JuLIA
Justice, Fundamental Rights and Artificial Intelligence

HANDBOOK
JuLIA (101046631)
DG Just –JUST-2021JTRA

ARTIFICIAL INTELLIGENCE, JUDICIAL DECISION-MAKING AND FUNDAMENTAL RIGHTS

SSM
Scuola Superiore della Magistratura
Università di Trento
upf. Universitat Pompeu Fabra Barcelona
Funded by the European Union

Published within the framework of project:
JuLIA (101046631) "Justice, fundamentaL rI
Co-funded by the European Commission (DG Just –JUST-2021-JTRA)

Coordinating Partner:
University of Pompeu Fabra

Partners:
University of Trento
Italian School for the Judiciary
Dutch National School of Magistracy
Consejo General del Poder Judicial
University of Paris Saclay (Versailles St Quintin)
University of Groningen
University of Coimbra
Polish Institute of Legal Studies of the Academy of Science
Bruno Kessler Foundation
LIBRe Foundation

First (draft) Edition: September 2023
Second Edition (published): October 2024

Scuola Superiore della Magistratura
ISBN 9791280600592

Scientific Coordinator of the JuLIA Project:

Mireia Artigot Golobardes

Coordinators of the team of legal experts on AI and Justice:

Gianluca Grasso - Paola Iamiceli

Project Manager:

Bruno Castro Coma

Co-editors and Co-authors of this Casebook:

Co-editors: *Mireia Artigot i Golobardes, Gianluca Grasso, Paola Iamiceli, Maria Rosaria Maugeri*

Co-authors: *Mireia Artigot Golobardes, Filiberto Brozzetti, Federica Casarosa, Mariavittoria Catanzariti, Maria Giuliana Civinini, Giuseppe Contissa, Antonio Corbo, Tommaso De Mari Casareto dal Verme, Filippo Donati, Paolo Ferragina, Alessia Fidelangeli, Benedetta Galgani, Federico Galli, Andrea Maria Garofalo, Marco Gioia, Gianluca Grasso, Paola Iamiceli, Giulia Lasagni, Marisaria Maugeri, Yannick Meneceur, Gaetana Natale, Nilcan Ozalp, Piera Santin, Francesco Perrone, Laura Piva, Silvio Ranise, Amedeo Santosuosso, Galileo Sartor, Etienne Vèrges, Géraldine Vial*

Assistant editors:

Beatrice Marone
Tommaso De Mari Casareto dal Verme

Note on national experts and contributors:

The JuLIA team would like to thank all the judges, experts, and collaborators who contributed to the project and to this Handbook by suggesting national and European case law.

Table of Contents

5.2. Analytics for Deciding Legal Cases: The ADELE Project

Federico Galli, Alessia Fidelangeli, Piera Santin, Galileo Sartor
–Università degli Studi di Bologna

Abstract: This contribution describes the ADELE research project, funded by the EU, that aims to use legal analytics to perform various activities in relation to different corpora of legal decisions belonging to different domains and written in different languages. The purpose is the implementation of tools and a methodological framework capable of being further transferred in other relevant fields.

1. The ADELE project

The ADELE Project ("Analytics for DEcision of LEgal cases") was co-funded by the European Union's Justice Programme (2014-2020). The project ran for 30 months from February 1, 2021 until July 31, 2023.

It was developed in cooperation with the Alma Mater University of Bologna, the Bulgarian legal informatics company APIS Europe, the University of Turin, the Centre for Judicial Cooperation of the European University Institute of Florence, the University of Luxembourg, the Bulgarian LIBRe Foundation, and the Bulgarian Union of Jurists.

The project developed Artificial intelligence (AI) and legal analytics (LA) methods to support legal research and decision-making processes in the judiciary. The project focused on two legal areas: value added tax (VAT) and trademarks and patents (T&P) and analysed both national and EU case law and legislation. It engaged in multiple tasks, such as the annotation of legal decisions, keyword and summary extraction, knowledge representation, network analysis, judicial argument extraction, and outcome prediction.

The project's main achievement was the development of an AI-driven tool that offered various functionalities and was embedded within an open platform available on the web.[159]

This contributions presents the different activities carried out during the project below.

2. Background

ADELE was built on a paradigm change in AI: techniques of knowledge representation and logical inference which are complemented by machine learning as applied to large datasets (Big Data). The application of ML techniques today enables the discovery of correlations, often yielding unexpected insights, and is giving rise to many new applications: conceptual retrieval, speech and image recognition, question-answering, translation, planning, etc. Such applications are transforming many aspects of life and delivering many benefits, i.e. in the fields of healthcare, commerce, and transportation.

[159] https://adele-tool.eu.

The way law is practiced is also being affected by AI as new techniques unfold for legal cognition and practice.[160] The emerging field of legal analytics (LA) for example develops applications in the legal domain to extract legal knowledge, infer undiscovered relations, and engage in data-driven predictions.[161]

So far, legal analytics applications have been developed in the following domains:

- the identification and representation of legal knowledge, defining types of annotations, concepts, links and their integration with conceptual ontologies;
- machine learning with legal texts, managing complex unstructured datasets, classifying paragraphs of texts, extracting legal rules, comparing rules across jurisdictions;
- the extraction of arguments from cases, including argument structure, claims and substantive legal factors;
- the connection of computational reasoning models and legal texts, annotating obligations in privacy policies, delivering Q&A services, making legal predictions in IPR cases, and providing visual maps linking cases and concepts.

