Reasoning with Legal Cases: A Hybrid ADF-ML Approach

Jack Mumford , Katie Atkinson , and Trevor Bench-Capon
Department of Computer Science, University of Liverpool, UK

Abstract. Reasoning with legal cases has long been modelled using symbolic methods. In recent years, the increased availability of legal data together with improved machine learning techniques has led to an explosion of interest in data-driven methods being applied to the problem of predicting outcomes of legal cases. Although encouraging results have been reported, they are unable to justify the outcomes produced in satisfactory legal terms and do not exploit the structure inherent within legal domains; in particular, with respect to the *issues* and *factors* relevant to the decision. In this paper we present the technical foundations of a novel hybrid approach to reasoning with legal cases, using Abstract Dialectical Frameworks (ADFs) in conjunction with hierarchical BERT. ADFs are used to represent the legal knowledge of a domain in a structured way to enable justifications and improve performance. The machine learning is targeted at the task of factor ascription; once factors present in a case are ascribed, the outcome follows from reasoning over the ADF. To realise this hybrid approach, we present a new hybrid system to enable factor ascription, envisioned for use in legal domains, such as the European Convention on Human Rights that is used frequently in modelling experiments.

Keywords. Abstract Dialectical Frameworks, Argumentation Frameworks, Reasoning with legal cases, Hybrid machine learning-argumentation

1. Introduction

Modelling legal case-based reasoning has been a central concern within the field of AI and Law from its early days e.g. [1]. Many of the approaches that have been proposed and developed over the past three decades have used symbolic techniques since, recognising that when humans make decisions on legal cases they call upon their domain expertise, it was natural to aim to model this expertise in legal AI systems. However, in the past decade there has been an increase in the application of data-driven methods aimed at predicting legal cases, aligning with developments in the general field of AI where machine learning (ML) applications have become increasingly prominent. The concerns about lack of interpretability and explainability of ML systems that are widespread within the general field of AI very much apply to tools that are built to perform legal reasoning; if such tools are to be trusted for deployment in real world scenarios, serious attention must be paid to the their ability to accurately capture laws and legal reasoning in the manner currently required of human experts. Moreover, in law, explanation is essential: natural justice requires that decisions are explained to the parties.

In this paper we present new work, extending on the outline presented in [2], that enables an important step within a wider programme that aims to demonstrate how hybrid approaches to modelling legal case-based reasoning can be developed to enable learning from large volumes of data, whilst retaining crucial domain knowledge that is needed to reason about and explain automated decisions on legal cases. In section 2 we provide a brief summary review of some of the key contributions to date within the field of AI and law that provide computational models of legal reasoning. In section 3 we provide an overview of how Boolean Abstract Dialectical Frameworks (ADFs) [3] can be used to model knowledge of a legal domain. In section 4 we set out the details of our new hybrid system in which an ADF provides the legal domain knowledge representation, and the H-BERT model is used to learn how to ascribe the base-level factor input for the ADF. Section 5 reports the results of experiments evaluating the performance of our hybrid system. We conclude in section 6 with reflections on our study and its results, along with the next steps for building upon the new foundational hybrid approach that we have presented.

2. Approaches to Modelling Legal Reasoning in AI and Law

The broad process of reasoning with legal cases in common law jurisdictions is undertaken within the context of a body of case law, comprising previous decisions, whereby a new case must be decided in the light of these precedents. At a hearing each side presents arguments as to why they should win and these arguments will typically be based on the precedent decisions and legislation.

Early work on modelling legal reasoning demonstrated how production rules [4] and logic programming [5] could be used to model and explain legal case-based reasoning. A shift of focus to identify the arguments involved in case based reasoning was brought through the development of the HYPO system [6]. This work was built upon in the CATO system [7], which was developed to assist law school students in forming better case-based arguments. A key idea within CATO was to describe cases in terms of *factors*, which are legally significant abstractions of patterns of facts found in the cases of a domain. These factors were then organised within a hierarchy of increasing abstraction, with factors labelled with a polarity to show whether they support or oppose the presence of their parent. Moving upwards through abstract factors within the hierarchy leads to the legal *issues* that have to be resolved to reach a decision in the particular legal domain (US Trade Secrets in the case of HYPO and CATO).

