Barrera, Maria Teresa Sagri, Daniela Tiscorni, and Giulia Venturi. When a framenet-style knowledge description meets an ontological characterization of fundamental legal concepts. In Monica Palmirani, Ugo Pagallo, Pompeu Casanovas, and Giovanni Sartor, editors, *AI Approaches to the Complexity of Legal Systems*, Lecture Notes in Computer Science, pages 93–112. Springer, 2009.
- Latifa Al-Abdulkarim, Katie Atkinson, and Trevor Bench-Capon. A methodology for designing systems to reason with legal cases using ADFs. *Artificial Intelligence and Law*, 24(1):1–49, 2016.
- Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro, and Vasileios Lampos. Predicting judicial decisions of the European Court of Human Rights: *PeerJ Computer Science*, 2:e93, 2016.
- Vincent Aleven and Kevin D Ashley. Doing things with factors. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 31–41, 1995.
- Giulia Andrighetto, Rosaria Conte, Eunat Mayor Villalba, and Giovanni Sartor, editors. *Artificial Intelligence and Law: Special Issue on Simulations, Norms and Laws*, volume 21:1. 2013.
- Michał Araszkiwicz, Enrico Francesconi, and Tomasz Zurek. Identification of contradictions in regulation. In *Proceedings of JURIX 2021*, pages 151–160, 2021.
- Kevin D Ashley and Stefanie Brüninghaus. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17(2):125–165, 2009.
- Tara Athan, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. Legalruleml: Design principles and foundations. In *Reasoning Web International Summer School*, pages 151–188. Springer, 2015.
- Shaun Azzopardi, Gordon J Pace, Fernando Schapachnik, and Gerardo Schneider. Contract automata. *Artificial Intelligence and Law*, 24(3):203–243, 2016.
- Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015.
- Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Benetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020.
- Trevor Bench-Capon. Neural networks and open texture. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 292–297, 1993.
- Trevor Bench-Capon and Frans Coenen. Isomorphism and legal knowledge based systems. *Artificial Intelligence and Law*, 1(1):65–86, 1992.
- Trevor Bench-Capon and Sanjay Modgil. Norms and value based reasoning: justifying compliance and violation. *Artificial Intelligence and Law*, 25(1):29–64, 2017.
- Trevor Bench-Capon and Marek Sergot. Towards a rule-based representation of open texture in law. In C. Walter, editor, *Computer Power and Legal Language: The Use of Computational Linguistics, Artificial Intelligence, and Expert Systems in the Law*, pages 39–60. Quorum Books, New York, 1988.
- Floris Bex and Henry Prakken. On the relevance of algorithmic decision predictors for judicial decision making. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 175–179, 2021.

- Carlo Biagioli. Towards a legal rules functional micro-ontology. In *Proceedings of the 1st International Workshop on Legal Ontologies, LEGONT 97*, 1997.
- Carlo Biagioli. *Modelli Funzionali delle Leggi. Verso testi legislativi autoesplicativi.*, volume 6 of *Legal Information and Communications Technologies Series*. European Press Academic Publishing, 2009.
- Avrim Blum. Random projection, margins, kernels, and feature-selection. In *International Statistical and Optimization Perspectives Workshop: Subspace, Latent Structure and Feature Selection*, pages 52–68. Springer, 2005.
- Guido Boella, Luigi Di Caro, Llio Humphreys, Livio Robaldo, Piercarlo Rossi, and Leendert van der Torre. Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. *Artificial Intelligence and Law*, 24(3):245–283, 2016.
- Alexander Boer, Rinke Hoekstra, Emile de Maat, Erik Hupkes, Fabio Vitali, and Monica Palmirani. CEN workshop agreement ‘Open XML Interchange Format for legal and legislative resources’. Technical Report CWA 15710:2010 E, CEN European Committee for Standardization, 2010.
- L. Karl Branting, editor. *Workshop on Law and Big Data*. International Association for AI and Law, 2015.