Many applications have been developed as private-sector commercial initiatives for lawyers or legal organisations.[162]

However, in the past few years, we have seen an increasing interest in developing and using AI in the judiciary and the CEPEJ launched an initiative to collect data and information on current applications in Europe.[163] This initiative resulted in an open database where helpful information can be found, i.e. the year the system became functional, a short description including the underlying technology, a link to an official public source/reference of the system, and the status of the system, i.e. whether the system is currently functional or is still in the pilot phase. The database includes applications that are strictly dedicated not only to domestic courts, but also to lawyers and law firms. Different categories of users have been identified, such as court users (general public), court management, judges, lawyers and prosecutors.

Several useful applications are being developed in this field, for example:

- management tools in administration activities, from setting up calendars of meetings and hearings to the more sensitive issue of allocation of cases depending on the topic and relevance;
- information retrieval, for instance, in the recognition of patterns in text documents and files, as eDiscovery, or argument extraction from cases;
- a triage function, helping judges to quickly assess the relevance of previous cases to the present decision, such as court cases to verify litigation risk assessment, or simply gathering relevant information for claims, such as the solution explorer at the Civil Resolution Tribunal;
- citation retrieval and management;

[160] R. SUSSKIND, Tomorrow's Lawyers, Oxford, 2017.

[161] K. D. ASHLEY, *Artificial intelligence and legal analytics: new tools for law practice in the digital age*, Cambridge, 2017.

[162] Among the most famous applications, we can cite Casetext, LexisNexis (incl. Lex Machina, Ravel Law), Premonition, ROSS Intelligence in the US; Case Law Analytics, Doctrine, Predictice for France; CourtQuant, Justis, Solomonic in the UK.

[163] Resource Centre on Cyberjustice and AI, https://www.coe.int/en/web/cepej/resource-centre-on-cyberjustice-and-ai.

- document-assisted generation systems, where the software automatically generates text that can help judges write judicial documents;
- speech-to-text applications, allowing the judge to receive the transcript of the hearing or courtroom record;
- risk prediction systems, where the algorithm is able to forecast a possible outcome of the case or an aspect related to it, which include compensation and litigation fees.

Policy-makers worldwide are increasingly recognizing the potential of these solutions and have raised government funding for initiatives and projects aimed at integrating AI into the judiciary.

3. Objectives

Against this background, the objectives of the ADELE project were to design a methodological framework of legal analytics (LA) for court decisions and to implement it in a pilot tool designed for Italian and Bulgarian case law in the fields of Intellectual Property Rights (IPR) and Value Added Tax (VAT).

The first objective was accomplished by designing a LA methodology consisting of: (1) a data collection and annotation phase, whereby relevant knowledge was added as metadata or machine-readable information to a particular document and integrated with legal ontologies; (2) a data processing phase, whereby machine learning and NLP techniques were applied to annotated legal texts to extract relevant information, i.e. arguments, claims, citations, sections, keywords, and summaries.

The aforementioned LA framework allowed for the development of the following functionalities:

- ontology-based searches, which provide a search function through a comprehensive view of the domain-specific maps with a representation of the most relevant concepts and their connections;
- citation extraction and network analysis, which enables judges to collect useful insights from previous judgments that are conceptually or functionally similar and to get a complete overview of the operative construction of the law;
- summary and keywords extraction, enabling the visualization of legal information to quickly understand the main factual and legal issues discussed in the decision;
- argumentation extraction, which allows the extraction of argumentative patterns from a decision. This application enables judges to establish the rationale of previous judgments and offers a conceptual and argumentative toolkit to make the final decision in the case at hand;
- outcome prediction, that enables the anticipation of what could be the decision to a specific case according to past case law.

In the following paragraphs, the methodology and practical steps taken to create the above-mentioned functionalities are discussed, starting with data collection and annotation, and then specifying how the data was used for each task presented above.

4. Data collection and annotations

Several judicial decisions were selected to create the source corpora for processing activities. In particular, approximately 500 decisions were collected between Italy and Bulgaria, both in the field of VAT and Trademark and Patents.

As a general rule, judicial decisions containing personal data were collected anonymously. For instance, both Bulgarian corpora were prepared by selecting anonymized decisions from public data sources. When possible, Italian corpora were also annotated by selecting decisions that were already anonymized. In cases where judicial decisions included personal data, anonymisation techniques were employed, accounting for the risk of data breaches balanced against the purposes of processing. In particular, any identification data of natural persons involved in proceedings were immediately anonymised. The names of legal persons were also anonymized as a precautionary approach, especially when the company name referred to natural-person entrepreneurs. In the Trademarks and Patents dataset, trademarks and patents were not anonymized –even when they refer to natural persons who are entitled to the relevant protection –as their name is often crucial for acquiring information about the dispute, which can also be important for machine learning purposes.

During the collection processes, we encountered different problems related to data accessibility, especially regarding Italian case-law. In particular, issues pertained to: (1) the multiplicity of databases through which decisions are accessible and whether a specific database includes all decisions or just a number of them; (2) the limited access to subscription-based databases or those restricted only to judicial professionals; (3) the lack of search filters in the database, which hampers the granularity of the search and thus the homogeneity of the created corpus; (4) the non-digital-native nature of documents and the machine-unreadability of formats used.