In later work based on CATO, Issue-Based Prediction (IBP) [8] was developed with the aim of not only discovering and presenting arguments, but also predicting the outcomes of cases. Evaluation results showed a good level of accuracy (over 90%) where the domain model relied upon manual analysis, but when machine learning was used for ascribing factors, the accuracy level decreased (to around 70%) [9].

In a further development in the spirit of CATO-style approaches, [10] set out a methodology, called ANGELIC, for capturing case law and explaining conclusions drawn through reasoning over the model. The methodology makes use of a well established knowledge representation technique, Abstract Dialectical Frameworks (ADFs) [3], to capture the factors and relationships between them within a domain of case law. Once defined for a domain, an ADF can easily be transformed into a logic [10] or JAVA [11] program that, when supplied with the facts of a case, can determine an outcome for the case and provide acceptable arguments leading to this decision. A success rate of over

96% accuracy in replicating past decisions was reported in [10], reflecting the high level of domain expertise captured within the ADFs through manual knowledge acquisition tasks undertaken to build the domain model.

In recent years, there has been an increase in work aimed at performing the task of case prediction through the use of data-driven methods that learn from the large datasets now available. Key representative examples are the work presented in [12], [13] and [14]. Problems with these approaches include lack of accuracy (typically around 70-80%), degradation of performance as the training set ages, and lack of explanations. State-of-the-art transformer-based models [15] for NLP tasks are unsuitable for long document classification, and proposed solutions [16] such as hierarchical transformer methods [17] rely on very large data sets and significant pre-training in order to produce decent results. None of these problems apply to symbolic models.

Our motivation for the work set out in this paper is to bring together into a hybrid system benefits found within symbolic and data-driven approaches to legal case-based reasoning. The domain expertise captured within ADFs is vital for grounding and explaining reasoning within the law, yet we wish to make use of machine learning approaches where they can be usefully deployed for the relatively expensive task of factor ascription. In the next section we show how ADFs are used to capture domains, prior to presenting, in Section 4, the use of ML for enabling the ascription of factors as an integral part of our new hybrid method for case classification.

3. Representing Legal Knowledge with ADFs

We represent the Legal Knowledge following the ANGELIC Methodology [10]. The methodology results in an instantiated Abstract Dialectical Framework (ADF) [3] and a set of questions. The instantiated ADF can be represented as a Table in which each node is associated with an ID, an informative label describing the node, a list of the children of the node, and a set of acceptance conditions. Sample nodes from the ADF modelling the European Convention on Human Rights (ECHR) in [18] are shown in Table 1. The acceptance conditions all take the form of "ACCEPT" or "REJECT" followed by a body expression containing only children of the node and the appropriate logical operators. The acceptance conditions are prioritised so that they will be tested in order and when one succeeds, the others will be ignored. The final acceptance condition for a node is always a default, so no node is undecided. The use of disjunctions means that the accept and reject conditions are interleaved. The ADF is accompanied by a set of questions which are posed to the user and which determine which leaf notes are accepted. Thus the leaf nodes have acceptance conditions in terms of the responses to one or more questions, such as "ACCEPT IF $Q7 \geq Q9$".

All the nodes in the ADF are Boolean, accepted or rejected, as in CATO [7]. There has been some work to model the degrees of acceptance (e.g. [19], [20]) but we see this as a matter of factor ascription [21]. The base level factors forming the leaf nodes all represent reasons to decide for one side or the other, and may represent a judgement. Suppose we have a factor *SignificantDelay*. This factor would have an associated question, such as *Q6: Enter the delay in months*. Precedents will have indicated the length of delay considered significant and so, if 18 months was considered significant, we have as the acceptance condition for this leaf node "ACCEPT IF $Q6 \geq 18$". Thus once the ADF is reached, matters of extent are settled and we can deal only with Booleans.

