Artificial Intelligence and Law (2022) 30:521–557
https://doi.org/10.1007/s10506-022-09326-7

Thirty years of *Artificial Intelligence and Law*: the second decade

Giovanni Sartor[1,2] · Michał Araszkiewicz[3] · Katie Atkinson[4] · Floris Bex[5,6] · Tom van Engers[7] · Enrico Francesconi[8,9] · Henry Prakken[5,10] · Giovanni Sileno[7] · Frank Schilder[11] · Adam Wyner[12] · Trevor Bench-Capon[4]

Accepted: 21 July 2022 / Published online: 8 August 2022

Abstract
The first issue of *Artificial Intelligence and Law* journal was published in 1992. This paper provides commentaries on nine significant papers drawn from the Journal's second decade. Four of the papers relate to reasoning with legal cases, introducing contextual considerations, predicting outcomes on the basis of natural language descriptions of the cases, comparing different ways of representing cases, and formalising precedential reasoning. One introduces a method of analysing arguments that was to become very widely used in AI and Law, namely argumentation schemes. Two relate to ontologies for the representation of legal concepts and two take advantage of the increasing availability of legal corpora in this decade, to automate document summarisation and for the mining of arguments.

Keywords Reasoning with cases · Argumentation · Ontologies · Document summarisation

1 Introduction

This article provides commentaries on nine papers taken from the second decade of *AI and Law* journal. They relate to a range of topics, all of which have been important to AI and Law throughout its history. These papers reflect particular points in their development.

Four of the papers relate to reasoning with legal cases. Sound foundations had been provided though systems such as HYPO (Rissland and Ashley 1987) and CATO (Aleven and Ashley 1995). These systems, however, did not consider context: cases were treated without regard to date or jurisdiction. Moreover, these systems produced arguments for and against but did not attempt to decide which were

✉ Trevor Bench-Capon
tbc@csc.liv.ac.uk

Extended author information available on the last page of the article

stronger. Two of the papers discussed here address these issues. The first, Hafner and Berman (2002), commented on by Giovanni Sartor, considers how temporal and procedural context can make a difference and how the purposes of law can be used to suggest how conflicts should be resolved. This last idea was particularly influential, giving rise to a substantial body of research on how social values can determine outcome (e.g. Bench-Capon and Sartor (2003) and Grabmair and Ashley (2011)). The second reasoning with cases paper, Ashley and Brüninghaus (2009), commented on by Henry Prakken, not only identifies conflicts, but attempts to resolve them. Moreover, Ashley and Brüninghaus (2009) starts with cases given in natural language rather than already analysed into dimensions (as in HYPO) or factors (as in CATO). The last aspect, moving from natural language to outcome, has been the subject of much interest in recent years (e.g. Medvedeva et al. (2020)[1]). Work on reasoning with cases has given rise to a variety of ways of representing the required knowledge and the associated arguments. A special issue (Atkinson 2012), commented on by its editor, Katie Atkinson, brought together several of the methods amd applied them to the same specific case, *Popov v Hayashi*, so that they could be compared. The fourth of the papers, Horty and Bench-Capon (2012), also commented on by Giovanni Sartor, gives a formal characterisation of precedential constraint, distinguishing two models, the results model and the reason model.

Case Based Reasoning involves argumentation. Throughout the first decade much of the modelling of argumentation had been carried out through dialogue [(e.g. Gordon (1993) and Hage et al. (1993)[2]]. Verheij (2003c), commented on by Floris Bex, introduces a different way of modelling arguments using argument *schemes*. This approach, based on the work of Doug Walton,[3] was to become a central method for modelling arguments in AI and Law (e.g. Bex et al. (2003), Gordon and Walton (2009), Bench-Capon and Prakken (2010), Grabmair (2017) and Atkinson and Bench-Capon (2021)).

The first of the ontologies papers (Breuker et al. 2004) describes two legal core ontologies developed at the Leibniz Center for Law: one describes and explains dependencies between types of knowledge in legal reasoning while the second captures the main concepts in legal information processing. The Leibnitz Center for Law at the University of Amsterdam was a pioneer of AI and Law ontologies, and the commentary is written by Tom van Engers and Giovanni Sileno, both of whom have worked for the Leibnitz Center. The other paper (Sartor 2006), commented on by Enrico Francesconi and Michał Araszkiewicz, gives an account of fundamental legal concepts intended to clarify some aspects of their logical structure and the role they play in legal reasoning.

The final two papers make use of corpora available through the World Wide Web. Hachey and Grover (2006), commented on by Frank Schilder, used a corpus of judgments of the UK House of Lords to investigate techniques for summarising such

1 See Sect. 6 of Villata et al. (2022), elsewhere in this issue.

2 See Sects. 6 and 7 of Governatori et al. (2022), elsewhere in this issue.

3 See Sect. 4 of Araszkiewicz et al. (2022), elsewhere in this issue, for a discussion of Walton's papers in *AI and Law* journal and Atkinson et al. (2020a) for his overall influence on the field.

documents. Mochales and Moens (2011), commented on by Adam Wyner, aimed to automatically detect, classify and structure argumentation in text, using a general purpose corpus produced at the University of Dundee, known as the Araucaria corpus, and a specifically legal corpus, a set of documents extracted from legal texts of the European Court of Human Rights (ECHR). Both of these papers were the precursors of significant research activity, as is made clear in the commentaries.

2 The role of context in case-based legal reasoning: Teleological, temporal, and procedural (Hafner and Berman 2002). *Commentary by Giovanni Sartor*

Hafner and Berman (2002) consolidates three papers by Don Berman and Carole Hafner, which they presented at three consecutive ICAIL conferences (1991, 1993, 1995). All three papers addressed the crucial concerns of the article, namely expanding the scope for case-based reasoning, on the basis of a representation of different contextual aspects. The first of these papers (Berman and Hafner 1991), deals with procedure, the second with purpose (Berman and Hafner 1993), and the third with time (Berman and Hafner 1995). In the consolidation paper these issues are presented in an order that differs from the sequence of the papers: purpose, time and procedure.

The analysis of these three contextual aspects addresses, from different perspectives, what Hafner and Berman saw as a fundamental deficiency in contemporary work on case based reasoning. Such approaches focus on features that are meant to capture, at various levels of abstraction, the facts of a legal case. On this basis such approaches aim at proposing reasoning moves that concern new cases, and ultimately, at predicting the outcome of such cases. This is exemplified by the HYPO system developed by Rissland and Ashley, which represented each case through a set of pro and con factors (Ashley 1990) and by Branting who represented cases using semantic networks (Branting 1991) and by linking outcomes to factual descriptions at different levels of abstraction (Branting 1993).

The first extension that Berman and Hafner propose consists in adding *purposes* to a fact-based representation of cases. More specifically, they propose to add to each factor an indication of "the legal purpose(s) which it affects, and each legal purpose in turn specifies whether it favours the plaintiff or defendant". They also propose to represent purposes through a knowledge graph that indicates relationships of subsumption and conflict between different purposes. They exemplify their approach by considering a set of cases dealing with situations in which hunters sue third parties whose intervention deprived such hunters of their catch. These cases are studied by many American law students in their introductory Property Law courses and their representation has, since Berman and Hafner (1993), become a paradigmatic test for computational models of legal reasoning (e.g. Bench-Capon and Sartor (2003); Atkinson (2012); Al-Abdulkarim et al. (2016)). For instance, the factor consisting in the fact that the prey had not been caught when the intervention took place is a factor that favours the defendant (the third party) and affects the value of legal certainty. In other words, by recognising this factor as favouring the defendant, i.e., by having decision in favour of the defendant when this factor is present, the law promotes legal certainty (if hunters who had not caught an animal could claim possession against third parties, unnecessary disputes would arise). Berman and Hafner argue that, by pointing to purposes, the parties may offer reasons to choose between competing arguments based on factual analogies, as determined by the factors in the cases. These will include arguments for preferring one purpose over another as modelled in Bench-Capon and Modgil (2009).

The second suggested extension is to model evolving legal doctrines. Their focus is on the process through which an argument can be weakened: "legal precedents are embedded in a temporal context of evolving legal doctrine, which can result in a strong precedent becoming weaker over time, to the point where a skillful attorney could reasonably **predict** that it will no longer be followed." (p31). To exemplify this dynamic, Berman and Hafner consider cases in which guest passengers sue host drivers for negligence, asking for damages, an issue that is differently regulated in different jurisdictions: New York allows for such claims while other jurisdictions, such as Ontario, restrict or exclude them. The traditional criterion for determining the applicable law was the territorial principle, requiring the application of the law of the place in which the accident (the event generating the legal claim) occurred. However, in other legal domains such as contracts, this rule has been abandoned, in favour of the center-of-gravity criterion, which focuses on the jurisdiction more closely connected to the interests at stake in the case. The latter criterion would indicate New York's law when both the driver and the passenger are from New York, even where the accident took place elsewhere. In predicting the outcome of a new case, involving New York residents having an accident in Ontario, precedents sharing features with the new case should only be considered with a caveat: the significance of the precedent which shares most factors with the new case may be diminished by more recent precedents, even when such precedents seem to address different matters. For instance, a precedent which applied to a contract the law of the place of performance (the center of gravity) rather than the law where the contract was executed (the place of the event), may weaken an earlier precedent which denied compensation to the New York passenger injured in Ontario by applying the territorial principle. The reason for the weakening is that the more recent precedent, while adjudicating a different claim, addresses a preliminary issue (the applicable law) which it shares with both the new case and the earlier precedent by applying a criterion (center of gravity) which diverges from the criterion adopted in the older precedent (the territorial principle).

To capture the weakening of precedents, a computational model is proposed where the representation of a case includes the following aspects: (a) both the final and the preliminary issues which were addressed by the judges; (b) the criteria according to which each issue was decided; (c) the values that are promoted by deciding an issues according to a certain criterion (e.g. according to the territorial principle rather than the center of gravity). Berman and Hafner also observe that the criterion adopted in a precedent to decide an issue can be challenged in different ways in newer precedents: the old criterion may be explicitly overruled, it may be implicitly overruled, it may be overruled in dissimilar cases, it may be subject to exceptions, or the values supporting that criterion may be subordinated to the values underlying alternative criteria. An algorithm is proposed to determine the extent to which the challenges ("red flags") just described may cumulatively contribute to the weakening of a precedent.

The third suggested extension concerns modelling the procedural context in which a precedent decision was generated. They distinguish the pleading state (where it is determined whether the alleged facts support, according to the law, a party's request), the pre-verdict stage (where the court can rule that as a matter of law the evidence offered is insufficient to create an issue for a trier of fact to decide), and the trial verdict stage (where the case is decided, on the basis of the facts, and possibly of the law). A further distinction is made between decisions on procedural matters and decisions on matters of fact: a decision in favour of the defendant party based on procedural matter (e.g., lack of evidence), may not support the same decision in a new case which shares the factual features of the precedent (e.g., the defendant won since there was not sufficient evidence for the facts alleged, while such evidence is available in the new case). On the basis of these distinctions, the support of a precedent decision is linked to the procedural settings of both that decision and the decision to be taken in the new case. For instance, it is argued that a decision at the pleading stage for one party strongly supports a corresponding new decision for that party at the pleading stage, and also at the trial stage. A trial verdict for a party only weakly supports a decision for that party at the pleading stage (as the decision at the trial phase may depend on factual issues, e.g., although the plaintiff lost, he would have won had he been able to prove the facts he alleged).

All in all, Hafner and Berman (2002) has a very rich content, and indeed the work that it reports, especially the work on purposes, has inspired much subsequent work on AI and Law.

The first suggestion in the paper namely, the idea to include purposes in the representation of legal cases, has been highly influential in subsequent work, although this later work tended to use the term "values" instead of purposes. Interest in values developed following the publication of three papers inspired by Berman and Hafner (1993): Bench-Capon (2002), Prakken (2002) and Sartor (2002), all of which appeared in the same issue as Hafner and Berman (2002). Here I will only mention some of the key AI and Law papers that pursued this direction, since there is an excellent summary in the comment of Trevor Bench-Capon to Berman and Hafner (1993) in Bench-Capon et al. (2012). These key papers included Bench-Capon and Sartor (2003), Greenwood et al. (2003), Chorley and Bench-Capon (2005), Sartor (2010) and Grabmair and Ashley (2011). Values have continued to be a topic of active research, including Al-Abdulkarim et al. (2015), Verheij (2016), Muthuri et al. (2017), Grabmair (2017) and Maranhão et al. (2021).

The second suggestion, namely, modelling the temporal dimension of cases, has led to two possible lines of development. The idea to include in the model of a case also the decisions on preliminary issues was earlier addressed by Branting (1993), and subsequently was retained in argumentation-based models of case based reasoning, such as Prakken and Sartor (1998), which is presented elsewhere in this issue.[4]

[4] See Sect. 9 of Governatori et al. (2022).