After the collection of data, a preliminary analysis of case-law was carried out to understand the recurrent structure of the decision, the type of language used, the recurrent argumentative patterns, the citation styles and overall, the content of the decisions.

Based on the aforementioned analysis, a set of annotation guidelines was developed. These are a set of rules or instructions defining the criteria legal experts use to annotate the text. They define what should be included in each annotation and can also provide a structure for annotation, such as a template. In particular, the annotation guidelines were developed to annotate the structural elements of decisions and their relationships and to annotate judicial arguments found in the court's decisions. The annotations were used for the automated extraction of arguments and outcome prediction. Citations were not annotated in the text, as a more traditional and less time-consuming method was adopted through regex.

Based on annotation guidelines, a corpus annotation was carried out in an incremental fashion. Annotation is the process of adding metadata or additional machine-readable information to a particular data or document. The annotation was carried out by experts in the analysed legal fields and consisted of the insertion of additional information within the texts of the analysed judgments. A part of the documents was tagged through a double-blind method, where two annotators were asked to tag the same documents and required to agree on the use of tags. Inter-annotator agreements were also measured to ensure the soundness and coherence of the datasets, i.e. to avoid the same legal information being treated differently by different annotators.

The corpora were annotated using XML text editors, including Sublime Text, NotePad++ and Visual Studio Code. Different annotation software (such as Gloss and Inception) was also experimented with and some developing with SenTag was also begun. This provided an environment for annotating and editing case law using XML tags, with several customized functionalities such as a semi-automated guided annotation procedure, a visualization space for the argumentative structure of decisions, and an automated agreement procedure.[164]

The annotated case law was contained in the ADELE platform as searchable and browsable case-law in which annotations can be highlighted through information retrieval functions. In addition to that, decisions were used to train the machine learning models for outcome prediction and argument mining, as well as to extract case law citations.

5. Citation extraction and network analysis

Given the large availability of data, the automated analysis of legal texts has become increasingly relevant in recent times. Network analysis, in particular, is being employed to analyse intricate legal domains by representing legal documents (judgements, statutes, regulation) in terms of nodes, i.e. cases, legislative documents, and their corresponding relationships or edges, such as citations and normative referrals. A network is particularly useful in understanding the overarching structure of a domain, allowing for the identification of specific characteristics, including the case most frequently cited, clusters of similar cases, which contribute to predictive analysis.[165] While most experiments in this field have been applied to EU case law, there are still ongoing efforts to apply the same methodology to national courts.[166]

In the context of the ADELE Project, network analysis in the field of VAT and T&P cases was carried out in Italy and Bulgaria. The pipeline included the automated extraction of cases and then the development of networks.

The automated extraction was performed by using a regular expression (regex) algorithm, namely a computational procedure used for matching patterns in strings based on pre-defined language rules. To do that, the first step was to manually detect judicial citation patterns, i.e. how judges usually refer to and cite other documents within a case. It was noticed, for example, that most Italian cases did not follow a common style of citations, as the same document could be referred to in multiple ways. For instance, the court identifier could be written fully, shortened, or, in some cases, skipped completely. The same occurs for the date. Through trial and error, we attempted to build regular expressions as high-level as possible to accommodate the automated extraction of all possible citation styles. For the purpose of the project, only citations in specific parts of the original decision text were searched, to include only the citations that were relevant in decision making, in particular those made by the judge.

---

164 A. ZERBINATI ET AL., *SenTag 2.0: A Cooperative Annotation Tool*, in *Proceeding of the 21st International Semantic Web Conference*, CEUR Workshop Proceeding, 2022.

165 M. DERLÉN, J. LINDHOLM, *Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments*, in *European Law Journal*, 2014, 20.5, 667.

166 A. LOUIS, G. VAN DIJCK, G. SPANAKIS, *Finding the Law: Enhancing Statutory Article Retrieval via Graph Neural Networks*, January 2023, available at: arXiv:2301.1284; R. WINKELS, J. DE RUYTER, *Survival of the Fittest: Network Analysis of Dutch Supreme Court Cases,* in M. PALMIRANI ET AL. (eds.), *AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents*, Lecture Notes in Computer Science, Berlin, Heidelberg, Springer, 2012, 106.

Once extracted, the citations were converted into a pre-established format: [Court][Branch][Date][Case number][Text section]. This made it possible to maintain a standard structure, with different elements directly retrieved in the text of the parsed document. A similar structure was extracted for legislative citations, while keeping the different legal sources in consideration. Unfortunately, citations were extracted only at the document level. This was due to the fact that judges, at least in Italy and Bulgaria, rarely refer to specific paragraphs of cited decisions in their judgments. For this reason two classes of citations could be identified: complete well-formed citations and incomplete ambiguous citations. Only complete citations permitted searching for the document and extracting additional information on the case, such as additional cited documents. However, in the Italian dataset, cited cases were often not publicly available. Unlike in EU law datasets which had been worked on previously,[167] this grossly limited the possibility of extracting complex networks of citations.

To enhance the output of the system, existing solutions were implemented that could be tailored to our use. For the Italian dataset in particular, the Linkoln system was used, a piece of open-source software developed at IGSG-CNR (*Istituto di Informatica Giuridica e Sistemi Giudiziari del Consiglio Nazionale delle Ricerche*) for the automated detection and linking of legal references contained in legal texts written in Italian.[168] With the capabilities of the library, which was itself built on the same methodology used (common patterns, regular expressions), it was possible to identify most citations and fix those that were still incorrect by modifying its output. Another module was then developed to check online sources for the document, validate the metadata extracted, and add more if necessary.