Table 1.: Sample nodes from ADF in [18]

ID	Description	Children	Acceptance Conditions
1	Violation of Article 6	2,3,8,20,21	REJECT IF NOT is a victim OR NOT case is admissible ACCEPT IF the case was not fair or public OR victim was presumed guilty OR the victim did not have the minimum rights REJECT otherwise
4	The case is admissible	4,5	REJECT IF NOT the case is well-founded OR there was no significant disadvantage ACCEPT otherwise

As noted in Section 2, legal domain knowledge can be seen as exhibiting a hierarchical structure, as represented by the factor hierarchy in CATO [7]. At the top is a statement expressing the decision (e.g. *The Plaintiff should win*) that is determined by resolution of the relevant issues (e.g. *There was a confidential relationship*). The issues provide necessary and sufficient conditions for the decision. The issues are resolved by considering the balance between pro and con factors. The factors themselves (e.g. *KnewInfoConfidential*) are ascribed on the basis of the facts of the case (e.g. *The defendant acquired the information while employed by the plaintiff*). These reasoning steps are based upon precedents. There are three different kinds of precedent, as explained in [21], each appropriate to a different statement type. *Framework* precedents, supplement legislation to identify the issues used to resolve the root node. *Preference* precedents determine which way the balance of factors should fall when resolving issues. *Ascription* precedents provide sufficient conditions for assigning factors to a case on the basis of its facts.

This hierarchy can be straightforwardly modelled as an ANGELIC ADF. For the node representing the decision, the necessary and sufficient conditions from legislation and framework precedents are stated in the required form. For the issue nodes, which require a balance of factors in accordance with precedents, the set of preference precedents are translated into rules in the manner devised in [22]. The pro reasons in a precedent form a condition with ACCEPT as head, the con reasons form a condition with REJECT as head and the outcome in the precedent determines the priority between them. The default here reflects the burden of proof for the particular issue. Finally the conditions for the leaf nodes, which have question answers as children, use the ascription precedents to determine whether the factor is present given the facts represented by these answers.

This method of representation has been applied in a number of domains, both academic [10] and practical e.g. [23]. Article 6 of the European Convention on Human Rights (ECHR), which concerns the right to a fair trial, used by the machine learning systems described in [12] and [14], was modelled as an ADF in [18] and [11], on which the work in this paper is based.

4. Hybrid ADF/H-BERT Method

Our approach is to leverage domain knowledge in conjunction with a state-of-the-art H-BERT architecture [24]. We use a Python implementation of the ADF constructed specifically for Article 6 of the ECHR [18] to provide intermediate classifications of base-level factors, and train an independent H-BERT model for each base-level factor.

The ADF provides domain knowledge representation of the legal reasoning process, and the H-BERT models are responsible for learning how to ascribe the base-level factor input for the ADFs from the case documentation. Intuitively, the ADF is interpreted as a final set of fixed weights that is added on top of the H-BERT models.

As we are interested in producing case classifications that can be understood and justified in terms of genuine domain knowledge, it is important that the H-BERT models ascribe base-level factors in a sensible and appropriate manner. In the legal reasoning hierarchy, base-level factors are ascribed according to the facts of the case. ECHR case summaries neatly present such facts as bullets within the document, which enables their extraction via the use of regular expressions whilst processing the document in HTML format. Each document is thus segmented in accordance with the bullets (taken only from the section of the document entitled “THE FACTS”, discussed in more detail below) and a pre-trained RoBERTa [25] model is used to produce an intermediate output representation for each bullet. The RoBERTa outputs for each bullet are then used as the inputs for training a subsequent BERT model that classifies the whole document for a particular base-level factor. Each document is encoded according to the RoBERTa model one time only, since the model is pre-trained. However, we must train a unique BERT model for each base-level factor in order to capture the interactions between the segmented fact encodings that are relevant for that specific factor.

The pipeline can be visualised in Figure 1, and can be understood as progressing through six stages:

***Stage 1*:** A corpus is obtained by scraping cases from the HUDOC website[1] – the principal repository for ECHR legal case documentation. Cases are scraped in HTML format in order to facilitate effective processing.