That issues should be modelled separately was also recently advocated in Bench-Capon and Atkinson (2021). The evolution of case law was explicitly addressed by Rissland and Xu (2011), Henderson and Bench-Capon (2019), Verheij (2016) and Zheng et al. (2021a).

The third suggestion, namely, modelling procedural contexts, as far as I know, has not much been taken up in analyses of case-based reasoning within AI and Law. This may be because Berman and Hafner's analysis is based on technicalities of the US legal procedure, and would need to be redefined within other legal contexts. Although there was considerable interest in the procedural aspects of law in the 1990s (e.g. Hage et al. (1993) and Gordon (1993)[5]), these tended to model the legal procedure as the protocol of a dialogue game. In particular Gordon (1993) modelled the the pleadings phase in this way. These accounts did not, however, consider the effects of context as it impacted on cases when used as precedents. Context has been explicitly modelled in Wyner and Bench-Capon (2009), which considers a variety of contexts including appealing a case, overruling a precedent and rehearing of a case as a civil rather than criminal proceeding, and Verheij (2016), which uses the cases from Berman and Hafner (1995) to illustrate its approach to the context dependence of ethical decision making.

3 Dialectical argumentation with argumentation schemes: an approach to legal logic (Verheij 2003c). *Commentary by Floris Bex*

Argumentation schemes are general patterns of reasoning that underlie arguments, with associated critical questions that point to sources of doubt for such arguments. They have become central to much of the research in AI and Law, having been used in modelling reasoning with evidence (Bex et al. 2003; Bex 2011), reasoning with cases (Prakken et al. 2015), e-democracy (Atkinson et al. 2006), Bayesian legal reasoning (Hahn et al. 2013), legal ontologies (Gordon 2008) and recently as a tool for statutory interpretation (Araszkiewicz 2021) and explainable AI in the law (Atkinson et al. 2020b). However, up until the early noughties, the use of argumentation schemes of the kind that Walton (1996) popularised was not widespread in AI and Law, with researchers interested in informal argumentation mostly using the general argument scheme of Toulmin (1958) (e.g. Marshall (1989) and Bench-Capon et al. (1993)). This changed when in 2003 an issue of AI and Law journal appeared with two papers on Walton-type argumentation schemes in a legal context. In this section, one of these papers will be discussed, namely Bart Verheij's *Dialectical argumentation with argumentation schemes: An approach to legal logic* (Verheij 2003c).[6]

In his paper, Verheij discusses argumentation schemes as the basis for a logic of law. The logic of law had at that time already received quite some attention, mainly

[5] See Sects. 6 and 7 of Governatori et al. (2022), elsewhere in this issue.

[6] The other article on argumentation schemes that appeared in the same issue as Verheij (2003c) is Bex et al. (2003), which is discussed elsewhere in this issue (Villata et al. (2022), Sect. 3).

from the argumentation in AI and Law community (for example, Bench-Capon et al. (1993), Gordon (1993), Loui and Norman (1995), Prakken and Sartor (1996), Hage (1997)[7]). However, these approaches mainly focus on more generic (defeasible) logical inference which, while relevant for a logic of law, does not consider the specific domain rules or contextual reasoning patterns in law. For this, Verheij argues, it is necessary to look at argumentation schemes as proposed in Walton (1996) and their associated critical questions. More specifically, Verheij analyses these schemes drawn from informal logic using formal methods using a concrete dialectical logic approach: *concrete* in that schemes can represent domain- or context-specific arguments, *dialectical* in that schemes are subject to counterarguments, and *logical* in that formal methods are used to analyse and represent the schemes.

The paper provides ample examples of argumentation schemes. One legal example is as follows.

AS Person *P* has committed crime *C*. Crime *C* is punishable by *n* years of imprisonment. *Therefore* Person *P* can be punished with up to *n* years in prison.

Typical critical questions for this scheme would be something like *Was it really person P who committed crime C?* and *Is there a justification for person P committing crime C?*. Other examples of argumentation schemes range from logical rules such as *modus ponens* to more specific pragmatic argumentation schemes from Walton (1996), such as the *Ad Hominem* and *Expert Opinion* schemes. And, while there are some argumentation schemes that look like small derivations (*A Therefore B Therefore C*) or pieces of dialogue (Proponent: *"A"*; Opponent: *"B Therefore not A"*), it can in general be argued that argumentation schemes have the form of (defeasible) inference rules, viz.

$Premise_1, \dots, Premise_n$ Therefore *Conclusion*.[8]

However, as Verheij argues, Walton's argumentation schemes have a certain "looseness" about them, which makes them ideal for analysing real-life arguments but less useful as a legal logic. Hence, he introduces a four-step method for investigating and formalising argumentation schemes.

The four-step method for argumentation scheme formalisation is as follows:

1. Determine the relevant types of sentences;
2. Determine the argumentation schemes;
3. Determine the exceptions blocking the use of the argumentation schemes;

[7] Several of these papers are discussed in Governatori et al. (2022), elsewhere in this issue.

[8] Although many authors, including Walton, have claimed something like this (cf. e.g. Bex et al. (2003)), Walton himself never considered argumentation schemes as purely (domain-specific) rules, but rather as dialogical or dialectical devices, where the critical questions are a key component of the scheme. Cf. Atkinson and colleagues' recent paper on the influence of Walton on AI and Law (Atkinson et al. 2020a). See also the discussion of Walton's papers in Sect. 4 of Araszkiewicz et al. (2022), elsewhere in this issue.

4. Determine the conditions for the use of the argumentation schemes.

Verheij first explains this method informally with examples, and then proceeds to formalise some argumentation schemes in his formal theory of dialectical argumentation, DEFLOG (Verheij 2003b). The language of DEFLOG consists of elementary logical sentences plus an implication symbol $\rightsquigarrow$ denoting that one statement (sentence) implies another statement, and a dialectical negation symbol $\times$ denoting that a sentence is defeated. So, for example, $(p \rightsquigarrow q)$ means that p implies q, $(p \rightsquigarrow \times q)$ means that p implies the defeat of q, $(r \rightsquigarrow (p \rightsquigarrow q))$ means that r implies that p implies q, and $(r \rightsquigarrow \times(p \rightsquigarrow q))$ means that r implies the defeat of p implies q. Verheij further provides argument diagrams based on DEFLOG, in which normal arrows stand for DEFLOG's implication and $\times$-headed arrows stand for the implication of dialectical negation. The above examples are then as in Fig. 1.[9]

In step 1, the logical language is defined. That is, the sentence types are determined, where sentence types contain variables and an instantiation of a sentence type gives rise to a sentence. Sentence types are the building blocks of argumentation schemes. For example, in the case of scheme *AS* we have at least the following sentence types *Person P has committed crime C*, *Crime C is punishable by n years of imprisonment* and *Person P can be punished with up to n years in prison*. Often, we will also take the negation of these as sentence types, i.e., *Person P has* not *committed crime C*. Sentences are then instantiations of sentence types, e.g., *The crime of murder is punishable by 20 years of imprisonment*.

In step 2 the argumentation schemes are defined given the sentence types of step 1, that is, which sentences can be inferred from which other sentences. In logic, the argumentation schemes would be the inference rules, and they can be represented in DEFLOG as $premise_1, \ldots, premise_n \rightsquigarrow conclusion$. The example scheme *AS* given above can be rendered as the ArguMed diagram in Fig. 2.

In step 3, the focus is on a common type of counterargument based on a critical question, namely one that provides an exception to the general rule the scheme captures. For example, whilst normally people are punishable when having committed a crime, if a person has a ground of justification for committing the crime, such as *force majeure*, they are not deemed punishable. This exception attacks (the application of) the general rule from premises to conclusion (Fig. 3), and thus such an attack is what Pollock first called an *undercutter* (Pollock 1987).[10]

Finally, step 4 looks at conditions for the use of an argumentation scheme. Because argumentation schemes are usually not universally valid, there are often conditions that must be fulfilled if they are to be applicable, and these may need to be made explicit. In our example scheme a condition might be that person *P* has been brought before a qualified judge (Fig. 4).

[9] This figure and the others in this section were made using Verheij's argumentation software ArguMed based on DEFLOG, which is still available from his website https://www.ai.rug.nl/~verheij/aaa/argumed3.htm (last accessed 12-2-2022). See also (Verheij 2003a).

[10] In formal argumentation, this notion of undercutting is now fairly standard, cf. (Prakken 2010).

After discussing the four-step method, Verheij discusses the role of critical questions, which point to counterarguments to the original argument based on the scheme. Two types of critical questions correspond to steps in the four-step method, namely questions that point to exceptional situations in which a scheme should not be used (step 3) and questions that point to conditions for a scheme's use (step 4). For example, the question *Is there a justification for person P committing crime C?* that was mentioned for scheme *AS* corresponds to the counterargument presented in Fig. 3, and a positive answer to the critical question *Has person P been brought before a qualified judge?* would give us the argumentation scheme presented in Fig. 4. Verheij does not discuss exactly what the difference is between conditions and exceptions, but we can consider this difference in terms of the dialectical process with a proponent, who in this case presents the argument based on *AS*, and an opponent, who criticizes this argument. Questioning a condition puts the burden of proof on the proponent to show that it is true, while questioning whether there is an exception puts the burden of proof on the opponent to show that this is indeed the case (Prakken et al. 2005). It is also possible to include the difference between exceptions and conditions explicitly in the logic itself: Gordon et al. (2007) propose different types of premises for argumentation schemes, namely exception premises (exceptions which are assumed to be false) and assumption premises (conditions which are assumed to be true).

Verheij further mentions critical questions that criticize a scheme's premises, and questions that point to (counter)arguments contradicting a scheme's conclusion. For example, the earlier question *Was it really person P who committed crime C?*, if answered negatively, would give us a counterargument *Person P has not committed crime C* that attacks the premise of the examples in Figs. 2, 3 and 4).[11]

Verheij finishes his paper by showing how arguments based on argumentation schemes thus constructed can be analysed in a concrete dialectical logic such as DEFLOG or Reason-Based logic (Hage 1997). With his structured and formal account of argumentation schemes as the basis of legal logic, Verheij popularised argumentation schemes in the AI and Law community. Interestingly, Bex et al. (2003), which is also about capturing and formally analysing and representing argumentation schemes, appeared in the same issue of *AI and Law*. This issue thus kickstarted the more widespread use of argumentation schemes in AI and Law (see the beginning of this section and Atkinson and Bench-Capon (2021)), as well as further research into logical interpretations of argumentation schemes (Gordon and Walton 2009). Furthermore, Verheij and Bex would later combine Verheijs approach to legal reasoning with argumentation schemes and Bex's approach to evidential reasoning with argumentation schemes in several articles that captures both legal and evidential reasoning (Bex and Verheij (2012), Bex and Verheij (2013)[12].

[11] In formal argumentation, such an attack on a premise is sometimes called *undermining*, and an argument that attacks the conclusion is called a *rebutter* (Prakken 2010).

[12] For a discussion, see Sect. 3 of Araszkiewicz et al. (2022), elsewhere in this volume.

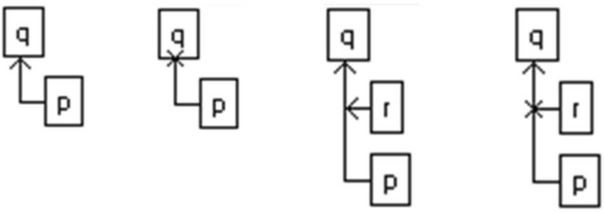

Fig. 1 ArguMed diagrams of the DEFLOG formulas (from left to right): $(p \rightsquigarrow q)$, $(p \rightsquigarrow \times q)$, $(r \rightsquigarrow (p \rightsquigarrow q))$, $(r \rightsquigarrow \times(p \rightsquigarrow q))$

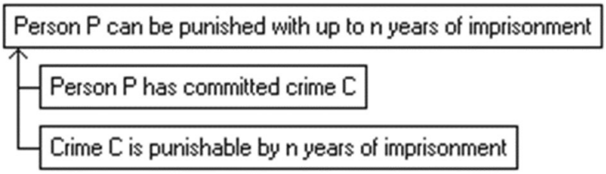

Fig. 2 ArguMed diagram of a legal argumention scheme

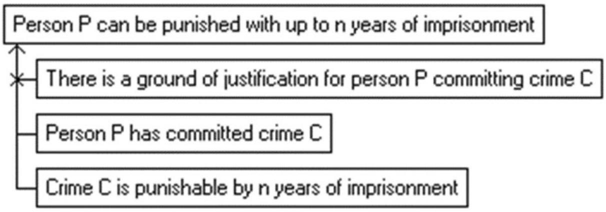

Fig. 3 A legal argumentation scheme with an exception

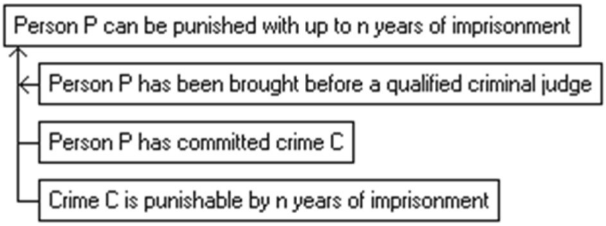

Fig. 4 A condition for the use of a legal argumentation scheme

4 Legal ontologies in knowledge engineering and information management (Breuker et al. 2004). *Commentary by Tom van Engers and Giovanni Sileno*

We selected the paper on legal ontologies by Breuker, Valente and Winkels (Breuker et al. 2004)[13], for its historical and methodological relevance, and for the vision of our discipline that it offers.