At this point, the extracted citations were ready to be converted to a graph structure and analysed further. In particular, the structure was imported into a graph that could be visualised. The result, while not very deep, still showed recurring patterns and groups of cases with common citations.

[167] G. SARTOR ET AL., *Automated Extraction and Representation of Citation Network: A CJEU Case-Study*, in R. GUIZZARDI, B. NEUMAYR (eds.), *Advances in Conceptual Modeling*, Lecture Notes in Computer Science, Cham: Springer International Publishing, 2022, 102.

[168] The software can be found at: https://gitlab.com/IGSG/LINKOLN/linkoln/

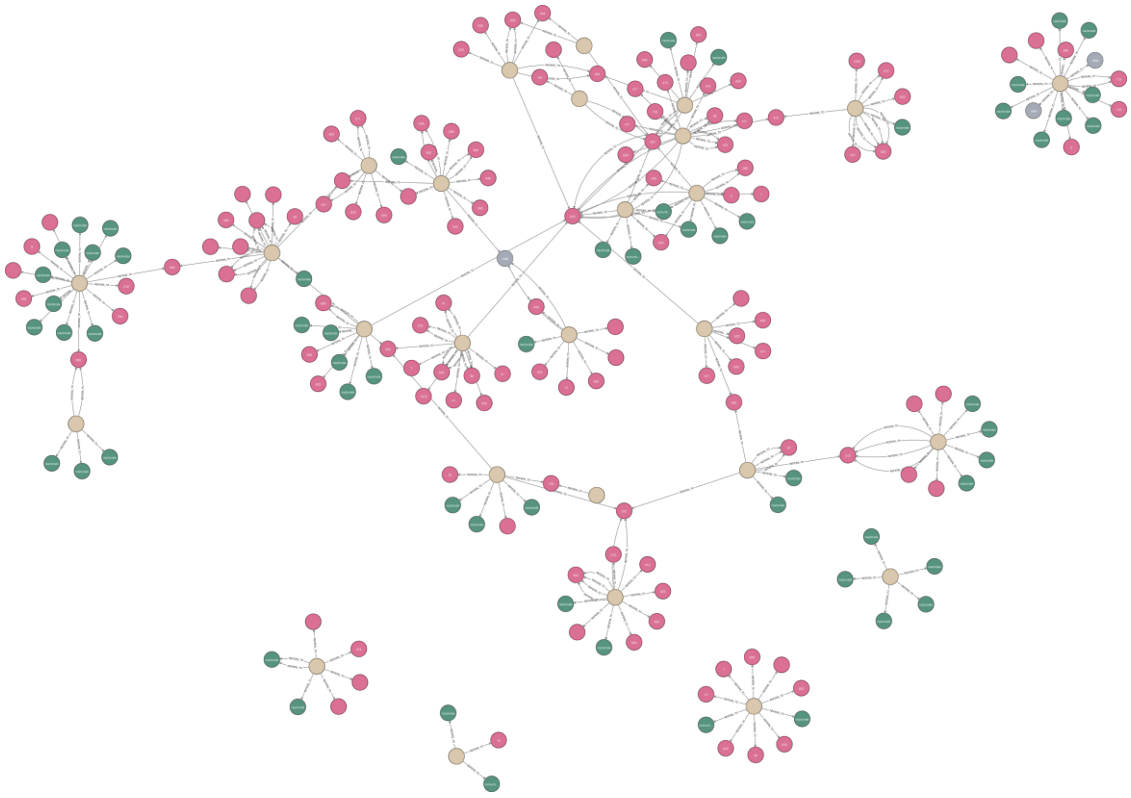

Figure 1. Section of the citation network

As we can see in Figure *1*, the ADELE cases, shown in light brown, are linked to the cited case law (pink) and legislative documents (green). We can clearly see that patterns, or clusters, can be identified and, even with incomplete data available, this is already a good result. The clusters have citations in common, and the citations that have many incoming edges, i.e. that are cited by many different cases, could be more relevant than others.

The graph can be further analysed with degree centrality metrics, measuring the number of incoming relationships for the citation nodes in the graph. With this information, a list of the more relevant (cited) cases was compiled and divided into four categories: national case law, European case law, national legislation, and European legislation.

This analysis could be enhanced further, with more information available, but as it is now, it can help identify the important cases or legislative documents, particularly when integrated with the ontology, that will be described in the following section, by linking the concepts in the ontology to the relevant legal documents and adding this information into the citation network.

6. Ontology framework

Ontologies are formal representations of structured and organised way of describing concepts and their relationships. They are used in AI, semantic web, natural language processing, and knowledge management systems to represent knowledge in a machine-readable and structured form. Such representations can then be used for information retrieval (enhancing the search and retrieval of legal

documents), facilitating the interoperability between different systems, to support automated reasoning and decision-making processes, and enabling the development of intelligent applications.[169]

In the ADELE project, two ontologies pertaining to the selected domain were built: (1) the OntoVAT, containing concepts relating to VAT taxable/exempt transactions 170 and (2) the PaTrOnto, integrating primary trademark and patent concepts, among which owner, validity, and transfer.[171]

The aim of the ontologies was twofold: (1) organizing and visualizing domain knowledge and (2) automatically linking the judgements in the dataset to the most relevant domain concepts.