***Stage 2*:** Each case in the corpus is processed and segmented to the fact level of representation, understood to correspond to the bulleted points in the document, which can be extracted from the HTML format. The text is further processed to remove irrelevant bullets, such as headings and enumeration markers. Only text from THE FACTS section is taken as pertaining to the fact level; other sections provide information either irrelevant to justifiable legal reasoning (e.g. the identities of the judges) or contain the reasoning itself, which would defeat the purpose if included.

***Stage 3*:** For each case in the corpus, each fact segment is tokenized and then encoded by a pre-trained RoBERTa model, where the same model is used for all segments, without fine-tuning.

***Stage 4*:** The RoBERTa encodings of a case document’s fact segments are used as input for the BERT models, where we have thirty-two BERT models – one for each base-level factor of the ADF (excluding admissibility related factors that are irrelevant for our corpus which consists solely of admissible cases). Each BERT model further encodes each RoBERTa input to provide overall document context.

***Stage 5*:** The BERT encodings are used as input in a feed-forward neural network (FFNN) that outputs a Boolean classification which is used to determine ascription or non-ascription of a particular base-level factor in the ADF.

***Stage 6*:** The Python program implementing the ADF then produces the outcome classification that follows from the base-level factor ascriptions.

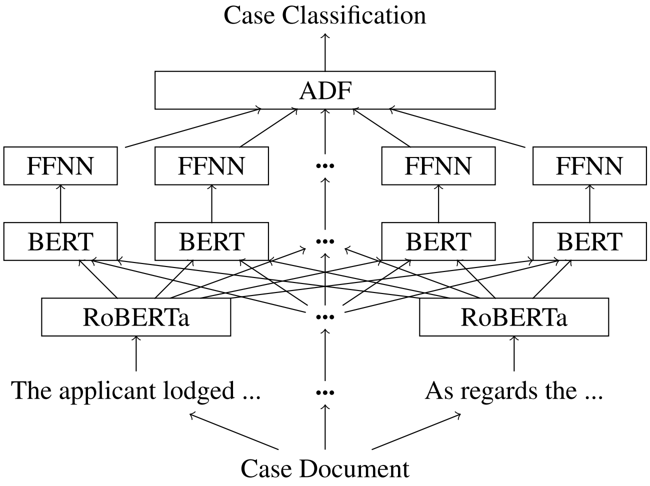

Figure 1. Feed-forward hybrid H-BERT/ADF model for ECHR Article 6 case classification.

During training, learning applies to adjusting weights only in the BERT models according to classifications provided from the ADF, which are in turn derived from processing the correct case outcome classification. The process extends through four steps:

***Step 1*:** Each case in the corpus is labelled according to its correct outcome classification: violation, or no-violation.

***Step 2*:** This classification is fed backwards through the ADF to produce probabilistic weights for the base-level factors, where the weight for ascription of any given base-level factor is the proportion of instances in which the factor is ascribed in correct outcome classification over all possible combinations of ascription.

***Step 3*:** Each base-level factor weight is used as the probability for ascribing the factor for classification with the relevant FFNN. For example if a base-level factor was assigned a probabilistic weight from the ADF of 0.2, then there would be a 20% chance that the FFNN would be provided a True Boolean classification, and 80% chance of a False Boolean classification.

***Step 4*:** The FFNN passes weight adjustments back to the appropriate BERT model, which in turn adjusts its weights accordingly. Learning does not propagate down to the RoBERTa models, which are fixed to their pre-training settings.

5. Experiments

In this section we assess the performance of our hybrid system on the legal case classification task. We first outline the details of the case corpus data set, and give the implementation details for our system[2]. We also compare the performance of the hybrid system against a H-BERT benchmark[3].

[1] hudoc.echr.coe.int/

[2] Undertaken on Barkla – High Performance Computing facilities, at the University of Liverpool, UK.

[3] Code available at: https://github.com/jamumford/LADF-HBERT.git

5.1. Data Set and Implementation Details

To build a relevant data set for experimentation necessitated the use of cases from January 2015 onwards, as the ADF was developed from official documentation and expert opinion relevant from this particular time point. As the law is subject to change over time, we decided to omit cases prior to January 2015 from analysis, due to the risk of incompatibility with the ADF model. We also restricted analysis to cases available only in English (note that when using the HUDOC site's filter for cases in English, roughly one in four of the filtered cases were incorrectly formatted and not available in English), which resulted in 575 cases between January 2015 and January 2022: 150 non-violation verdicts, and 425 violation verdicts.