The paper presents itself as aiming to collect and synthesize experience of at least a decade of research within the Leibniz Center for Law, the former AI and Law department at the University of Amsterdam[14]. Like other departments in the UK, Germany, France and Italy, the Leibniz Center came into existence with the increase in public attention towards the potential use of technology applied to law (at the end of eighties/beginning of nineties)[15]. In particular, the authors elaborate on the three areas of research: FOLaw, a functional ontology of law (Valente 1995; Valente et al. 1999); LRI-core, a core ontology for law (Breuker et al. 2004a); and (more briefly) NLP/information retrieval methods in legal documents leveraging these ontologies (e.g. the CLIME project (Winkels et al. 2002)).

From a temporal point of view, and for the subjects it touches upon, the paper can be placed in the long wave of expert systems (eighties), of multi-agent systems (nineties), of statistical NLP and IR (nineties), and of semantic web technologies (noughties). This framing makes it particularly distant from the recent wave of deep learning and related technologies occurring in the last decade, and indeed its content may plausibly not be immediately recognized as important for most AI practitioners today. However, the history of AI has plenty of cases in which previous points of view would offer, in some revisited form, pointers to how current undefeated challenges might be addressed.

For its summarizing/reflective purpose, the paper has a historicist attitude in itself. It offers (Sect. 2) a short but reasoned overview of the different traditions on ontology (in philosophy, knowledge engineering, and software engineering), of the often mixed levels of abstractions to which ontologies are generally associated, and of the specific roles of ontologies in AI and Law (to organize and structure information, for reasoning and problem-solving, for semantic integration/interoperation, and for understanding a domain). The paper then provides a succinct but functional summary of theories of law (Sect. 3), concerning validity and coherence, the main ontological commitments of law expressed in the theories of Kelsen, Hart, Bentham and Hohfeld, and how they are received in AI and Law.

[13] The paper appeared as part of a special issue on *Ontologies for Law* (Breuker et al. 2004), which indicates the great interest in the topic at the time. Elsewhere in this issue, Sect. 2 of Araszkiewicz et al. (2022) discusses the development of ontolgies in AI and Law over the decades.

[14] This department, part of the Law faculty, has hosted for two decades a handful of researchers coming from Psychology, Legal studies, Artificial Intelligence and Computer Science. In 2017 its flag passed to the Leibniz Institute, spanning over the faculties of Law and Science of UvA, and TNO, the Dutch organization for applied research.

[15] In terms of interest, today we live a similar heyday (Francesconi 2022), albeit very different approaches are being used: what is understood today by a general audience by the term *Artificial Intelligence* is most probably some machine-learning-based, data-driven approaches, whereas RegTech and similar technologies are much more related to distributed systems than normative systems research.

4.1 Methodological content

The central focus of the paper lies, however, in Sects. 4–6. The FOLaw ontology (Sect. 4) inherits the most characteristic aspect of expert systems research: the focus on human problem-solving. The authors themselves recognize (although only retrospectively) that FOLaw has a strong connection with CommonKADS, a set of design methods for building knowledge-based systems (Schreiber et al. 1994). FOLaw identifies a number of knowledge types relevant for legal activities (normative knowledge, world knowledge, responsibility knowledge, reactive knowledge, creative knowledge and meta-legal knowledge) and their inter-dependencies, related to their role in legal reasoning. In various attempts to reuse FOLaw, it turned out that most effort in modeling went on world knowledge, which suggested moving the focus from epistemological frameworks to ontological frameworks.

With hindsight, the development of the consequent LRI-core ontology (Breuker et al. 2004a) can be placed in the golden age of ontologies. Compared to other similar efforts, it has a unique characteristic: it considers the specificity of legal reasoning, and, in doing so, reaffirms the importance of common-sense reasoning for legal understanding, rather than logical-mathematical, physical, or more generally scientific models[16].

The main categories that LRI-core describes are physical and mental concepts, roles, abstract concepts and terms for occurrences. We will cite here two particularly interesting observations made by the authors: how law concerns essentially "roles" and not agents, and that objects (as concepts) need to be separated from their occurrences, similarly to the semantic memory/episodic memory distinction observed in psychology. In Sect. 6, the authors elaborate on an application of LRI-core; first as anchoring an ontology for Dutch criminal law, CRIME.NL, developed as part of e-COURT project (Breuker et al. 2002), and then used to support the retrieval of information contained in the hearing session documents of criminal cases. For example, a query using only keywords given by a user usually returns unsatisfactory results; however, as ontologies provide e.g. subsumption relationships, queries can be expanded e.g. to subsuming and subsumed classes.

4.2 Vision

The follow-up to a certain research trajectory may not necessarily occur immediately after it is identified. FoLAW did not have any descendants. LRI-core furnished the basis for LKIF-core ontology (Hoekstra et al. 2007), which arguably remains one of the most complete examples of a core ontology designed purposely for law[17]. But LKIF-core arrived a couple of years before the rise of the deep-learning wave,

[16] For instance, to accept that humans do not typically reflect on their conduct before taking decisions (e.g. the neurological evidence in Daniel (2002)) would map our view of the world to some form of emotional determinism, which would undermine many of the (fictional, possibly illusory) constructs that allow our societies to be maintained.

[17] A more recent proposal in this direction is UFO-L (Griffo et al. 2016).

which washed away most of the attention paid to knowledge-driven approaches.[18] The introduction of pre-trained models (e.g. Devlin et al. (2019)) providing word/sentence embeddings, and of transformers (Wolf et al. 2020) have today absorbed much of the space of traditional NLP/IR techniques.

Breuker et al. (2004) was published well before any of these aspects were manifest. Looking at contemporary research in AI, including AI and Law, the efforts of Breuker et al. (2004) sound extraordinarily distant, but their vision is still worthy of consideration. First, there is a genuine attempt to reconstruct an architecture of cognitive functions operating for performing legal activity out of human experiences, but not conflating this inquiry to empirical induction. Second, there is a constitutive attempt to define what is "correct" as a foundation for proper legal reasoning. *Constitutive* because it is not defined with respect to a definite outcome, but discusses the principles (and constraints) this outcome should be based on, and why those principles are right. Constructing a legal core ontology as a latent space of some empirical observations (of text? of behaviour?) would fail to satisfy the requirements that such an ontology must have for humans: providing and maintaining means to make sense of the world (including responsibility, will, etc.). Third, it reminds us how information processing can and should to some extent be guided by humans.

4.3 Reappraisal of knowledge-driven AI in the realm of data-driven AI

The currently mainstream data-driven AI approaches have shown to be successful in several areas, but they raise critical concerns when applied in domains where humans are being subjected to computational 'decisions.'[19] The call for explainability, fairness and transparency has grown over the last years, leading to new legislation putting constraints on the use of AI and impacting its construction.[20] If we look at demands for explainability it is easy to see that, for any application in the legal domain, that explanation should be built on legal arguments and a clear argument construction and epistemological derivation mechanisms. Popular applications such as risk assessment in banking, taxes, social benefits etc., may use some data-driven components, comparing the data of clients to other clients to classify cases into risk categories, but, even if this categorization may be argued for in statistical terms, the consequent treatment of the clients should always be backed-up by legal

[18] Furthermore, a general disillusionment emerged, even more in practical settings, towards semantic web technologies, for their inability to handle (normative) reasoning in a scalable way.

[19] A number of criticisms have been put forward: e.g. Bench-Capon (2020) cites, as well as lack of explanations, the bias and mistakes present in past cases, the fact that the law may have evolved so that past decisions may have been made with different understandings of the law at different times, and the fact that the law is subject to change in the future. Medvedeva et al. (2020) shows that performance degrades as the dataset ages. Bex and Prakken (2021) demonstrate that it is not rational to follow predictions blindly, even given a high level of accuracy, and Steging et al. (2021) show that high accuracy can be achieved even when the underlying rationale is flawed. Many of these problems can be mitigated if the predictions are *explainable*, by giving a justification in legal terms, but this requires in principle some form of knowledge model.

[20] e.g. in Europe, the Artificial Intelligence Act, https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD) &l=en.

arguments (e.g. (Bex and Prakken 2021)). This challenge makes the work of Breuker et al. (2004), and of other researchers that work on (legal) knowledge-driven AI, still relevant today.

5 Fundamental legal concepts: a formal and teleological characterisation (Sartor 2006). *Commentary by Enrico Francesconi and Michał Araszkiewicz*

Sartor (2006) presents a set of definitions of normative positions relevant in the field of law (deontic obligations, different types of rights) as well as some other fundamental legal notions, such as normative conditionals and legal powers. The method he adopts may be referred to as *semi-formal modeling*: the paper's aim is not to develop a fully-fledged formal account of fundamental legal concepts, but rather to clarify some aspects of their logical structure and the role they play in legal reasoning. In this sense, this contribution may be classified as foundational - it provides a theoretical grounding for formal models of (different domains of) legal knowledge and reasoning. The topics presented in the paper were previously discussed in several chapters of Sartor's comprehensive monograph (Sartor 2005).

Concepts referring to normative positions, such as "duty" or "right" have been the subject of inquiry in legal philosophy since the early days of the discipline. The modern approach towards this problem has been shaped, to a large extent, by the influential work by Hohfeld (1913) who introduced the basic logical relations between these concepts, captured in the well-known *Hohfeldian squares*. The progress of logic in the 20th century enabled a more nuanced analysis of the formal featres of Hohfeldian concepts, including the influential work of Kanger (1972) and Lindahl (1977); this work has become a standard reference and is often labeled the *Kanger–Lindahl theory* (Sergot 2013). In parallel, the problem of fundamental legal concepts has been the subject of legal-philosophical investigations including Kelsen (1967), Ross (1968) and Spaak (1994). Normative concepts have also become a subject of interest of AI and Law research (McCarty 1986). It has been argued that robust legal knowledge and reasoning modeling requires a theory and a language extending beyond standard deontic logic to include Hohfeldian concepts (Allen and Saxon (1986) and Allen and Saxon (1993)). Of course, this claim is related to such elements of legal knowledge (and associated reasoning) which makes substantial use of the notions of rights, duties, powers and similar concepts. These are not needed for many expert systems type applications (Bench-Capon (1989), Jones and Sergot (1992)) which simply apply the law without questions of violation arising. The development of a complete and coherent model of normative positions, however, enabling legal reasoning across different domains, does require such fundamental concepts to be addressed. The Kanger–Lindahl theory has been formalized and adopted for computational purposes (Sergot (2001) and Sergot (2013)). The theory of normative positions has also been applied to domains other than law (Jones and Parent 2008), but it remains essential for legal knowledge modeling when this concerns rights. duties and violations. Sartor (2006) is methodologically situated between strict formal and computational approaches on the one hand, and informal legal-theoretical investigations on the other hand.

The conceptual scheme developed by Sartor is based on the notion of action. Actions may be obligatory, prohibited or permitted. However, in order to account for more complex normative positions, we need to introduce further types of knowledge, going beyond classical deontic modalities. The paper introduces an additional, teleological, layer to the analysis: indication of a typical interest (goal, value) realized if a given normative proposition is adopted. This interest may be assigned to an individual or a group of individuals, including society as a whole. These considerations enable Sartor to define *directed deontic modalities*, in particular directed obligations, where the performance of an obligatory action advances the interest of an individual or group of individuals. For instance, if Tom has an obligation towards Anne to pay her 100 Euro then it is obligatory for Tom to pay Anne 100 Euro, for the advancement of Anne's interest. Directed prohibitions and permissions are also considered. The introduction of a teleological characterization of deontic modalities enriches the legal knowledge modeling by introducing a slot representing agents towards whom an action subject to a deontic modality is directed. The notion of a directed obligation is further used to define an obligative right, which is directed obligation's correlate: if Tom has an obligation towards Anne to pay her 100 Euro, it means that she has an obligative right that Tom pays her 100 Euro. This approach enables Sartor to define the remaining concepts of the first Hohfeldian square – noright and privilege – as well as permissive rights, *erga omnes* rights and exclusionary rights.