The first of those aims was achieved by embedding a tool that provides the user with a comprehensive overview of a conceptual map of the respective legal areas in the ADELE platform. Each of the two ontologies offered a structure of terms with the respective legal definition provided in the legislation or case-law. Links between concepts are also built to allow the identification of synonyms and related terms between national and European legal concepts in the relevant field. To build the ontology, the most important concepts based on a twofold approach were selected: (i) top-down, starting from the pertinent legal sources, i.e. Directive 2006/ 112/CE for VAT, and (ii) bottom-up, based on the most recurrent concepts contained in the collected Italian and Bulgarian VAT decisions. Based on this analysis, a team comprised of lawyers and computer scientists provided a seminal representation of the domain's notions for each element of the examples, synonyms and related concepts, as well as the relevant European and national legislation in which the concepts were mentioned or defined and the most common examples instantiating that concept. For this reason, a multilingual OWL ontology enriched with a SKOS lexicalization was built and implemented in English, Italian, and Bulgarian. Finally, the gathered results were validated by the legal team that returned them to the technical team, who implemented the new information in the ontology. These last three steps were then iterated several times to refine the ontologies.

Different approaches were used, instead, to achieve the second aim, namely, to link judgements in the dataset to the most relevant domain concepts. The first layer was represented by the link between the ontological concepts and the citation automatically extracted from the datasets. The pertinent legislation associated with each ontological concept was then benchmarked with the citation extracted from cases. If the decision in the dataset cited one or more legal provisions contained in the ontology, it was connected to the related ontological concept. This means that if a decision cited a legal provision, and that legal provision contained a definition of a concept, then the case was marked as "rel a certain concept. This link was based on the assumption that if a decision cited one or more

[169] P. CASANOVAS ET AL., *Introduction: Theory and methodology in legal ontology engineering: Experiences and future directions*, Springer, 2011.

[170] D. LIGA, A. FIDELANGELI, R. MARKOVICH, *OntoVAT, an ontology for knowledge extraction in VAT-related judgments*, in *New Frontiers in Artificial Intelligence: JSAI-isAI* 2023 Workshops, AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers, Springer, 2024.

[171] D. LIGA, D. AMITRANO, R. MARKOVICH, *PaTrOnto, an ontology for patents and trademarks*, in *New Frontiers in Artificial Intelligence: JSAI-isAI 2023 Workshops,* AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers. Springer. 2024.

legal provisions containing the definition of a legal concept, then the decision was also relevant to that concept.

A more refined NLP pipeline is also in the process of being implemented which uses the two ontologies to determine whether an ontological concept is relevant for a specific judgement, i.e. if a specific decision deals with one or more of the ontological concepts. For this reason, mixed techniques were used, relying on: (1) supervised learning based on expert annotations and (2) unsupervised learning based on the ontology content itself. Regarding the first, legal experts were asked to select concepts from the ontology they considered more relevant to the cases collected in the dataset. Then, legal experts were asked to manually annotate nearly 70% of the judgements by including the information of whether each selected concept was relevant in each judgement by associ… was "1" … relevant" and "not relevant". The concept was considered relevant if the decision of the court concerned the concept from a substantial point of view. A designed algorithm then encoded the information contained in the ontology to predict whether or not a was relevant, comparing the results with the gold standard defined in the previous step.

Regarding the second step, the algorithm used synonyms and related concepts in the ontology to provide an expanded knowledge base to run NLP techniques which were able to link the ontological concept to case-law. For instance, the presence of definitions, examples, and related terms permitted ontological concepts like "supply of good…" to be linked to judgements in which the "transfer of…" was mentioned without some explicit reference to the term "sup…".

The ontologies built in the ADELE project can also be used for multiple purposes. For the organisation of judgements according to structured taxonomies that are more precise than those currently in use for example, which could help database searches for work and study purposes; or the creation of intelligent mind maps that allow for easy access to specific topics on a subject matter with conspicuous references to other areas of Law such as taxation; or for the extraction of keywords in judgements that are particularly homogeneous from a lexical point of view; or as a navigation and visualisation tool in intelligent databases.

7. Summarization and keyword extraction

Providing a set of relevant key terms (keywords) and summaries facilitates the retrieval of legal information by quickly assisting users to understand the main factual and legal issues discussed in a particular case without reading its text. In the context of ADELE, only extractive, not abstractive, summaries were worked with. Extractive summarisation selects the most meaningful sentences in the input text and combines them to form a summary. No change was made to the textual content of the extracted sentences.[172]

The automated key term extraction was created based on a trained spaCy model based on Named Entity Recognition (NER). Two specialised dictionaries were used for the training of the dataset, which contained legal terms in the relevant fields of Law – one in the field of VAT and one in the field of TM&P. Both were compiled by legal experts. In addition, a few more general law-related dictionaries, such as EuroVoc, were exploited. The applied model labels the key terms recognized in the text of a decision in two categories: the terms from the specialised dictionaries were labelled as priority key terms, whereas those contained in other

[172] Abstractive summarisation on the other hand generates new text which aims to provide a synoptic statement of the content of input documents, without reproducing their wording.

dictionaries were labelled as non-priority key terms. The terms that the trained model classified as priority key terms were directly included in the final selection, whereas the KeyBERT library[173] was used for the non-priority keywords to extract only those keywords that were most relevant to the processed document. However, before using KeyBERT, additional filtering was applied for non-priority key terms. For instance, only non-priority key terms that had at least two occurrences within the text of the processed document were included.