The scarcity of the data is in stark contrast to the relatively vast data sets that are usually employed for NLP tasks. We focus on two classification approaches, a state-of-the-art hierarchical BERT approach developed specifically for small data sets which we refer to as H-BERT [24], and our hybrid system which uses the aforementioned H-BERT architecture in conjunction with the ADF layer as outlined in section 4. Both approaches use the same fact-level pre-trained RoBERTa model encodings using 512 tokens, and both use 256 tokens for document BERT model encoding.

Three metrics were selected for analysis as appropriate for the binary classification task: accuracy, macro F1 score, and MCC (Matthews correlation coefficient) score. Since the data set is unbalanced between non-violation and violation verdicts, the macro F1 and the MCC scores are of clear relevance, with the MCC score particularly suited to providing a representative score for performance over the full distribution.

In total forty experiments were conducted, twenty for each approach. Each experiment was randomly split 80% into training and 20% into test data, and underwent 30 epochs of training. Four Nvidia Tesla P100 GPUs were used for fine-tuning.

For the hybrid H-BERT/ADF system, in each epoch the training data was further split such that 10% was randomly assigned to a validation data set. At the end of an epoch the validation data set's MCC score was used to scale the factor ascription weights from the ADF layer. Hence a higher validation set MCC score would increase the likelihood of preserving the previous epoch's factor ascriptions for the subsequent epoch, whereas a lower MCC score would decrease the likelihood.

5.2. Results and Evaluation

The experimental results are summarised in Table 2 and provide a comparison of the performance of the H-BERT model against our hybrid H-BERT/ADF system. For all metrics, the hybrid system demonstrably outperforms the benchmark H-BERT model. These results can be reliably inferred as statistically significant, with negligible Mann-Whitney p values returned for each metric. The values $p < 0.001$ indicate that the hypothesis that the distributions for the two classification approaches (across all experiments) belong to the same population can be rejected at the 99.9% confidence level (and indeed the p values are so low that the hypothesis can be rejected at even higher confidence levels).

There is a greater degree of variance of results for the hybrid system in comparison with the H-BERT results, as indicated by the negative and positive range values. This greater variance is likely due to the inherent uncertainty of the thirty-two classification targets derived from the factor ascription weights produced by the ADF layer. Since the

Table 2.: Comparison of standard H-BERT model performance against Hybrid H-BERT/ADF system on ECHR Article 6 case outcome classification task. Results reported to 2d.p.

	Accuracy	Macro F1	MCC
H-BERT [24]	$66.78^{+0.17}_{-0.70}$	$60.16^{+0.23}_{-0.04}$	$17.90^{+0.03}_{-0.10}$
Hybrid H-BERT/ADF	$\mathbf{72.00}^{+8.87}_{-7.65}$	$\mathbf{67.47}^{+8.72}_{-5.58}$	$\mathbf{33.98}^{+18.66}_{-11.48}$
Mann-Whitley *p* values	< 0.001	< 0.001	< 0.001

factor ascription weights are treated as probabilities to ascribe a factor for a given case, each of the classification targets for each BERT model in the hybrid system will likely change over epochs, whereas they remain constant in the benchmark H-BERT approach.

Each H-BERT experiment required roughly five minutes of training for our given data set. Our hybrid system required roughly four hours, due mainly to the thirty-two H-BERT models (one for each base-level factor) that were necessary to train. However, it is worth noting that training only needs to be done once, and the resulting model could be used without the need for frequent updates, since the law changes relatively slowly in comparison to these processing times. Once training is completed, running a single case for classification requires negligible time (less than one second) when considered against practical timescales for evaluation.

6. Discussion and Summary

In our approach to argument-based modelling of legal reasoning, we have identified a particular role for machine learning: factor ascription rather than predicting cases as a whole, with ascription done by Hierarchical BERT applied to natural language descriptions of the case facts. Prediction is still possible, since the factors in a case determine the outcome. Law changes over time; models that cover a wide period of time are unlikely to be of great relevance in terms of producing justifiable outcomes. However, most ML approaches to case classification have focused on large data sets, which is understandable since state-of-the-art performance in NLP typically requires an abundance of data. But this is likely to set an upper limit on the usefulness of such models.