- L. Karl Branting, editor. *Workshop on Legal Text, Document, and Corpus Analytics*. University of San Diego Law School, 2016.
- L Karl Branting. Data-centric and logic-based models for automated legal problem solving. *Artificial Intelligence and Law*, 25(1):5–27, 2017.
- L Karl Branting, Craig Pfeifer, Bradford Brown, Lisa Ferro, John Aberdeen, Brandy Weiss, Mark Pfaff, and Bill Liao. Scalable and explainable legal prediction. *Artificial Intelligence and Law*, 29(2):213–238, 2021.
- Stefanie Brüninghaus and Kevin D Ashley. Predicting outcomes of case based legal arguments. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 233–242. ACM, 2003.
- Joanna J Bryson, Mihailis E Diamantis, and Thomas D Grant. Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law*, 25(3):273–291, 2017.
- Bruce G Buchanan and Thomas E Headrick. Some speculation about artificial intelligence and legal reasoning. *Stanford Law Review*, 23(1):40, 1970.
- Pompeu Casanovas, Monica Palmirani, Silvio Peroni, Tom Van Engers, and Fabio Vitali. Semantic web for the legal domain: the next step. *Semantic Web*, 7(3): 213–227, 2016.
- Ilias Chalkidis and Dimitrios Kamps. Deep learning in law: early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence and Law*, 27(2):171–198, 2019.
- Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. Neural legal judgment prediction in English. *arXiv preprint arXiv:1906.02059*, 2019.
- Seokkyu Choi, Hyeonju Lee, Eunjeong Park, and Sungchul Choi. Deep learning for patent landscaping using transformer and graph embedding. *Technological Forecasting and Social Change*, 175:121413, 2022.
- Alison Chorley and Trevor Bench-Capon. Agatha: Using heuristic search to automate the construction of case law theories. *Artificial Intelligence and Law*, 13(1): 9–51, 2005a.
- Alison Chorley and Trevor Bench-Capon. An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelli-*

- gence and Law*, 13(3):323–371, 2005b.
- Jack G. Conrad and L. Karl Branting, editors. *Artificial Intelligence and Law: Special Issue on Legal Text Analytics*, volume 26:2. 2018.
- Giuseppe Contissa, Francesca Lagioia, Marco Lippi, Hans-Wolfgang Micklitz, Przemyslaw Palka, Giovanni Sartor, and Paolo Torroni. Towards consumer-empowering artificial intelligence. In *International Joint Conference on Artificial Intelligence*, pages 5150–5157, 2018.
- Hans Crombag. On the artificiality of Artificial Intelligence. *Artificial Intelligence and Law*, 2(1):39–49, 1993.
- Faraz Dadgostari, Mauricio Guim, Peter A Beling, Michael A Livermore, and Daniel N Rockmore. Modeling law search as prediction. *Artificial Intelligence and Law*, 29(1):3–34, 2021.
- Anthony D’Amato. Can/should computers replace judges? *Georgia Law Review*, 11:11–36, 1977.
- Lieven de Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4): 1253–1278, 2000.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Volume 1*, pages 4171–4186, Minneapolis, MN, USA, 2019. Association for Computational Linguistics.
- Changshun Du and Lei Huang. Text classification research with attention-based recurrent neural networks. *International Journal of Computers Communications and Control*, 13(1):50–61, 2018.
- D. Engstrom, D.F. Ho, C. Sharkey, and M.F. Cuellar. Government by algorithm: Artificial Intelligence in Federal Administrative Agencies. Technical Report ACUS-AI-Report, Stanford Law and Policy Lab, 2010.
- Mohammad Falakmasir and Kevin D Ashley. Utilizing Vector Space Models for identifying legal factors from text. In *Proceedings of JURIX 2017*, pages 183–192, 2017.
- Jerry A Fodor. *The language of thought*. Harvard University Press, 1975.