Similar to automatically extracted key terms, summaries were extracted following a supervised learning approach. The experiments with text summarisation began by using the TextRank model,[174] but the results were unsatisfactory. It was therefore decided to train a spaCy model[175] on a dataset of about 40,000 Bulgarian court decisions summarized manually by legal experts. The model was based on sentence text categorization, whereas the aforementioned manually annotated data were used to generate a training dataset containing two categories of sentences: summary and not_summary. The applied approach led to a significant improvement in results for the Bulgarian decisions. For Italian decisions, initially, the TextRank model was also attempted. Since the results were unsatisfactory as well, the trained "Bulgarian" spaCy model was also used, as it was based on a multilingual transformer. In this way, far better results were achieved.

In the ADELE tool, key terms and sentences are displayed under the title of judicial decisions, in the lists of documents corr in an opened document. Depending on the length of the decision, between 5 and 15 key terms are usually shown.

8. Argument extraction

The argument extraction module sought to identify arguments and their features. For this purpose, we deployed and refined state-of-the-art techniques in argument mining, i.e. the process of automatically detecting arguments from natural language texts to identify and analyse their structure and content.

Experiments were initially developed for and applied to CJEU decisions in the field of European Tax Law, specifically on fiscal State aids appeal decisions, and then were tested on national VAT and T&P datasets.[176]

Within CJEU decisions, sections devoted to the descri ruling were focused on in particular.[177] For the annotation process, sentences independent of one another were considered as tagging units. If a sentence contained more than one argument, usually separated by semicolons, each argument was considered and labelled separately. Two main elements were identified in particular: 1) the premise(s) and 2)

[173] M. GROOTENDORST, *KeyBERT: Minimal keyword extraction with BERT*, in *Zenodo*, 2020.
[174] https://cran.r-project.org/web/packages/textrank/vignettes/textrank.html.
[175] https://spacy.io/usage/models.
[176] Given the complexity of the task at hand, the decisions of the CJEU provided two advantages: (1) the standard (although not fixed) structure of CJEU decisions, in which complex and highly variable arguments are embedded; and (2) the English language, which prevented possible issues deriving from linguistic differences between Bulgarian and Italian judgments, possibly resulting in different argumentation structures (which may hinder a harmonised annotation process).
[177] These sections are usual.l"y denoted as "Findings

conclusion. When an argument consisted of multiple steps, the conclusion of each step was the premise for the next, so each intermediate step was labelled as a premise and only the last step as a conclusion.

In the annotation, some mandatory or optional properties were also attributed to each part of the arguments. For example, a distinction was made between legal and factual premises, the former on applicable law or relevant case-law and the latter related to statements of facts of the case. Legal premises were also characterised by argumentation schemes, which identified the stereotypical structure of an argument.[178] The following classes of argument were included: authoritative, verbal classification, interpretation, literal interpretation, precedent, principle, intention of the legislator, rule, systematic interpretation, and teleological argument.

An additional layer of complexity pertained to linking each premise to other premises or the conclusions through a logical connection. These were: (1) support, indicating a support relationship between a set of contiguous premises or between a premise and its conclusion; (2) attack, indicating either a rebuttal or an undercutting relation; (3) rephrasing, indicating that a particular premise was entirely rephrased by another premise, both having the same semantic meaning, even though they were differently formulated.

From a machine learning point of view five tasks were defined: (1) argument detection, i.e. classifying a sentence as premise, conclusion, or neither; (2) argument classification, i.e. classifying a sentence that is known to be argumentative as premise or conclusion; (3) type classification, a multi-label classification problem where a sentence that is known to be a premise is classified as legal and/or factual; (4) scheme classification, i.e. a multi-label classification task where a sentence, known to be a legal premise, is classified according to its argumentative scheme; and (5) link prediction, where two inputs, i.e. the source and the target that belong to the same document and are known to be argumentative, predict whether there is a link from the source to the target.

For tasks 1-4, we adopted three different representations of the input text: TF-IDF, Sentence-BERT (SBERT),[179] and Legal-BERT.[180] A set of traditional machine learning techniques that have a low computational footprint were chosen as classifiers, i.e. Linear SVC, Random Forest, and K Neighbors. For task 5, the neural architecture comprising stacks of dense and Long Short-Term Memory (LSTM) layers , a co-attention module, and residual connections were experimented with.[181] Moreover, 300-dimensional GloVe pre-trained embeddings were used for text representation. The model was also designed to encode the distance between the source and the target as a 10-bit array and a complete description of techniques and results was presented in two previously published papers.[182]

---

[178] D. WALTON, C. REED, F. MACAGNO, *Argumentation schemes*, Cambridge, 2008; D. WALTON, F. MACAGNO, G. SARTOR, *Statutory interpretation: pragmatics and argumentation*, Cambridge, 2021.

[179] N. REIMERS, I. GUREVYCH, Sentence-bert: Sentence embeddings using siamese bert-networks, 2019, available at: arXiv preprint arXiv:1908.10084.

[180] I. CHALKIDIS ET AL., *LEGAL-BERT: The muppets straight out of law school*, 2020, in *arXiv preprint* arXiv:2010.02559.