The paper follows with a classification of normative conditionals. The conditional relation between an antecedent and consequent is not subject to any particular formalization; it is only assumed that this relation enables the consequent to be derived (at least defeasibly) if the antecedent holds. The focus is on the characteristics of the antecedents and consequents of normative conditionals. As a result, nine types of normative conditionals are distinguished. This systematization is followed by an account of the well-known category of constitutive rules and the counts-as relation. Further, the paper discusses a specific meaning of the word "must" in legal texts, which does not indicate an obligation, but rather a relative necessity: certain conditions that must be met in order for some result to obtain. This use of the word "must" is referred to by Sartor as the use in *anancastic* sense.

The next part of the paper is devoted to the complex notion of power. The classification of normative conditionals enables Sartor to distinguish and define five different sub-types of power. On this basis, the paper can account for the concepts represented in the second (potestative) Hohfeldian square and provide definitions and comments on the notions of enabling powers and potestative rights. The notions of obligative rights and potestative rights are related to each other: a violation of an obligation (a correlate of an obligative right) typically activates a certain agent's potestative right to impose a sanction. The catalogue of defined concepts is further enriched by liability rights. This contribution of the paper goes beyond the scope of normative positions captured by standard Kanger–Lindahl theory, as the latter does not address the concept of power or competence (Sergot 2013).

Finally, the paper discusses the notion of *result-declaration*. This notion is a modification of an idea taken from the speech act theory (Searle 1969) where a certain result is brought about by (occurs) in social reality through an utterance. Sartor's notion of result-declarations encompasses acts performed by the addressees of legal norms (for instance, a party terminating a contract), legal officials applying the law (court issuing a judgment) and law-making institutions (parliament enacting a statutory rule). The concept of result-declaration is discussed in connection with the concept of validity which is seen here as a meta-linguistic predicate. Eventually, the notion of declarative power is formulated: an agent has a declarative power to realize A if, if this agent declares A, then A is legally valid. The notion of source of law is defined as any such fact which embeds normative statements and makes them legally valid in virtue of such embedding.

This paper is a good example of the use of modest formal tools for the sake of disambiguation and explication of legal-theoretical concepts which are subject to a heated debate. Importantly, the author does not ground the proposal on any pre-existing robust legal-theoretical conception, but provides original solutions. Some of these results have been applied in the modeling of regulatory compliance for software requirements (Ingolfo et al. 2013) or referred to in the works in the field of legal ontologies (Agnoloni et al. 2009). The paper demonstrates clearly the richness and diversity of legally relevant normative positions and substantiates the claim that it is necessary to go beyond the basic deontic modalities if we aim at developing legal knowledge bases and reasoning systems capturing the use of rights, duties and cognate concepts. Importantly, the fundamental character of the explicated concepts means that they are to a large extent jurisdiction- and domain neutral. Therefore, Sartor (2006) may be considered a substantial contribution to the debate directed towards the standardization of legal knowledge representation. It was used in Francesconi (2014), and referenced in Sileno et al. (2015) and Pascucci and Sileno (2021). Some of its results have also been discussed in recent legal-theoretical literature (de Oliveira Lima et al. 2021). Sartor's paper shows the importance of development of a foundational ontology for the purposes of law, with the definitions of the discussed concepts making use of more abstract, undefined concepts such as the notions of action, state of affairs or events.

6 Extractive summarisation of legal texts (Hachey and Grover 2006). *Commentary by Frank Schilder*

Lawyers often rely on summaries of long legal texts and also produce condensed descriptions of complex legal issues for themselves. The AI and Law community has been exploring systems for legal summarization for a while now and one of the first approaches addressing legal summarization is the seminal work Hachey and Grover (2006). Subsequent work took their techniques and applied them to text coming from different legal frameworks (e.g., Saravanan and Ravindran (2010) and Yamada et al. (2019)) and for identifying text segments in legal information retrieval tasks (Tran et al. 2020).

Hachey and Grover's approach and subsequent work were *extractive* by nature, and identified important sentences or clauses in the original text. More recently, neural network approaches have become more dominant and address the question of *abstractive* summarization (de Vargas Feijo and Moreira 2021). Although the field has seen some progress in this respect, it also now faces issues from generated text that was not in the input text and may contain potentially contradictory or harmful language (a.k.a. *hallucinations*).

In 2006, Hachey and Grover proposed a novel approach to summarizing legal documents that utilizes the work by Teufel and Moens (1997, 2002) on argumentative zones. Text summarization is seen as domain dependent and by identifying important zones of a text, the authors show that a more effective summary is generated by taking those zones into account. Automatic summarization approaches to other domains such as news may not require such treatment because news text is often created according to the inverted pyramid structure (i.e., the most important information is stated at the beginning of the article) making a baseline system that takes the first *n* sentences of a text a highly competitive summarization system.

Hachey and Grover's work addresses the task of legal summarization as a machine learning problem of identifying different zones that are relevant for writing a legal text (e.g., Framing, Fact, Disposal). They show superior results to summarization utilizing the set of machine learning algorithms (SVMs, Decision trees etc.) predominantly used at the time. Since their overall approach is extractive, they identify sentences in the judgment as representative for the respective zone. They also provide an annotated corpus of judgments of the House of Lords (HOLJ) for training and testing their approach.[21]

The impact of their work can be clearly seen in subsequent work that identified other ML approaches for this task and also expanded the way legal text can be annotated and legal concepts can be captured.

– Saravanan and Ravindran (2010) followed in their footsteps and provided annotation guidelines for analyzing three sub-domains of Indian caselaw. They showed in particular that Conditional Random Fields (CRFs, Lafferty et al. (2001)) can capture the sequential nature of a legal case.
– Work by Yamada et al. (2019) showed another application of rhetorical structure to legal text. The authors developed a novel annotation scheme for Japanese judgments that is also influenced by the annotation schema of Hachey and Grover (2006). They started with rhetorical classification and combined these classes with relation-based argument mining tasks leading to a new Issue Topic concept.

Yamada et al.'s annotation work also gives way to testing various automatic means to identify the rhetorical structure and topics. Similar to previous work up to that

[21] The HOLJ corpus comprises 188 judgments from the years 2001–2003 from the House of Lords website. The authors extracted the judgements, removed the HTML tags, and assigned two types of label to each sentence: the rhetorical role and a relevance metric.

point, classical ML methods such as SVMs or CRFs using features including n-grams and various other syntactic and semantic features showed decent performance when classifying rhetorical relations and identifying topics.

The last ten years have seen the reawakening of neural network methods. Approaches to legal summarization have been adapting those techniques as well. For example, Tran et al. (2020) utilize a neural summarization approach in order to improve a legal retrieval system. They use various word or document embeddings (i.e., word2vec, GloVe and doc2vec) that feed into convolutional layers of their neural network architecture.

In contrast to most previous work, more recent work by de Vargas Feijo and Moreira (2021) is addressing the more challenging task of *abstractive* summarization. Traditionally, *extractive* summarization had been the main focus of previous approaches as in extracting sentences or phrases verbatim from the original text ([Nallapati et al. (2017) and Cheng and Lapata (2016)).

de Vargas Feijo and Moreira (2021) use the current state-of-the-art approach for many NLP tasks based on a transformer architecture. However, in order to process documents that are longer than the maximum tokens that BERT (Devlin et al. 2019) can process (i.e., 512), they explore various strategies to split a ruling into different chunks of coherent text. Those chunks are run through a transformer-based architecture generating a summary. Chunk and summary pairs are then used as input for BERT in order to score the summaries. The summary with the highest score is selected as the final summary of the entire case.

The proposed system by de Vargas Feijo and Moreira (2021) shows that summarizing legal text is clearly more complex than summarizing news messages because of the length of the text and the current limitations of language models. Even language models that allow for longer input text need to solve the problem of identifying different parts of the legal documents and their functions. It is unlikely that scaling up the language models with more parameters and longer inputs will improve the quality of a legal summarization system.

It is my belief that the neural network approaches will dominate the summarization research for the coming decade. Currently, language model systems such as PEGASUS (Zhang et al. 2020) and BART (Lewis et al. 2020) are showing top results in summarizing text including legal text. In particular, more corpora have been recently made available for testing various summarization approaches, as in European legislative documents (Steinberger et al. 2006), U.S. congressional bills (Kornilova and Eidelman 2019), contracts (Manor and Li 2019), Indian Supreme Court cases[22] (Bhattacharya et al. 2019), U.S. Federal cases (Gargett et al. 2020). However, they will likely hit a ceiling if they do not incorporate more legal knowledge that draws from the relevant legal system and the legal concepts required to understand legal reasoning. Current legal language models may incorporate legal knowledge in only a superficial way and the danger of hallucination may prevent further progress. Returning to the roots of legal summarization as in Hachey and Grover (2006) may then be the way forward by incorporating higher-level

[22] Data is not publicly available.

rhetorical structures and legal reasoning into the next generation of neural network architectures.

7 Automatically classifying case texts and predicting outcomes (Ashley and Brüninghaus 2009). *Commentary by Henry Prakken*

Machine learning is currently hot in AI and Law, as in many other areas of AI. Some recent research on the prediction of case outcomes has even made it into the mainstream press[23] and headlines like *Lawyers could be the next profession to be replaced by computers*[24] and *The robot lawyers are here - and they're winning*[25] have been used.

AI and Law researchers know that such claims are exaggerated but still the prospects of practical application of AI and Law research have never been better (Francesconi 2022). Yet a main limitation of much recent work on case outcome prediction is that the predictions are hard to explain, while explanation is of paramount importance in the legal field. Generally, knowledge-based approaches are better for explanation than data-driven ones but they suffer from the knowledge acquisition bottleneck since the manual encoding of a knowledge or a case base can be very labour-intensive. So is it is important to investigate whether the input to a symbolic reasoning model can be automatically extracted from natural-language sources. Ashley and Brüninghaus (2009) were the first in our field to address this issue. Given the current explosion of interest in legal text analytics, there is every reason to discuss this paper's influence on and relevance for current research.

In doing so, I will not comment on the natural-language processing aspects of the paper, not only because that is not my expertise but also because the developments in NLP are going so fast that the conclusion that Ashley and Brüninghaus's NLP methods are now outdated is inevitable (which, however, does not at all detract from the paper's value). Rather, I will take the perspective of someone who is interested in how AI and Law models of legal reasoning could be used to support legal professionals. One reason for taking an application-oriented approach is the above-noted recent general interest in our field. Unlike in 2009, both the legal world and the general public now pay attention to our research and often our results are (rightly or wrongly) interpreted as indicating practical applicability. This means that we now have an additional responsibility besides the usual methodological ones, namely, to explain the societal relevance of our results.

[23] https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-college-london-computer-scientists about Aletras et al. (2016), https://www.nrc.nl/nieuws/2020/12/30/robot-weet-welke-uitspraak-het-hof-zal-doen-a4025683 about Medvedeva et al. (2020) (see Sect. 6 of Villata et al. (2022), elsewhere in this issue) and quite recently https://www.dailymail.co.uk/news/article-10346933/China-develops-AI-prosecutor-press-charges-97-accuracy.html.

[24] https://www.cnbc.com/2017/02/17/lawyers-could-be-replaced-by-artificial-intelligence.html.

[25] https://www.bbc.com/news/technology-41829534.

7.1 Summary of the paper

IBP (Brüninghaus and Ashley 2003) is a descendant of the HYPO (Ashley 1990) and CATO (Aleven 2003) systems and predicts outcomes of US trade secret misappropriation cases. In doing so, it combines rule-based and factor-based reasoning. Cases are represented as two sets of factors favouring, respectively, the plaintiff and the defendant. IBP's knowledge model combines a logical decision tree with lists of pro-plaintiff and pro-defendant factors for each of the five leaves of the tree, called the *issues* (e.g. *did the plaintiff maintain secrecy?*, and *did the defendant obtain the secret by improper means?*). Issues are addressed with a prediction model that according to Ashley and Brüninghaus applies a kind of scientific evidential reasoning. Roughly, if all factors in the case favour the same side for that issue, then IBP predicts a win for that side on the issue (unless all these factors are 'weak'). Otherwise it retrieves precedents that contain all case factors on that issue. If all have the same outcome, then IBP predicts that outcome, otherwise it tests the hypothesis that the side that won the majority of precedents will win, by trying to explain away each precedent won by the other side; this attempt succeeds if the precedent contains a 'knock-out' factor that is not in the current case. IBP's notions of weak and knock-out factors are a refinement of the CATO factor model and are defined in terms of low, respectively, high predictive power for the side they favour. Finally, IBP's predictions on all the issues are combined in an overall prediction. In an evaluation experiment IBP outperformed 11 other outcome predictors and achieved a high accuracy score of 92%, which compares very favourably with many current machine-learning approaches.