Factors are vital for acceptable explanations, since the justification of a case outcome is given in terms of how the issues were resolved considering the balance of factors present in the case. Use of factors for explaining the output of machine learning systems has been advocated in [26] and [27]. Exploiting the domain structure and learning factor ascription rather than case prediction should also improve performance. While formal approaches summarised in [28] treat cases as collections of factors, empirical approaches such as [8] and [29] decompose cases into issues. As shown in [21] applying precedential constraint at the level of cases rather than issues results in cases which should be constrained being not constrained, because they can be distinguished on factors unrelated to the issue in dispute in the case. This is also borne out by empirical work. In [8], while issue based prediction abstained on only one of the 186 test cases enabling accuracy of 91.4%, a similar system which treated the outcome as a single issue abstained on 50, reducing accuracy to 68.3%, a level in line with the legal machine learning approaches reported in [30]. A further reason for providing explanation in terms of factors is provided by [31], in which it was shown that a good level of performance by a machine learning system was no guarantee that it was applying the correct rationale when making its predictions. In the light of this, avoiding injustices and securing trust in the predictions demands an explanation, couched in legal terms, of the outcome. This in turn demands that the factors in a case be identified. The approach taken in this paper is broadly aligned with the observations made in [32], that cognitive computing for legal application will involve the interaction between an expert-derived factor-based model for higher reasoning and ML for lower-level processing of the raw document text. In future research we would like to assess our hybrid system in terms of explainability and justifiability of any given classification/prediction, via the fact-level attention weights produced by the individual H-BERT models. These attention weights might be useful for improving performance, presenting a feedback loop where experts approve or reject the facts selected by the system to justify factor ascription.

It should be noted that the effort involved in constructing the domain model has been found to be not disproportionate when applied to real world problems working with legal practitioners (e.g. [23]). An obstacle to practical deployment, however, is the effort required for manual analysis of cases into factors, both for constructing the case base and representing the new cases. Thus using machine learning to address the labour intensive task of ascribing factors to cases, while reserving the construction of the domain for experts with the appropriate knowledge, seems the sensible way to allocate resources. The time taken to train and use our hybrid system is fully in keeping with these expectations.

Our new H-BERT/ADF system lays the foundation for a hybrid approach to automating reasoning about legal cases, using both symbolic and data-driven techniques. Structured legal domain expertise is captured using the ANGELIC ADFs described in Section 3, which enable explanations of the outcomes of reasoning. Our ultimate aim is to use machine learning for the task of factor ascription, since factors must be ascribed for every case. We consider the results outlined in this paper as encouraging in moving towards this aim. Our next steps will be to incorporate higher levels of domain knowledge via the use of annotated data sets consisting of cases labelled by domain experts with respect to factor ascription. These annotations will be used to remove some of the uncertainty of the classification labels presented to the H-BERT models, both directly for the labelled cases, and indirectly by suggesting better priors to guide the ADF layer's weight propagation for factor ascription in accordance with the background distribution of factor ascriptions from the annotated data set. Comparison to alternative ML approaches in the literature was not conducted in the analysis due to divergent data sets and scope of focus. However, future work would benefit from direct comparison against a wider array of benchmarks. We also want to explore other means of capturing domain knowledge in the learning process, which we argue is likely to be essential for deriving effective and justifiable models, such as via the incorporation of semantic-search methods.

References

[1] McCarty LT. Reflections on TAXMAN: An experiment in Artificial Intelligence and legal reasoning. Harvard Law Review. 1976;90:837.

[2] Mumford J, Atkinson K, Bench-Capon T. Explaining Factor Ascription. In: Proceedings of JURIX 2021. IOS Press; 2021. p. 191-6.

[3] Brewka G, Ellmauthaler S, Strass H, Wallner J, Woltran P. Abstract Dialectical Frameworks revisited. In: Proceedings of the 23rd IJCAI. AAAI Press; 2013. p. 803-9.