[181] A. GALASSI, M. LIPPI, P. TORRONI, *Multi-task attentive residual networks for argument mining*, in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[182] G. GRUNDLER ET AL., *Detecting Arguments in CJEU Decisions on Fiscal State Aid*, in *ArgMining 2022: 9th Workshop on Argument Mining*, 2022, 143; P. Santin et al., *Argumentation Structure Prediction in CJEU Decisions Fiscal State Aids*, in *Proceeding of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023)*, June 19–23, 2023, Braga, Portugal. ACM, 2023.

9. Outcome prediction

The number of studies where machine learning and NLP techniques have been applied to judicial decisions to predict the outcome of cases has greatly increased in the past several years. Most of these studies adopt judicial decisions –or features extracted therefrom – and their related outcomes as examples for training machine learning classifiers. Roughly three kinds of approaches have been identified:[183] (1) approaches based on features unrelated to the merits of the case;[184] (2) approaches based on legally relevant factors;[185] and (3) approaches based on the textual description of a case.[186]

In the ADELE project, the third approach was followed by applying machine learning and NLP techniques only to a portion of cases, i.e. the requests of parties (request, claims, and arguments). Based on a request, the task was to predict the likely outcome of a decision on that specific request. Questions were raised about whether such a task could actually be described as predictive since the system was trained to "predict" decisions that were already made. If correctly trained, the system should be able to classify decisions based on outcomes. This task was mainly useful for identifying predictors like facts, arguments, and judges of court decisions within the text of judgements, i.e. possible correlations between the requests of parties and the court's dec requests. However, whether such detected predictors could be used to anticipate future cases based on new requests was controversial and subject to discussion.[188] The aforementioned task was realised in ADELE using supervised learning, in which the system was provided with the relevant annotated text together with the corresponding outcomes.

The annotation guidelines reflected the structure of the decisions and their portions. The guidelines focused on the identification of the following elements in partiuclar: (i) the parties, (ii) the related requests, claims, and arguments; (iii) the Cour decisions. Such information could be of different lengths and details and was often enclosed within the same portion of text. For this reason, hierarchical levels of annotation were identified. For example, the section containing the parties' submissions <partreq> label) included (a) specific requests (denoted with a <req> label), i.e. the judicial measures sought by the party or the counter-party; (b) the claims (denoted with a <claim> label), i.e. the ultimate reasons for grounding a request, usually supported by premises and which may have concerned procedural or substantive facts; (c) arguments (denoted with a <arg> label), i.e. statements that supported or attacked a claim and which could be legal or factual. The annotation guidelines were developed with the aim of making them applicable

[183] F. GALLI, G. SARTOR, *AI Approaches to Predictive Justice: A Critical Assessment*, in *Humanities and Rights Global Network Journal* 5.2, 2023

[184] D. M. KATZ, M. J. BOMMARITO, J. BLACKMAN, *A general approach for predicting the behaviour of the Supreme Court of the United States*, in *PloS one*, 2017, 12.4, e0174698.

[185] K. D. ASHLEY, S. BRÜNINGHAUS, *Automatically classifying case texts and predicting outcomes*, in *Artificial Intelligence and Law*, 2009, 17.2, 125.

[186] N. ALETRAS ET AL., *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective*, in *PeerJ Computer Science*, 2016, 2, e93.

[187] M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction of court decisions*, in *Artificial Intelligence and Law*, 2022, 1.

[188] M. MEDVEDEVA, P. MCBRIDE, *Legal Judgment Prediction: If You Are Going to Do It, Do It Right*, in D. PREOTIUC-PIETRO ET AL. (eds.), *Proceedings of the Natural Legal Language Processing Workshop* 2023, Singapore: Association for Computational Linguistics, December 2023, 73.

not only to the specific domains of he projects, but possibly to all legal domains, with minimal adjustments. Moreover, they were designed to be applicable to different legal systems.

Afterwards, NLP and machine learning techniques were applied to the texts of decisions. The aim was to establish whether, based on the requests and arguments of the parties, the system was able to predict the outcome of individual requests and with what margin of error. Two representations of the input text were adopted: TF-IDF and Sentence-BERT (SBERT). As classifiers, Linear SVC, SVC, Random Forest, Gaussian Naïve Bayes, and K-Neighbours were chosen. The F1 score obtained for each class and their macro-average were also measured. The task of determining the decision outcome based only on the claims and arguments of the parties reached a maximum score of 0.68 with Linear SVC and SBERT. The introduction of the motivation and decision sections gave conflicting results, improving some classifiers but worsening others. A complete description of techniques and results was presented and discussed in a previously published paper.[189]

The results obtained were in line with those of other projects, but insufficient to imagine a real use in application hypotheses. Moreover, the experiments concerned judgments that had already been written and, thus, in order to assess any further application, a first assessment of the appropriateness of the use of the methodology adopted for the acts of the parties would be necessary.

10. Validation and feedback

Besides relying on common validation techniques for machine learning, testing and validation events with judges were planned. Two series of events were organised in particular to test the beta version of the platform and to validate the final version. The events were organised in parallel in Italy and in Bulgaria, with representative judges from the two selected domains. The two national validation events were organized to assess the technical and user-operative capacity of the final version of the tool. The Italian validation event was held on May 12, 2023, as a one-day event, at the European University Institute in Florence. The 15 participants were all Tax Law judges working in different jurisdictions and at different levels of proceedings. The Bulgarian validation event took place on May 17, 2023, in a hybrid mode. 50 participants took part in it, of which 16 were judges and judicial assistants. Lawyers and tax consultants also showed great interest in the event.