In Brüninghaus and Ashley (2003) the presence of factors in a case was identified by human knowledge engineers. SMILE instead tries to automatically learn these factors from manually annotated 'case squibs', which are summaries of case texts like those that first year law students prepare in briefing cases. SMILE applies a supervised-learning method to three alternative text representations: the general bag-of-words method and two more knowledge-intensive methods of the authors' own making. In a validation experiment SMILE best recognised factors with the authors' methods but with all three methods its performance was quite modest. Moreover, when IBP was applied to the case base as classified by SMILE instead of by human knowledge engineers, its outcome-prediction accuracy dropped from 92 to 64%.

7.2 Discussion

Ashley and Brüninghaus acknowledge that the performance of SMILE+IBP is modest and that their use of squibs in the SMILE experiments instead of full-text case opinions limits the generality of their results. They nevertheless claim that their paper is a " ... milestone in the field of AI and Law; it marks the first time to our knowledge that a program can reason automatically about legal case texts." This is not at all exaggerated. The authors realised the importance of combining NLP and symbolic AI at a time when this was not at all obvious, and the scholarly quality of their paper is impressive. Having said so, the authors hardly comment on the practical applicability of their work. In 2009 this was understandable, since the big-data and machine-learning revolution had not yet started and the legal world hardly paid attention to our field. However, as I noted above, this has changed, so a discussion of practical applicability is in order.

Unlike most current legal case outcome predictors, SMILE-IBP is not data-driven but knowledge-based: it applies a reasoning model with legal background knowledge to a factor representation of a case to predict its outcome. Ashley and Brüninghaus claim that the explanations of SMILE+IBP are intelligible to legal professionals and that this is a step towards practical applicability. However, understandability does not imply usefulness while, moreover, neither the understandability of SMILE+IBP's output nor its usefulness for legal professionals has been experimentally tested. The same holds for Grabmair (2017)'s VJAP system, a descendant of SMILE+IBP. One thing to note here is that, strictly speaking, IBP does not reason about what to *decide* but about what to *predict* as a decision: witness its use of majority rules and the predictive power of weak and knock-out factors. So its arguments cannot be directly used in court.

Ashley (2019), quoting Aleven (2003), claims that predictive accuracy is a good (although not the only) measure of the reasonableness of a computational model of argument, but like understandability, reasonableness does not imply usefulness. Moreover, it has been argued that a prediction on its own, even if the algorithm has high predictive accuracy, does not give legal professionals any useful information, since it does not reflect the probability that a judge in the new case will take the predicted decision (Bex and Prakken 2021). This makes explanations on legal grounds for a prediction essential for the prediction's usefulness, which in turn calls for validation studies of such usefulness. However, this requires more than applying numerical performance measures like accuracy, precision and recall; what is needed are empirical validation studies with potential or actual users of the system.

Having said so, clearly the higher a system's predictive accuracy the better, regardless of the nature of the application. Here an issue is whether a combination of a symbolic reasoner with NLP for providing the inputs can outperform the pure NLP approaches that have recently become popular (e.g. Aletras et al. (2016) and Medvedeva et al. (2020)). That this may be the case is indicated by the fact that in the first experiment IBP significantly outperformed all methods that did not employ a model of legal argument. So employing a reasoning model in outcome prediction may benefit not only explainability but also predictive accuracy.

However, here an issue is whether the proposed factor-based reasoning models are general enough, even if enriched with value trade-offs as in VJAP. Both SMILE-IBP and VJAP implement rather specific reasoning models that may suit particular common-law domains but that may be hard to generalise to other legal domains or jurisdictions.

Finally, if the aim is to "predict and explain the outcomes of new cases" (Ashley and Brüninghaus's final words) then an often overlooked issue arises, namely, where to find the representation of the new case. In all current validation studies the cases in the test set are past cases like the training data, with the only difference that the outcome of the test case is hidden from the system. However, if the outcome of a new case is to be predicted, there is no case decision yet, so other case-related documents must be used, but these will often give much less reliable information about the new case than the final verdict. For example, Medvedeva et al. (2020), who apply NLP to the full text of cases of the European Court of Human Rights, use as the new case the so-called communicated case sent by the Court to the government that is sued which contains a summary of the facts of the case and questions to the Government. These communications are available well before the Court's decision. As shown in Medvedeva et al. (2022), however, the accuracy of forecasting outcomes when using this information is considerably lower than when classifying the cases according to outcome using information available after the trial.

7.3 Conclusion

Ashley and Brüninghaus's paper was indeed a milestone in AI and Law, being the first paper in our field that studied how symbolic models of legal reasoning can be applied directly to case texts instead of to human-crafted case representations. It thereby made an important step in overcoming the notorious knowledge acquisition bottleneck of symbolic approaches without giving up their advantages. However, much research is still needed to make this approach practically applicable. Clearly, advances in NLP are needed. Research on this is already being done, (Branting et al. 2021),[26] including the first legal applications of currently fashionable language models like BERT (Chalkidis et al. 2020; Zheng et al. 2021b; Tagarelli and Simeri 2021). Moreover, given the arguably limited generality of the symbolic legal-reasoning models underlying SMILE-IBP and VJAP, further work on developing such models is also needed. Or will an author celebrating the 50th anniversary of this journal in 2042 conclude that hybrid NLP-symbolic approaches were not the way to go and that purely data-driven approaches could solve the explanation problem on their own?

8 Argumentation mining (Mochales and Moens 2011). *Commentary by Adam Wyner*

The aim of Mochales and Moens (2011) is to automatically detect, classify and structure argumentation in text. The approach taken may be described as *sequential* and *top-down*, starting from *coarse-grained* and *independent* categorisations of propositions within a text to more *fine-grained* and *relational* categorisations. The techniques range across machine learning and rule-based context-free grammars. In developing the analyses and presenting the results, basic ideas are briefly reviewed, identifying what is relevant to their study and what is auxiliary or challenging, leaving such issues for future work.

[26] See Sect. 7 of Villata et al. (2022), elsewhere in this issue.

As part of argument and argumentation analysis, argumentation mining focusses on the identification, classification, and interrelationships amongst textual portions that represent arguments. The end result should not only be highly accurate categorisations of textual portions, but some reconstruction of the overall argumentation put forth. The discussion is highly scoped: for example, it does not tie the extracted arguments to abstract argumentation, e.g., Dung (1995), nor is any attempt made to translate expressions to a logic (Wyner et al. 2016). Succinct background overviews of some relevant areas in argumentation are given, covering formal logic, informal logic, psychology, philsophy, and computer science. Several applications are discussed.

The aim of argumentation mining is to automatically detect the argumentation of a document. Questions that need to be addressed include: *What are the units of individual arguments and argumentation?*; *What are the relations that hold between two arguments units and/or argumentation?*; *Can the units of arguments and/or argumentation be determined automatically?* and *Can relations amongst arguments be determined automatically?* From the review of related literature, pragmatic answers are provided so as to focus on argumentation mining itself. The elementary units are propositions that serve as premises and conclusions. Argumentation schemes may provide some further internal structure to arguments, by characterising the propositions and their roles in an argument, though the identification of argumentation schemes is not central to the work. Similarly, there are several relationships between arguments, such as coordination or subordination, which are beyond the scope of the article. For the purposes of the article, an argument can be extended from particular premises to a conclusion to interlinked arguments, where the conclusion of one argument is the premise of another, leading to a tree structure.

To carry out the experiments, two corpora were used. One was the Araucaria (Reed and Rowe 2004) corpus, which is a heterogeneous collection of texts across domains that are manually annotated, providing something of a gold standard. The other was European Court of Human Rights (ECHR) corpus, described in detail in Palau and Moens (2008), which represents complex legal arguments that have been manually annotated to create a gold standard.

On these corpora, several classification experiments are carried out. In the first experiment, the task is to classify each proposition as either argumentative or non-argumentative. As such, it is the first step in the analysis pipeline. Statistical classifiers were applied and the contributory features identified. A model was generated over the gold standard; high accuracy is reported for both corpora. Having classified a proposition as argumentative, the next experiment looks to classify statements as either a premise or a conclusion. There were two phases of classification, where different features are used that highlight the different roles. High levels of accuracy are reported against the benchmark. Continuing with the sequence of analysis, the next step is to detect the relations between arguments, in other words, how arguments chain together. For this, a different tack is taken, turning to rule-based parsing with a context-free grammar (CFG), starting from the sentence level. Keywords, rhetorical phrases, or grammatical categories might be taken to signal the argumentative role, which is then used in higher level grammatical rules. Some categories segment the text while allowing a conclusion of one argument to serve as the premise of another argument. In such a way, the tree structure of a larger scale argument can be structured from the text. The rationale for this change of technique (from statistical to rule-based) is not made entirely clear. Evaluated against the benchmark corpora, good results are reported for detecting argumentation structures.

The applications for argumentation mining in law are briefly discussed, e.g., managing cases, annotating conceptual portions of cases into issues, facts, etc., linking statutes with case law and legal arguments, and extracting legal arguments for reuse.

8.1 Discussion

As outlined above, the paper provides a helpful synopsis of issues, approaches, theories, and techniques relating to argumentation mining, both in general and for law, as well as presenting some details and novel results. Particularly helpful is the *sequential* and *top-down* approach, starting from *coarse-grained* and *independent* categorisations of propositions within a text to more *fine-grained* and *relational* categorisations.

One can say that the article was published at the start of what became a significant wave of research, resources, projects, and implementations on argumentation mining.[27] The (non-exhaustive) list below includes several subsequent state-of-the-art articles, meetings, and a book:

- Peldszus and Stede (2013);
- Argument mining workshop series (2014–2020);[28]
- Since 2014, the IBM Debater[29] project has been developing argumentation technologies at industrial scale with a constant publication or resources and articles;
- Lippi and Torroni (2016);
- A 2016 Dagstuhl Seminar on Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments;[30]
- Stede and Schneider (2018);
- Lawrence and Reed (2019).

Of course, as the work progressed, one could expect that areas would be filled in with greater detail or a more critical eye. For instance, there have been efforts to standardise some aspects of argument analysis with the Argument Interchange Format (Rahwan and Reed 2009). The development, delivery, and reporting of machine learning studies has become more sophisticated and detailed, enabling validation. Well-structured and carefully curated corpora have been open-sourced for ongoing research developments.[31] There has been work to deepen our understanding of

[27] Google Scholar gives 419 citations to the paper. Date of access: 16 May 2022.

[28] https://2021.argmining.org/index.html#previous_workshops.

[29] https://research.ibm.com/interactive/project-debater/.

[30] https://www.dagstuhl.de/en/program/calendar/semhp/?semnr=16161.

[31] See particularly the IBM Debater Datasets: https://research.ibm.com/haifa/dept/vst/debating_data.shtml.

rhetorical markers (Peldszus and Stede 2016) as well as the scope and variety of propositions (Jo et al. 2019). The analysis of features, which identify those most significant to the model, has progressed {Zheng and Casari (2018) and Dong and Liu (2018)). While the article cannot lay claim to have created all these topics, it would appear fair to say that it gave researchers a clear line-of-sight on what was feasible.

9 Popov v Hayashi special issue (Atkinson 2012). *Commentary by Katie Atkinson*

It is ten years since the journal of *Artificial Intelligence and Law* published a special issue, for which I served as guest editor, on approaches to modelling the case of *Popov v. Hayashi* (Atkinson 2012). The motivation for the special issue was to enable comparison of the use of a range of different AI techniques that could each be applied to the task of modelling a specific legal case, *Popov v. Hayashi*. The case concerned a dispute over ownership of a baseball that had financial value when it was struck during a game to score a home run by a famous player who broke a prominent record in doing so.[32] Looking back on the journal special issue and the papers it contains provides a timely opportunity to reflect on the techniques that were prominent at the time for modelling legal case-based reasoning and to consider how approaches to the task have evolved over the past decade.

The special issue contained four different papers that each used different AI techniques for the modelling task: the first paper by Bench-Capon (2012) makes use of dimensions (Rissland and Ashley 1987) and factors (Aleven and Ashley 1995); the second paper by Gordon and Walton (2012) makes use of Carneades, a tool to enable argument construction, evaluation and visualisation; (Gordon et al. 2007); the third paper by Prakken (2012) makes use of ASPIC+ (Prakken 2010), an abstract framework using strict and defeasible rules; the final paper by Wyner and Hoekstra (2012) is a demonstration of an ontology-based approach to capture conceptual knowledge and reasoning about the domain. Whilst each approach has different merits, as discussed in the individual papers, reflecting on the collection as a whole, a notable feature is that all the papers make use of symbolic AI techniques, which have a long tradition of being used to represent and reason about legal cases. Since the special issue was published, there have been many more papers appearing on the general topic of reasoning about legal cases. Some of these papers build on and extend the use of symbolic AI techniques, with a particular focus on the explanation features that these yield. However, in the past decade a whole new strand has opened up focusing on the use of machine learning techniques for undertaking the task of legal case prediction.