- Enrico Francesconi. The “norme in rete” project: Standards and tools for Italian legislation. *International Journal of Legal Information*, 34(2):358–376, 2006.
- Enrico Francesconi. A description logic framework for advanced accessing and reasoning over normative provisions. *Artificial intelligence and Law*, 22(3):291–311, 2014.
- Yoav Goldberg and Omer Levy. word2vec explained: deriving Mikolov et al.’s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.
- Guido Governatori, Florian Idelberger, Zoran Milosevic, Régis Riveret, Giovanni Sartor, and Xiwei Xu. On legal contracts, imperative and declarative smart contracts, and blockchain systems. *Artificial Intelligence and Law*, 26(4):377–409, 2018.
- Guido Governatori, , Michał Bart Verheij Araszkievicz, Trevor Bench-Capon, Enrico Francesconi, and Grabmair Matthias. Thirty years of AI and Law: The first decade. *Artificial Intelligence and Law*, 30(4), 2022.

- Matthias Grabmair. *Modeling Purposive Legal Argumentation and Case Outcome Prediction using Argument Schemes in the Value Judgment Formalism*. PhD thesis, University of Pittsburgh, 2016.
- Cees Groendijk and Anja Oskamp. Case recognition and strategy classification. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 125–132, 1993.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Douglas R Hofstadter. *Gödel, Escher, Bach*. Harvester Press London, 1979.
- Wesley N Hohfeld. Some fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal*, 23(1):16–59, 1913.
- David I Holmes and Judit Kardos. Who was the author? an introduction to stylometry. *Chance*, 16(2):5–8, 2003.
- John Horty. Reasoning with dimensions and magnitudes. *Artificial Intelligence and Law*, 27(3):309–345, 2019.
- John F Horty. Reasons and precedent. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, pages 41–50, 2011.
- John F. Horty. Modifying the reason model. *Artificial Intelligence and Law*, 29: 271–83, 2021.
- John F Horty and Trevor Bench-Capon. A factor-based definition of precedential constraint. *Artificial Intelligence and Law*, 20(2):181–214, 2012.
- Bipin Indurkha, Jaap Hage, and Bartosz Brozek, editors. *Artificial Intelligence and Law: Special Issue on Machine Law*, volume 25:3. 2017.
- Donghong Ji, Peng Tao, Hao Fei, and Yafeng Ren. An end-to-end joint model for evidence information extraction from court record documents. *Information Processing and Management*, 57(6):102305, 2020.
- Andrew Jones and Marek Sergot. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law*, 1(1):45–64, 1992.
- A Kaur and B Bozic. Convolutional neural network-based automatic prediction of judgments of the European Court of Human Rights. In *27th AIAI Irish Conference on AI and Cognitive Science*, pages 458–469. CEUR 2563, 2019.
- Patrick D Kelley. Emerging patterns in arbitration under the uniform domain-name dispute-resolution policy. *Berkeley Tech. LJ*, 17:181, 2002.
- Yoon Kim. Convolutional neural networks for sentence classification, 2014.
- Yoon Kim, Carl Denton, Luong Hoang, and Alexander M Rush. Structured attention networks. *arXiv preprint arXiv:1702.00887*, 2017.
- Maciej Klodawski. Referring phrases with deictic indication and the issue of comprehensibility of texts of normative acts: The case of Polish codes. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, 34:497–524, 2021.
- Robert Kowalski and Akber Dato. Logical English meets legal English for swaps and derivatives. *Artificial Intelligence and Law*, pages 1–35, 2021.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*, pages 282–289, 2001.
- Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *28th AAAI conference on Artificial Intelligence*, 2015.

- Greg Leibon, Michael Livermore, Reed Harder, Allen Riddell, and Dan Rockmore. Bending the law: geometric tools for quantifying influence in the multinet network of legal opinions. *Artificial Intelligence and Law*, 26(2):145–167, 2018.
- Leonardo Lesmo. The Turin University parser at Evalita 2009. In *Proceedings of EVALITA 2009*, 2009.