[4] Gardner A. An Artificial Intelligence Approach to Legal Reasoning. MIT Press; 1987.

[5] Sergot M, Sadri F, Kowalski R, Kriwaczek F, Hammond P, Cory H. The British Nationality Act as a logic program. Communications of the ACM. 1986;29(5):370-86.

[6] Ashley KD. Modeling legal arguments: Reasoning with cases and hypotheticals. MIT press; 1990.

[7] Aleven V. Teaching case-based argumentation through a model and examples. Univ. of Pittsburgh; 1997.

[8] Bruninghaus S, Ashley K. Predicting outcomes of case based legal arguments. In: Proceedings of the 9th ICAIL; 2003. p. 233-42.

[9] Ashley KD, Brüninghaus S. Automatically classifying case texts and predicting outcomes. AI and Law. 2009;17(2):125-65.

[10] Al-Abdulkarim L, Atkinson K, Bench-Capon T. A methodology for designing systems to reason with legal cases using ADFs. AI and Law. 2016;24(1):1-49.

[11] Atkinson K, Collenette J, Bench-Capon T, Dzehtsiarou K. Practical tools from formal models: the ECHR as a case study. In: Proceedings of the 18th ICAIL; 2021. p. 170-4.

[12] Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V. Predicting judicial decisions of the ECHR: A natural language processing perspective. PeerJ Computer Science. 2016;2:e93.

[13] Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I. LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:201002559. 2020.

[14] Medvedeva M, Vols M, Wieling M. Using machine learning to predict decisions of the European Court of Human Rights. AI and Law. 2019:1-30.

[15] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.

[16] Dai X, Chalkidis I, Darkner S, Elliott D. Revisiting Transformer-based Models for Long Document Classification. arXiv preprint arXiv:220406683. 2022.

[17] Zhang X, Wei F, Zhou M. HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. arXiv preprint arXiv:190506566. 2019.

[18] Collenette J, Atkinson K, Bench-Capon T. An Explainable Approach to Deducing Outcomes in European Court of Human Rights Cases Using ADFs. In: Proc. of COMMA 2020; 2020. p. 21-32.

[19] Horty JF. Reasoning with Dimensions and Magnitudes. In: Proceedings of the 16th ICAIL; 2017. p. 109-18.

[20] Bench-Capon T, Atkinson K. Lessons from Implementing Factors with Magnitude. In: Proceedings of JURIX 2018; 2018. p. 11-20.

[21] Bench-Capon T, Atkinson K. Precedential constraint: The role of issues. In: Proceedings of the 18th ICAIL; 2021. p. 12-21.

[22] Prakken H, Sartor G. Modelling reasoning with precedents in a formal dialogue game. AI and Law. 1998;6(3-4):87-231.

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

[24] Lu J, Henchion M, Bacher I, Namee BM. A sentence-level hierarchical bert model for document classification with limited labelled data. In: Proceedings of DS 2021. Springer; 2021. p. 231-41.

[25] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:190711692. 2019.

[26] Branting LK, Pfeifer C, Brown B, Ferro L, Aberdeen J, Weiss B, et al. Scalable and explainable legal prediction. AI and Law. 2021;29(2):213-38.

[27] Prakken H, Ratsma R. A top-level model of case-based argumentation for explanation: Formalisation and experiments. Argument & Computation. 2021:1-36.

[28] Prakken H. A formal analysis of some factor-and precedent-based accounts of precedential constraint. AI and Law. 2021:1-27.

[29] Grabmair M. Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs. In: Proceedings of the 16th ICAIL; 2017. p. 89-98.

[30] Chalkidis I, Jana A, Hartung D, Bommarito M, Androutsopoulos I, Katz D, et al. LexGLUE: A Benchmark Dataset for Legal Language Understanding in English. arXiv preprint:211000976. 2021.

[31] Steging C, Renooij S, Verheij B. Discovering the Rationale of Decisions: Experiments on Aligning Learning and Reasoning. arXiv preprint arXiv:210506758. 2021.

[32] Ashley KD. Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge University Press; 2017.