The application of data-driven techniques follows the trends that have been seen within the general field of AI whereby the availability of large data sets and increases in computational power enable established machine learning algorithms to

[32] The case was the subject of a 2004 comic documentary film, *Up For Grabs*, https://www.imdb.com/title/tt0420356/.

be trained on these data sets. In the case of law, legal data sets are now available for many different jurisdictions and over the past decade there has been a multitude of papers published whose objective is to predict the outcomes of legal cases, having been trained on data sets of case law. Representative examples from this literature can be found in Aletras et al. (2016), Medvedeva et al. (2020),[33] Şulea et al. (2017), Zhong et al. (2018).

A key evaluation metric that is frequently used to assess the performance of the machine learning techniques on case prediction is the accuracy level of the algorithms in correctly deciding the cases, as compared to the human decision that was made on the reported cases. Thus, in such approaches the focus shifts away from faithfully modelling the legal reasoning being undertaken when making a decision on a case and instead concentrates focus more narrowly on getting the outcome correct. However, beyond the evaluation metric of *accuracy* of the predictive systems, another key criterion for evaluation that has recently come into more focus is the *explainability* of the systems for giving the conclusions they produce. Similar to debates within the field of general AI, there are concerns about the lack of explanation features that are provided by machine learning approaches, particularly since there is growing interest in making use of AI techniques in real world legal practice where explanations from automated tools are vital.

Looking back to the set of papers appearing in the special issue on modelling *Popov v. Hayashi*, a notable feature of the symbolic approaches that have been used for modelling the case is the focus on capturing the legal reasoning undertaken by the judge to reach a decision on the case outcome. All four papers explicitly note that the models used serve the purpose of reconstructing the legal reasoning, albeit through the use of different techniques to achieve this, covering factor-based reasoning, argumentation and legal concept modelling. Such approaches naturally lead to an ability for the explanations to be produced about *how* the conclusions were reached through reasoning over the models. Although explanation facilities were not a primary focus of this set of papers, the reasoning captured in the models can be inspected to check on the fidelity to the human judge's reasoning. This is important to not only check that sound reasoning has been undertaken, but also to ensure that the reasoning has jurisprudential grounding.

The demand for explanation facilities within automated reasoners has become more vocal in recent years with AI now starting to be deployed in practice. In a recent paper (Atkinson et al. 2020b), a comprehensive survey was produced reviewing the past literature on explanation in AI and law, current developments and future challenges around this topic. There is interest from legal practitioners in investigating the viability of deploying such approaches in practice (see e.g. Al-Abdulkarim et al. (2019) for a description of such a pilot application) and the growing backlog of cases in courts around the world is urging consideration of how new technologies could be harnessed to tackle such issues that are leading to delayed delivery of justice.

[33] See Sect. 6 of Villata et al. (2022), elsewhere in this issue.

Whilst there has been a wealth of new research over the past decade on modelling, reasoning about, and predicting outcomes of, legal cases, the special issue on modelling the case of *Popov v. Hayashi* stands as a marker within the literature demonstrating appreciation of AI and law researchers' varied approaches to legal knowledge representation and reasoning. The research on this topic has since evolved, recognising different motivations and a wider array of approaches, as discussed more broadly in (Ashley 2019). Whereas the main concern of the approaches in the special issue had been to model legal reasoning, current approaches are now more motivated by the considerations that could lead to practical deployment, most importantly *scalability* and *explanation*. Data-driven approaches (e.g. Medvedeva et al. (2020)) directly address the scalability aspects, but are as yet deficient in explanation. Methodologies to improve scalablity in symbolic approaches have been proposed (e.g. Al-Abdulkarim et al. (2016)). There is also currently a keen interest in pursuing development of hybrid systems that can combine the scalablity of data-driven approaches with the explanations available from symbolic approaches; see Branting et al. (2021) and Mumford et al. (2021) for recent proposals along these lines.

Curating a collection of papers today on the topic of modelling legal cases would likely yield work covering a broad set of approaches using a wide range of AI techniques. Among those that have been advanced since the appearance of the special issue are: machine learning approaches, using both Support Vector Machines (Medvedeva et al. 2020) and neural networks (Chalkidis et al. 2019); symbolic approaches based on use of social values (Grabmair 2017), argument schemes (Prakken et al. 2015), and abstract dialectical frameworks (Al-Abdulkarim et al. 2016); and hybrid systems using both machine learning and symbolic models (Branting et al. 2021). However, there are many insightful conclusions from the papers in the 2012 special issue that still stand as important contributions and enablers for the new developments within today's research on the topic: the use of precedent cases, the role of argument visualisation, approaches to domain conceptualisation, and the fidelity of the reconstruction of the judge's decision, are a representative sample of the issues drawn out across the papers appearing in the special issue.

As a final reflection, the case of *Popov v. Hayashi*, introduced into the AI and Law literature in Wyner et al. (2007), remains a personal favourite given the rich collection of literature that study of the case has led to, as exemplified by the 2012 special issue of the AI and Law journal.

10 A factor-based definition of precedential constraint (Horty and Bench-Capon 2012). *Commentary by Giovanni Sartor*

This article provides a theory of precedential constraint, namely, of the way in which a set of precedents constrains the space for new decisions which are consistent with that set. This outcome is achieved by consolidating and developing previous lines of research in AI and Law and legal theory: the factor-based approach to case-based reasoning of HYPO and CATO (Ashley (1990) and Aleven (2003)), its reinterpretation/extension through logics for defeasible argumentation (Prakken and Sartor 1998) and its representation as partial orders on sets of factors (Bench-Capon 1999), the evolution of case law (Bench-Capon and Sartor 2003), and the analysis of the connection between *rationes decidendi* and cases in legal theory (Raz 1979; Alexander 1989; Lamond 2005). This paper develops ideas presented by John Horty in earlier contributions, where two models of precedential constraints were presented: the results model, where precedential constraint only depends the factors in the precedent (Horty 2004) and the reason model, where the constraint also depends on a rule in the precedent (Horty 2011). The two approaches correspond to traditional approaches to precedent in legal theory, one focusing on the facts of the precedent (so that a precedent governs new cases sharing the same facts), e.g. Alexander (1989), the other on the reason for the decision or *ratio decidendi* stated by judges (so that a precedent governs all new cases to which the precedent's rule can be applied), e.g. Lamond (2005).

The paper focuses on the reason model, assuming that the representation of a precedent includes the following:

- an outcome o (which is assumed to be binary, either for the plaintiff (π) or for the defendant (δ))
- a set of factors, consisting of two disjoint subsets, the factors F^{π} for the outcome and the factors F^{δ} against the outcome.
- a reason, which includes a subset $S^{o} \subseteq F^{o}$ of the factors for the outcome.

The key message of the precedent $< F^{\pi} \cup F^{\delta}, o, S^{o} \rightarrow o >$ concerns the relation between these sets of factors. In the results model, the message of a case having outcome o (which may be π or δ) is that the set of all factors supporting that outcome (respectively F^{π} or F^{δ}) outweighs the set of all factors against the outcome (F^{δ} or F^{π}). The fact that the first set outweighs the second is assumed to explain the decision. In the reason model the message is different: the set of factors S^{o} providing the reason for o—which is a subset of all factors supporting that outcome—is deemed to be sufficiently strong to outweigh on its own all factors in the case against that outcome.

Consider a case $c =< \{f_1^{\pi}, f_2^{\pi}, f_1^{\delta}, f_2^{\delta}\}, \pi, \{f_1^{\pi}\} \rightarrow \pi >$. For c, the results model tells us that $\{f_1^{\pi}, f_2^{\pi}\}$ outweighs $\{f_1^{\delta}, f_2^{\delta}\}$, while the reason model tells us that f_1^{π} alone is sufficient to outweigh $\{f_1^{\delta}, f_2^{\delta}\}$. Thus in the reason model, the reason indicated by the judges may include only a strict subset of the factors for the outcome and so provides a stronger message, having a broader scope of application. In both models the message that is extracted from a case is extended a fortiori: if a case tells us that a set of factors X for an outcome prevails over a set Y for the opposite outcome, this entails that any superset of X also prevails over any subset of Y. This corresponds to the fact that by adding new factors favouring an outcome to a set of factors favouring that outcome, we should obtain a stronger set of factors for that outcome; and similarly, by subtracting some factors favouring the same outcome from a set of factors favouring that outcome, we should obtain a weaker set of factors for that outcome.

The priorities between sets of factors extracted from cases have implications for the consistency of case-bases: a case-base is inconsistent when the precedents in it entail conflicting preferences: e.g. it contains both a case c_1 telling us that the set of factors X for an outcome prevails over a set of factors Y for the opposite outcome, and a case c_2 telling us that on the contrary Y prevails X.

The requirement of consistency within a case-base (a set of precedents) may dictate the decision in certain new cases (unless the decision maker decides to overrule some precedents). Other cases may not be so constrained: the new decision maker will then be fully free, or may have a choice between following or distinguishing the precedent.

For instance, assume that a case base contains a precedent $c_1 =< F^{\pi} \cup F^{\delta}, \pi, S^{\pi} \rightarrow \pi >$. The addition of a new case $c_2 =< F'^{\pi} \cup F'^{\delta}, \delta, S'^{\delta} \rightarrow \delta >$, such that S^{π} is included in F'^{π} ($S^{\pi} \subseteq F'^{\pi}$) and F'^{δ} is included in F^{δ} ($F'^{\delta} \subseteq F^{\delta}$), would lead to an inconsistency. In fact c_1 indicates that S^{π} outweighs F^{δ}, while the c_2 indicates that a subset of F^{δ} outweighs a superset of S^{π}: whereas if S^{π} outweighs F^{δ}, equally or *a fortiori* S^{π} should outweigh F'^{δ}.

In contrast, if the new case c_2 is such that either S^{π} is not included in F'^{π} or that F'^{δ} is not included in F^{δ} then c_2 can be decided for δ consistently with c_1. If S^{π} is not included in F'^{π} (some decisive factors supporting π are missing in c_2) then the judge of the new case is not constrained at all by the precedent, and can decide as it sees fit. If c_2 includes S^{π}, but F'^{δ} is not included in F^{δ}, the new decision maker has the choice between following the precedent, i.e., deciding c_2 for π on the basis of S^{π} or a superset of it, or making a distinction, i.e., deciding c_2 for δ, based on a reason S'^{δ} containing some factors for δ not included in c_1. These ideas in the paper are expanded by taking into account whole sets of precedents (case-bases) and the preferences between sets of factors implied by all of them, an aspect which cannot be developed here.

The paper has the merit of synthesising and clarifying several strands of previous work on case-based reasoning within the AI and Law community, and linking this work to legal theories of precedent. It set the scene for multiple subsequent developments within the AI and Law research.

In Rigoni (2015) the notion of a *framework* precedent was introduced. Such precedents do not determine preferences between sets of factors, but instead set out the issues that must be considered.

There followed the development of models of precedential constraints that take into account not only binary factors but also multivalued dimensions, called *factors with magnitude* by Horty (2017, 2021), Rigoni (2018) and Bench-Capon and Atkinson (2018). These developments enabled the consideration of cases where the extent to which a factor was satisfied was important in determining its significance. A formal comparison of the result and reason models and several approaches to modelling dimensions which summarises all this work was given in Prakken (2021).

In Bench-Capon and Atkinson (2021) it was argued that there was a third kind of precedent, *ascription* precedents, which expressed not preferences between sets of factors but whether a particular dimension favoured the plaintiff or the defendant. Reasoning with cases can be seen as a two stage process: first the facts of the case must be qualified to identify the factors, and then these factors balanced to give the outcome. Precedents constrain both stages: only the second is considered in Horty and Bench-Capon (2012).

Bench-Capon and Atkinson (2021) also suggested that precedential constraint should be applied at the level of issues rather than to cases as a whole, recalling the insistence on the representation of *multi-step* arguments in cases in Prakken and Sartor (1998).[34] Applying the constraint at the level of cases rather than issues means that irrelevant distinctions may prevent the precedent constraining cases which should be constrained.

The development of understanding of reasoning with precedent cases has been a major success story in AI and Law: Horty and Bench-Capon (2012) made an important contribution to this understanding by providing an initial formal account of how precedents constrain future decisions.