- Edward Levi. An introduction to legal reasoning. *The University of Chicago Law Review*, 15(3):501–574, 1948.
- Marco Lippi, Przemysław Pałka, Giuseppe Contissa, Francesca Lagioia, Hans-Wolfgang Micklitz, Giovanni Sartor, and Paolo Torroni. CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. *Artificial Intelligence and Law*, 27(2), 2019.
- Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 2873–2879, 2016.
- Samhar Mahmoud, Nathan Griffiths, Jeroen Keppens, Adel Taweel, Trevor Bench-Capon, and Michael Luck. Establishing norms with metanorms in distributed computational systems. *Artificial Intelligence and Law*, 23(4):367–407, 2015.
- Mitch Marcus, editor. *HLT '02: Proceedings of the Second International Conference on Human Language Technology Research*. Morgan Kaufmann Publishers Inc., 2002.
- Masha Medvedeva, Michel Vols, and Martijn Wieling. Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28(2):237–266, 2020a.
- Masha Medvedeva, Ahmet Üstun, Xiao Xu, Michel Vols, and Martijn Wieling. Automatic judgement forecasting for pending applications of the European Court of Human Rights. In *Proceedings of the 5th workshop on Automated Semantic Analysis of Information in Legal text*, 2021.
- Masha Medvedeva, Martijn Wieling, and Michel Vols. Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law*, pages 1–18, 2022.
- MashaM Medvedeva, Xiao Xu, Michel Vols, and Martijn Wieling. JURI SAYS: An automatic judgement prediction system for the European Court of Human Rights. In *Proceedings of JURIX 2020*, pages 277–280, 2020b.
- Leora Morgenstern. Toward automated international law compliance monitoring (tailcm). Technical report, Leidos Hldingso Inc, Reston Va, 2014.
- Adeline Nazarenko, François Lévy, and Adam Wyner. A pragmatic approach to semantic annotation for search of legal texts—an experiment on GDPR. In *Proceedings of JURIX 2021*, pages 23–32. IOS Press, 2021.
- Truong-Son Nguyen, Le-Minh Nguyen, Satoshi Tojo, Ken Satoh, and Akira Shimazu. Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts. *Artificial Intelligence and Law*, 26(2):169–199, 2018.
- Adam R Pah, David L Schwartz, Sarath Sanga, Zachary D Clopton, Peter DiCola, Rachel Davis Mersey, Charlotte S Alexander, Kristian J Hammond, and Luís A Nunes Amaral. How to build a more open justice system. *Science*, 369(6500): 134–136, 2020.
- Monica Palmirani. Lexdatafication: Italian legal knowledge modelling in Akoma Ntoso. In *AI Approaches to the Complexity of Legal Systems XI-XII*, pages 31–47. Springer, 2020.

- Ananddeep S Pannu. Using genetic algorithms to inductively reason with cases in the legal domain. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 175–184, 1995.
- Henry Prakken. A formal analysis of some factor-and precedent-based accounts of precedential constraint. *Artificial Intelligence and Law*, 29(4):559–585, 2021.
- Adam Rigoni. An improved factor based approach to precedential constraint. *Artificial Intelligence and Law*, 23(2):133–160, 2015.
- Adam Rigoni. Representing dimensions within the reason model of precedent. *Artificial Intelligence and Law*, 26:1–22, 2018.
- Edwina L Rissland and Kevin D Ashley. A case-based system for Trade Secrets law. In *Proceedings of the 1st International Conference on Artificial Intelligence and Law*, pages 60–66, 1987.
- Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.
- Federico Ruggeri, Francesca Lagioia, Marco Lippi, and Paolo Torroni. Detecting and explaining unfairness in consumer contracts through memory networks. *Artificial Intelligence and Law*, 30(1):59–92, 2022.