During the events, the features of the ADELE pilot tool were demonstrated using examples from the fields of VAT and TM&P Law. After that, a moot case was presented and participants were required to solve it by using the different platform functionalities. Finally, a discussion was held with participants in which they shared their thoughts and experiences regarding the use of AI systems in the practice of law. At the end of the event, feedback surveys on the platform modules[190] were collected by questionnaire. In total, 31 feedback questionnaires were collected.

The results showed a high level of satisfaction with the tool on average (see Figure) and an overall positive stance towards certain uses of AI by the judiciary (see Figure). Judges seem to have a highly positive estimation of AI-powered functionalities in particular that supports the easy and effective retrieval and summarisation of information in texts. This

[189] F. GALLI ET AL., *Predicting outcomes of Italian VAT decisions*, in *Jurix 2022, Proceeding of the 35th International Conference on Legal Knowledge and Information Systems*, 2022, 188.

[190] The "Database" module of the legal database, visualization of annotated judgements, automatically extracted keywords and summaries and extracted citation and analysis.

included automated summaries and key term extraction, citation analysis, ontology visualization and search, and the partially automated extraction of arguments. The largest doubts pertained to the outcome prediction.

Figure 2. Effectiveness of the ADELE modules

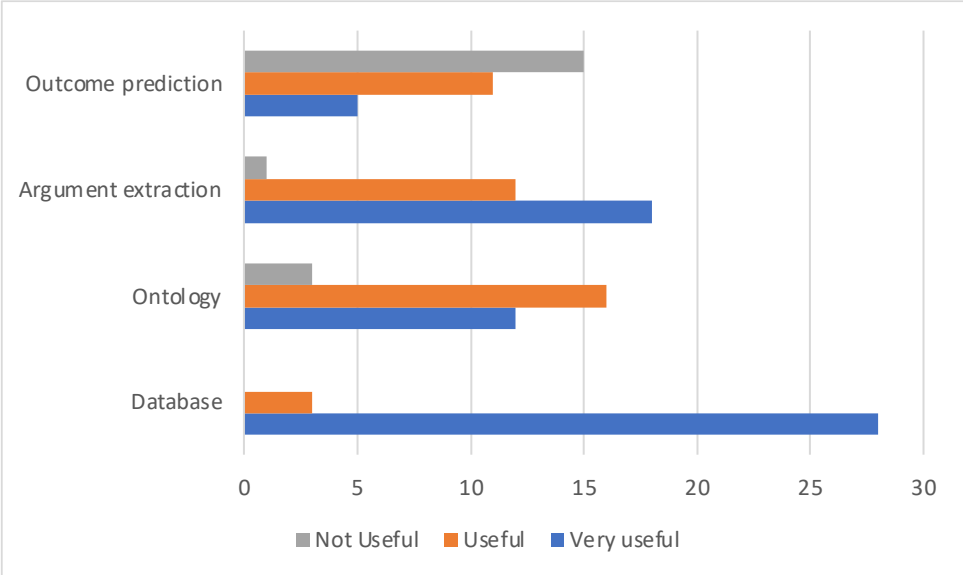

Figure 3. Usefulness of the ADELE modules

11. Ethics

During the project, an ethical self-assessment was carried out by the Consortium regarding the development of the AI-powered pilot tool. The tool was developed in compliance with the EU Ethical Guidelines on AI, adopted in 2019 by the European Commission through

the High-Level Expert Group on Trustworthy AI[191] and the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment, adopted in 2018 by the Council of Europe acting through the European Commission for the Efficiency of Justice.[192]

The tool was designed with human oversight, traceability, and auditability in mind, serving as a support instrument for legal professionals rather than a stand-alone decision-making tool. Its use was triggered by professionals for exploratory purposes, with outcomes intentionally initiated by users for spec drafted for the testing events, provided clear guidance on functionalities and machine learning model assumptions. Guidelines for annotating case-law were published and tested by legal professionals. Moreover, the tool prioritised compliance with fundamental rights, emphasising adherence to the European Convention on Human Rights (ECHR) and data protection laws. It incorporated privacy by design and default principles, by way of restricting processing to anonymized data. While designed to align with the ECHR, further investigation is still required to assess its compatibility w voluntary and exploratory nature, along with judicial responsibility for its use, mitigates concerns about fundamental rights. The independence of judges is highlighted, with measures taken to inform judges about the to use it or not to their judgment.

Furthermore, the tool adhered to fairness principles in AI design and usage, ensuring unbiased input data and algorithm design. Legal domains chosen for piloting (VAT and TM&P) also avoided sensitive personal data. Training data was properly anonymized, and rigorous testing revealed no discrimination tendencies. Efforts were also made to inform judges and legal professionals about the model limitations and training specifics.

Finally, users of the tool were informed of its use of AI, with clear explanations of its abilities, limitations, risks and benefits. The tool, designed for scientific debate in the legal use of AI, emphasised purposeful limitations. While th was not fully explainable, users received information about its AI algorithms, training data, and the non-binding nature of its predictions. The disclaimer specified that neither the ADELE consortium partners, nor the European Commission were responsible for provision.

[191] EUROPEAN COMMISSION HIGH-LEVEL EXPERT GROUP, *Ethics guidelines for trustworthy AI*, April 2019, available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[192] CEPEJ, *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, 2018.