Funding No funding was received for conducting this study. No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Agnoloni T, Barrera MF, Sagri MT, Tiscorni D, Venturi G (2009) When a framenet-style knowledge description meets an ontological characterization of fundamental legal concepts. In: Palmirani M, Pagallo U, Casanovas P, Sartor G (eds) AI approaches to the complexity of legal systems. Lecture notes in computer science. Springer, Berlin, pp 93–112

Al-Abdulkarim L, Atkinson K, Bench-Capon T (2015) Factors, issues and values: revisiting reasoning with cases. In Proceedings of the 15th international conference on artificial intelligence and law, pp 3–12

Al-Abdulkarim L, Atkinson K, Bench-Capon T (2016) A methodology for designing systems to reason with legal cases using ADFs. Artif Intell Law 24(1):1–49

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

Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V (2016) Predicting judicial decisions of the European Court of Human Rights. PeerJ Comput Sci 2:e93

Aleven V (2003) Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. Artif Intell 150(1–2):183–237

Aleven V, Ashley KD (1995) Doing things with factors. In: Proceedings of the 5th international conference on artificial intelligence and law, pp 31–41

Alexander L (1989) Constrained by precedent. South Calif Law Rev 63:1

Allen LE, Saxon CS (1986) Analysis of the logical structure of legal rules by a modernized and formalized version of Hohfeld legal conceptions. In: Martino A, Natali F (eds) Automated analysis of legal texts. North Holland, Amsterdam

[34] See Governatori et al. (2022), Sect. 9, elsewhere in this issue.

Allen LE, Saxon CS (1993) A-Hofeld: a language for robust structural representation of knowledge in the legal domain to build interpretation-assistance expert systems. In: Meyer J-J, Wieringa R (eds) Deontic logic in computer science: normative system specification. Chichester, Wiley, pp 205–224
Araszkiewicz M (2021) Critical questions to argumentation schemes in statutory interpretation. J Appl Log 8(1):291–320
Araszkiewicz M, Bench-Capon T, Francesconi E, Lauritsen M, Rotolo A (2022) Thirty years of AI and law: overviews. Artif Intell Law 30(2):147–161
Ashley KD (1990) Modeling legal arguments: reasoning with cases and hypotheticals. MIT Press, Cambridge
Ashley KD (2019) A brief history of the changing roles of case prediction in AI and law. Law Context Socio-Legal J 36:93
Ashley KD, Brüninghaus S (2009) Automatically classifying case texts and predicting outcomes. Artif Intell Law 17(2):125–165
Atkinson K (ed) (2012) Artificial intelligence and law: special issue on modelling Popov v Hayashi vol 20:1
Atkinson K, Bench-Capon T (2021) Argumentation schemes in AI and Law. Argum Comput 12(3):417–434
Atkinson K, Bench-Capon T, McBurney P (2006) PARMENIDES: facilitating deliberation in democracies. Artif Intell Law 14(4):261–275
Atkinson K, Bench-Capon T, Bex F, Gordon TF, Prakken H, Sartor G, Verheij B (2020) In memoriam Douglas N. Walton: the influence of Doug Walton on AI and Law. Artif Intell Law 28(3):281–326
Atkinson K, Bench-Capon T, Bollegala D (2020b) Explanation in AI and law: past, present and future. Artif Intell, p 103387
Bench-Capon T (1989) Deep models, normative reasoning and legal expert systems. In: Proceedings of the 2nd international conference on artificial intelligence and law, pp 37–45
Bench-Capon T (1999) Some observations on modelling case based reasoning with formal argument models. In Proceedings of the 7th international conference on artificial intelligence and law, pp 36–42
Bench-Capon T (2002) The missing link revisited: the role of teleology in representing legal argument. Artif Intell Law 10(1):79–94
Bench-Capon T (2012) Representing Popov v Hayashi with dimensions and factors. Artif Intell Law 20(1):15–35
Bench-Capon T (2020) The need for Good Old Fashioned AI and Law. In: Hötzendorfer W, Tschohl C, Kummer F (eds) International trends in legal informatics: a Festschrift for Erich Schweighofer. Weblaw, Bern, pp 23–36
Bench-Capon T, Atkinson K (2018) Lessons from implementing factors with magnitude. In: Proceedings of Jurix-2018, pp 11–20. IOS
Bench-Capon T, Atkinson K (2021) Precedential constraint: the role of issues. In: Proceedings of the 18th international conference on artificial intelligence and law, pp 12–21
Bench-Capon T, Modgil S (2009) Case law in extended argumentation frameworks. In: Proceedings of the 12th international conference on artificial intelligence and law, pp 118–127
Bench-Capon T, Prakken H (2010) Using argument schemes for hypothetical reasoning in law. Artif Intell Law 18(2):153–174
Bench-Capon T, Sartor G (2003) A model of legal reasoning with cases incorporating theories and values. Artif Intell 150(1–2):97–143
Bench-Capon T, Coenen F, Orton P (1993) Argument-based explanation of the British Nationality Act as a logic program. Inf Commun Technol Law 2(1):53–66
Bench-Capon T, Araszkiewicz M, Ashley K, Atkinson K, Bex F, Borges F, Bourcier D, Bourgine P, Conrad JG, Francesconi E et al (2012) A history of AI and Law in 50 papers: 25 years of the international conference on AI and law. Artif Intell Law 20(3):215–319
Berman DH, Hafner CL (1991) Incorporating procedural context into a model of case-based legal reasoning. In: Proceedings of the 3rd international conference on artificial intelligence and law, pp 12–20
Berman DH, Hafner CL (1993) Representing teleological structure in case-based legal reasoning: The missing link. In: Proceedings of the 4th international conference on artificial intelligence and law, pp 50–59
Berman DH, Hafner CL (1995) Understanding precedents in a temporal context of evolving legal doctrine. In: Proceedings of the 5th international conference on artificial intelligence and law, pp 42–51

Bex F (2011) Arguments, stories and criminal evidence: a formal hybrid theory. Springer, Berlin
Bex F, Prakken H (2021) On the relevance of algorithmic decision predictors for judicial decision making. In: Proceedings of the 18th international conference on artificial intelligence and law, pp 175–179
Bex F, Verheij B (2012) Solving a murder case by asking critical questions: an approach to fact-finding in terms of argumentation and story schemes. Argumentation 26(3):325–353
Bex F, Verheij B (2013) Legal stories and the process of proof. Artif Intell Law 21(3):253–278
Bex F, Prakken H, Reed C, Walton D (2003) Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. Artif Intell Law 11(2):125–165
Bhattacharya P, Hiware K, Rajgaria S, Pochhi N, Ghosh K, Ghosh S (2019) A comparative study of summarization algorithms applied to legal case judgments. In: Advances in information retrieval—41st European conference on IR research, ECIR 2019, pp 413–428
Branting LK (1991) Reasoning with portions of precedents. In: Proceedings of the 3rd international conference on artificial intelligence and law, pp 145–154
Branting LK (1993) A computational model of ratio decidendi. Artif Intell Law 2(1):1–31
Branting LK, Pfeifer C, Brown B, Ferro L, Aberdeen J, Weiss B, Pfaff M, Liao B (2021) Scalable and explainable legal prediction. Artif Intell Law 29(2):213–238
Breuker J, Elhag A, Petkov E, Winkels R (2002) IT support for the judiciary: Use of ontologies in the e-court project. In: Proceedings of the 10th international conference on conceptual structures, integration and interfaces, pp 15–19
Breuker J, Hoekstra R et al. (2004a) Core concepts of law: taking common-sense seriously. In: Proceedings of formal ontologies in information systems, pp 210–221
Breuker J, Tiscornia D, Winkels R, Gangemi A (eds) (2004) Artificial intelligence and law: special issue on ontologies for law 12:4
Breuker J, Valente A, Winkels R (2004) Legal ontologies in knowledge engineering and information management. Artif Intell Law 12(4):241–277
Brüninghaus S, Ashley KD (2003) Predicting outcomes of case based legal arguments. In: Proceedings of the 9th international conference on artificial intelligence and law, pp 233–242
Chalkidis I, Androutsopoulos I, Aletras N (2019) Neural legal judgment prediction in english. arXiv preprint arXiv:1906.02059
Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I (2020) Legal-bert: the muppets straight out of law school. arXiv preprint arXiv:2010.02559
Cheng J, Lapata M (2016) Neural summarization by extracting sentences and words. In: Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 1: Long Papers), Berlin, Germany, pp 484–494
Chorley A, Bench-Capon T (2005) An empirical investigation of reasoning with legal cases through theory construction and application. Artif Intell Law 13(3):323–371
de Oliveira Lima JA, Griffo C, Almeida JPA, Guizzardi G, Aranha MI (2021) Casting the light of the Theory of Opposition onto Hohfeld's fundamental legal concepts. Legal Theory, pp 1–34
Devlin J, Chang M-W, Lee K, Toutanova K (2019) 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, vol 1, pp 4171–4186
Dong G, Liu H (eds) (2018) Feature engineering for machine learning and data analytics. CRC Press, Boca Raton
Dung PM et al (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif Intell 77(2):321–357
de Vargas Feijo D, Moreira VP (2021) Improving abstractive summarization of legal rulings through textual entailment. Artificial Intelligence and law, pp1–23
Francesconi E (2014) A description logic framework for advanced accessing and reasoning over normative provisions. Artif Intell Law 22(3):291–311
Francesconi E (2022) The winter, the summer and the summer dream of artificial intelligence in law. Artif Intell Law, pp 1–15
Gargett A, Firth R, Aletras N (2020) Legalops: a summarization corpus of legal opinions. In: IEEE international conference on big data 2020, pp 2117–2120
Gordon TF (1993) The Pleadings game. Artif Intell Law 2(4):239–292
Gordon TF (2008) Constructing legal arguments with rules in the legal knowledge interchange format (LKIF). In: Computable models of the law. Springer, Berlin, pp 162–184

Gordon TF, Walton D (2009) Legal reasoning with argumentation schemes. In: Proceedings of the 12th international conference on artificial intelligence and law, pp 137–146
Gordon TF, Walton D (2012) A Carneades reconstruction of Popov v Hayashi. Artif Intell Law 20(1):37–56
Gordon TF, Prakken H, Walton D (2007) The Carneades model of argument and burden of proof. Artif Intell 171(10–15):875–896
Governatori G, Bench-Capon T, Verheij B, Araszkiewicz M, Francesconi E, Matthias G (2022) Thirty years of AI and law: the first decade. Artif Intell Law 30(4)
Grabmair M (2017) Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs. In: Proceedings of the 16th international conference on artificial intelligence and law, pp 89–98
Grabmair M, Ashley KD (2011) Facilitating case comparison using value judgments and intermediate legal concepts. In: Proceedings of the 13th international conference on artificial intelligence and law, pp 161–170
Greenwood K, Capon TB, McBurney P (2003) Towards a computational account of persuasion in law. In: Proceedings of the 9th international conference on artificial intelligence and law, pp 22–31
Griffo C, Almeida JPA, Guizzardi G (2016) Legal relations in a core ontology of legal aspects based on Alexy's theory of constitutional rights. In: Proceedings of JURIX 2016
Hachey B, Grover C (2006) Extractive summarisation of legal texts. Artif Intell Law 14(4):305–345
Hafner CD, Berman DH (2002) The role of context in case-based legal reasoning: teleological, temporal, and procedural. Artif Intell Law 10(1):19–64
Hage J (1997) Reasoning with rules: an essay on legal reasoning and its underlying logic. Kluwer Academic Publishers, Dordrecht
Hage J, Leenes R, Lodder AR (1993) Hard cases: a procedural approach. Artif Intell Law 2(2):113–167
Hahn U, Oaksford M, Harris A (2013) Testimony and argument: a bayesian perspective. In: Bayesian argumentation. Springer, Berlin, pp 15–38
Henderson J, Bench-Capon T (2019) Describing the development of case law. In: Proceedings of the 17th international conference on artificial intelligence and law, pp 32–41
Hoekstra R, Breuker J, Di Bello M, Boer A (2007) The LKIF core ontology of basic legal concepts. Proc LOAIT 07:43–63
Hohfeld WN (1913) Some fundamental legal conceptions as applied in judicial reasoning. Yale Law J 23(1):16–59
Horty J (2017) Reasoning with dimensions and magnitudes. In: Proceedings of the 15th international conference on artificial intelligence and law
Horty JF (2004) The result model of precedent. Leg Theory 10:19–31
Horty JF (2011) Rules and reasons in the theory of precedent. Leg Theory 10:1–33
Horty JF (2021) Modifying the reason model. Artif Intell Law 29:271–83
Horty JF, Bench-Capon T (2012) A factor-based definition of precedential constraint. Artif Intell Law 20(2):181–214
Ingolfo S, Siena A, Mylopoulos J, Susi A, Perini A (2013) Arguing regulatory compliance of software requirements. Data Knowl Eng 87:279–296
Jo Y, Visser J, Reed C, Hovy EH (2019) A cascade model for proposition extraction in argumentation. In: Proceedings of the 6th workshop on argument mining, pp 11–24
Jones A, Parent X (2008) Normative-informational positions: a modal-logical approach. Artif Intell Law 16(1):7–23
Jones A, Sergot M (1992) Deontic logic in the representation of law: towards a methodology. Artif Intell Law 1(1):45–64
Kanger S (1972) Law and logic. Theoria 38(3):105–132
Kelsen H (1967) Pure theory of law. University of California Press, Berkeley
Kornilova A, Eidelman V (2019) Billsum: a corpus for automatic summarization of us legislation. arXiv preprint arXiv:1910.00523
Lafferty JD, McCallum A, Pereira FCN (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th international conference on machine learning (ICML 2001), pp 282–289
Lamond G (2005) Do precedents create rules? Leg Theory 11(1):1–26
Lawrence J, Reed C (2019) Argument mining: a survey. Comput Linguist 45(4):765–818
Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L (2020) BART: denoising sequence-to-sequence pre-training for natural language generation, translation,