- Ali Sadeghian, Laksshman Sundaram, Daisy Zhe Wang, William F Hamilton, Karl Branting, and Craig Pfeifer. Automatic semantic edge labeling over legal citation graphs. *Artificial Intelligence and Law*, 26(2):127–144, 2018.
- Giovanni Sartor. Fundamental legal concepts: A formal and teleological characterisation. *Artificial Intelligence and Law*, 14(1-2):101–142, 2006.
- Giovanni Sartor, Monica Palmirani, Enrico Francesconi, and Maria Angela BIASIOTTI. *Legislative XML for the semantic web: principles, models, standards for document management*. Springer Science & Business Media, 2011.
- Giovanni Sartor, Michał Araszkiwicz, Katie Atkinson, Trevor Bench-Capon, Floris Bex, Tom van Engers, Enrico Francesconi, Henry Prakken, and Giovanni Sileno. Thirty years of AI and Law: The second decade. *Artificial Intelligence and Law*, 30(4):This issue, 2022.
- Marek J Sergot, Fariba Sadri, Robert A Kowalski, Frank Kriwaczek, Peter Hammond, and H Terese Cory. The British Nationality Act as a logic program. *Communications of the ACM*, 29(5):370–386, 1986.
- David B Skalak and Edwina L Rissland. Arguments and cases: An inevitable intertwining. *Artificial Intelligence and Law*, 1(1):3–44, 1992.
- Sheikh M Solaiman. Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. *Artificial Intelligence and Law*, 25(2):155–179, 2017.
- Nguyen Truong Son, Viet-Anh Phan, and Le-Minh Nguyen. Recognizing entailments in legal texts using sentence encoding-based and decomposable attention models. In *COLIEE 2017, The 4th Competition on Legal Information Extraction and Entailment*, 2017.
- Cor Steging, Silja Renooij, and Bart Verheij. Discovering the rationale of decisions: towards a method for aligning learning and reasoning. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 235–239, 2021.
- Andrew Stranieri, John Zeleznikow, Mark Gawler, and Bryn Lewis. 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, 1999.

- Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. *Advances in neural information processing systems*, 28, 2015.
- Andrea Tagarelli and Andrea Simeri. Unsupervised law article mining based on deep pre-trained language representation models with application to the Italian Civil Code. *Artificial Intelligence and Law*, pages 1–57, 2021.
- Oanh Thi Tran, Bach Xuan Ngo, Minh Le Nguyen, and Akira Shimazu. Answering legal questions by mining reference information. In Yukiko Nakano, Ken Satoh, and Daisuke Bekki, editors, *New Frontiers in Artificial Intelligence*, pages 214–229. Springer International Publishing, 2014.
- Anthony Trippe. Guidelines for preparing patent landscape reports. Technical report, World Intellectual Property Organization, Geneva, 2015.
- H. Jaap van den Herik. *Kunnen Computers Rechtspreken? (Can Computers Decide Legal Cases?)*. Gouda Quint, Arnhem, 1991.
- Marc Van Opijnen. European Case Law Identifier: Indispensable asset for Legal Information Retrieval. In *From Information to Knowledge: Online access to legal information: methodologies, trends and perspectives*, pages 91–103. IOS Press, 2011.
- Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- Bart Verheij. A second coffeehouse conversation on the Van den Herik Test. In *Liber amicorum ter gelegenheid van het emeritaat van prof. dr. Jaap van den Herik, hoogleraar Recht en Informatica te Leiden*, pages 101–114. Ipskamp Publishing, 2021.
- Adam Wyner and L. Karl Branting, editors. *Workshop on Applying Human Language Technology to the Law*. International Association for AI and Law, 2011.
- Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*, pages 1480–1489, 2016.
- Lucia Zheng, Neel Guha, Brandon R Anderson, Peter Henderson, and Daniel E. Ho. When does pretraining help? Assessing self-supervised learning for law and the CaseHOLD dataset of 53,000+ legal holdings. In *Proceedings of the 18th International Conference on Artificial intelligence and Law*, pages 159–168, 2021.