and comprehension. In: Proceedings of the 58th annual meeting of the association for computational linguistics, ACL 2020, pp 7871–7880
Lindahl L (1977) Position and change—a study in law and logic. Reidel, Dordrecht
Lippi M, Torroni P (2016) Argumentation mining: state of the art and emerging trends. ACM Trans Internet Technol 16(2):10:1-10:25
Loui RP, Norman J (1995) Rationales and argument moves. Artif Intell Law 3(3):159–189
Manor L, Li JJ (2019) Plain (e)nglish summarization of contracts. arXiv preprint arXiv:1906.00424
Maranhão J, de Souza EG, Sartor G (2021) A dynamic model for balancing values. In: Proceedings of the 18th international conference on artificial intelligence and law, pp 89–98
Marshall CC (1989) Representing the structure of a legal argument. In: Proceedings of the 2nd international conference on artificial intelligence and law, pp 121–127
McCarty LT (1986) Permissions and obligations: an informal introduction. In: Martino A, Socci F (eds) Automated analysis of legal texts. North Holland, Amsterdam
Medvedeva M, Vols M, Wieling M (2020) Using machine learning to predict decisions of the European Court of Human Rights. Artif Intell Law 28(2):237–266
Medvedeva M, Wieling M, Vols M (2022) Rethinking the field of automatic prediction of court decisions. Artif Intell Law, pp 1–18
Mochales R, Moens M-F (2011) Argumentation mining. Artif Intell Law 19(1):1–22
Mumford J, Atkinson K, Bench-Capon T (2021) Explaining factor ascription. In: Proceedings of JURIX 2021. IOS Press, pp 191–196
Muthuri R, Boella G, Hulstijn J, Capecchi S, Humphreys L (2017) Compliance patterns: harnessing value modeling and legal interpretation to manage regulatory conversations. In: Proceedings of the 16th international conference on articial intelligence and law, pp 139–148
Nallapati R, Zhai F, Zhou B (2017) Summarunner: a recurrent neural network based sequence model for extractive summarization of documents. In: Proceedings of the 31st AAAI conference on artificial intelligence
Palau RM, Moens M-F (2008) Study on the structure of argumentation in case law. In: Proceedings of JURIX, pp 11–20
Pascucci M, Sileno G (2021) The search for symmetry in Hohfeldian modalities. In: International conference on theory and application of diagrams, pp 87–102
Peldszus A, Stede M (2013) From argument diagrams to argumentation mining in texts: a survey. Int J Cogn Inform Nat Intell 7(1):1–31
Peldszus A, Stede M (2016) Rhetorical structure and argumentation structure in monologue text. In: Proceedings of the 3rd workshop on argument mining, pp 103–112
Pollock JL (1987) Defeasible reasoning. Cogn Sci 11(4):481–518
Prakken H (2002) An exercise in formalising teleological case-based reasoning. Artif Intell Law 10(1):113–133
Prakken H (2010) An abstract framework for argumentation with structured arguments. Argum Comput 1(2):93–124
Prakken H (2012) Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and Dungean semantics. Artif Intell Law 20(1):57–82
Prakken H (2021) A formal analysis of some factor-and precedent-based accounts of precedential constraint. Artif Intell Law 29(4):559–585
Prakken H, Sartor G (1996) A dialectical model of assessing conflicting arguments in legal reasoning. Artif Intell Law 4(3–4):331–368
Prakken H, Sartor G (1998) Modelling reasoning with precedents in a formal dialogue game. Artif Intell Law 6(2–4):231–287
Prakken H, Reed C, Walton D (2005) Dialogues about the burden of proof. In: Proceedings of the 10th international conference on artificial intelligence and law, pp 115–124
Prakken H, Wyner A, Bench-Capon T, Atkinson K (2015) A formalization of argumentation schemes for legal case-based reasoning in ASPIC+. J Log Comput 25(5):1141–1166
Rahwan I, Reed C (2009) The Argument Interchange Format. In: Argumentation in artificial intelligence, pp 383–402. Springer
Raz J (1979) The authority of law: essays on law and morality. Clarendon, Oxford
Reed C, Rowe G (2004) Araucaria: software for argument analysis, diagramming and representation. Int J Artif Intell Tools 13(04):961–979
Rigoni A (2015) An improved factor based approach to precedential constraint. Artif Intell Law 23(2):133–160

Rigoni A (2018) Representing dimensions within the reason model of precedent. Artif Intell Law 26:1–22
Rissland EL, Ashley KD (1987) A case-based system for Trade Secrets law. In: Proceedings of the 1st international conference on artificial intelligence and law, pp 60–66
Rissland EL, Xu X (2011) Catching gray cygnets: an initial exploration. In: Proceedings of the 13th international conference on artificial intelligence and law, pp 151–160
Ross A (1968) Directives and norms. Routledge and Kegan Paul Ltd., London
Saravanan M, Ravindran B (2010) Identification of rhetorical roles for segmentation and summarization of a legal judgment. Artif Intell Law 18(1):45–76
Sartor G (2002) Teleological arguments and theory-based dialectics. Artif Intell Law 10(1):95–112
Sartor G (2005) Legal reasonimg: a cognitive approach to law. Springer, Dordrecht
Sartor G (2006) Fundamental legal concepts: a formal and teleological characterisation. Artif Intell Law 14(1–2):101–142
Sartor G (2010) Doing justice to rights and values: teleological reasoning and proportionality. Artif Intell Law 18(2):175–215
Schreiber G, Wielinga B, de Hoog R, Akkermans H, Van de Velde W (1994) Commonkads: a comprehensive methodology for KBS development. IEEE expert 9(6):28–37
Searle JR (1969) Speech acts: an essay in the philosophy of language. Cambridge University Press, Cambridge
Sergot MJ (2001) A computational theory of normative positions. ACM Trans Comput Log 2(4):581–622
Sergot MJ (2013) Normative positions. In: Gabbay D, Horty J, Parent X, van der Meyden R, van der Torre L (eds) Handbook of deontic logic and normative systems. College Publications, Rickmansworth, pp 353–406
Sileno G, Boer A, van Engers T (2015) Revisiting constitutive rules. In: 6th workshop on AI approaches to the complexity of legal systems, pp 39–55
Spaak T (1994) The concept of legal competence: an essay in conceptual analysis. Dartmouth Pub Co, Hanover, NH
Stede M, Schneider J (2018) Argumentation mining, synthesis lectures on human language technologies. Morgan & Claypool Publishers, San Rafael
Steging C, Renooij S, Verheij B (2021) 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, pp 235–239
Steinberger R, Pouliquen B, Widiger A, Ignat C, Erjavec T, Tufis D, Varga D (2006) The JRC-Acquis: a multilingual aligned parallel corpus with 20+ languages. In: Proceedings of the fifth international conference on language resources and evaluation, LREC 2006, pp 2142–2147
Şulea O-M, Zampieri M, Vela M, van Genabith J (2017) Predicting the law area and decisions of French Supreme Court cases. In: Proceedings of the international conference recent advances in natural language processing, RANLP 2017, pp 716–722
Tagarelli A, Simeri A (2021) Unsupervised law article mining based on deep pre-trained language representation models with application to the Italian Civil Code. Artif Intell Law, pp 1–57
Teufel S, Moens M (1997) Sentence extraction as a classification task. In: Proceedings of the ACL/EACL'97 workshop on intelligent scalable text summarization
Teufel S, Moens M (2002) Summarizing scientific articles: experiments with relevance and rhetorical status. Comput Linguist 28(4):409–445
Toulmin SE (1958) The uses of argument. Cambridge University Press, Cambridge
Tran V, Le Nguyen M, Tojo S, Satoh K (2020) Encoded summarization: summarizing documents into continuous vector space for legal case retrieval. Artif Intell Law 28(4):441–467
Valente A (1995) Legal knowledge engineering: a modelling approach. IOS Press, Amsterdam
Valente A, Breuker J, Brouwer B (1999) Legal modeling and automated reasoning with ON-LINE. Int J Hum Comput Stud 51(6):1079–1125
Verheij B (2003) Artificial argument assistants for defeasible argumentation. Artif Intell 150(1–2):291–324
Verheij B (2003) Deflog: on the logical interpretation of prima facie justified assumptions. J Log Comput 13(3):319–346
Verheij B (2003) Dialectical argumentation with argumentation schemes: an approach to legal logic. Artif Intell Law 11(2):167–195
Verheij B (2016) Formalizing value-guided argumentation for ethical systems design. Artif Intell Law 24(4):387–407

Villata S, Araszkiewicz M, Ashley K, Bench-Capon T, Branting LK, Conrad JG, Wyner A (2022) Thirty years of AI and Law: The third decade. Artif Intell Law 30(4)
Walton D (1996) Argumentation schemes for presumptive reasoning. Lawrence Erlbaum Associates, New York
Wegner DM (2002) The illusion of conscious will. Bradford Books, MIT Press. Cambridge, ISBN 9780262731621
Winkels R, Boer A, Hoekstra R (2002) Clime: lessons learned in legal information serving. In: ECAI, pp 230–234
Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M et al (2020) Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, pp 38–45
Wyner A, Bench-Capon T (2009) Modelling judicial context in argumentation frameworks. J Log Comput 19(6):941–968
Wyner A, Hoekstra R (2012) A legal case OWL ontology with an instantiation of Popov v. Hayashi. Artif Intell Law 20(1):83–107
Wyner A, Bench-Capon T, Atkinson K (2007) Arguments, values and baseballs: representation of Popov v. Hayashi. In: Proceedings of JURIX 2007, vol 165, pp 151–160
Wyner A, van Engers T, Hunter A (2016) Working on the argument pipeline: through flow issues between natural language argument, instantiated arguments, and argumentation frameworks. Argum Comput 7(1):69–89
Yamada H, Teufel S, Tokunaga T (2019) Building a corpus of legal argumentation in Japanese judgement documents: towards structure-based summarisation. Artif Intell Law 27(2):141–170
Zhang J, Zhao Y, Saleh M, Liu P (2020) PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. In: Proceedings of the 37th international conference on machine learning, pp 11328–11339
Zheng A, Casari A (2018) Feature engineering for machine learning: principles and techniques for data scientists. O'Reilly Press, Sebastopol
Zheng H, Grossi D, Verheij B (2021a) Hardness of case-based decisions: a formal theory. In: Proceedings of the 18th international conference on artificial intelligence and law, pp 149–158
Zheng L, Guha N, Anderson BR, Henderson P, Ho DE (2021b) 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, pp 159–168
Zhong H, Guo Z, Tu C, Xiao C, Liu Z, Sun M (2018) Legal judgment prediction via topological learning. In: Proceedings of the 2018 conference on empirical methods in natural language processing

Authors and Affiliations

Giovanni Sartor[1,2] · Michał Araszkiewicz[3] · Katie Atkinson[4] · Floris Bex[5,6] · Tom van Engers[7] · Enrico Francesconi[8,9] · Henry Prakken[5,10] · Giovanni Sileno[7] · Frank Schilder[11] · Adam Wyner[12] · Trevor Bench-Capon[4]

1 University of Bologna, Bologna, Italy

2 European University Institute, Fiesole, Italy

3 Jagiellonian University, Krakow, Poland

4 University of Liverpool, Liverpool, UK

[5] Utrecht University, Utrecht, The Netherlands

[6] Tilburg University, Tilburg, The Netherlands

[7] University of Amsterdam, Amsterdam, The Netherlands

[8] National Research Council of Italy (IGSG-CNR), Florence, Italy

[9] European Parliament, Luxembourg, Luxembourg

[10] University of Groningen, Groningen, The Netherlands

[11] Thomson Reuters, Saint Paul, USA

[12] University of Swansea, Swansea, UK