Technical Specifications

Grant Agreement No.:	101087342
Project Acronym:	POLINE
Project Title:	Principles of Law in National and European VAT
Website:	https://site.unibo.it/poline/en
Contributing WP:	WP4
Deliverable ID:	D4.1
Contractual delivery date:	31/07/2024
Actual delivery data:	25/07/2024
Dissemination level:	PU
Deliverable leader:	APIS

This project is funded by the European Union's Justice Programme (2022)

1. Document History

Version	Date	Author	Partner	Description
1.0	13/7/2024	Hristo Konstantinov, Vasil Oreshenski	APIS	Definition of the document structure and first draft of Section 8.2
2.0	14/7/2024	Federico Galli	UNIBO	Contribution of Section 8.3
3.0	16/7/2024	Federico Galli, Giulia Grundler	UNIBO	Contribution of Section 10.3.3
4.0	19/7/2024	Rachele Mignone, Ivan Spada	UNITO	Contribution of Sections 8.4, 8.5 and 8.6
5.0	23/7/2024	Hristo Konstantinov, Dilyana Bozhanova, Lilia Kachoreva, Vasil Oreshenski	APIS	First draft
6.0	24/7/2024	Federico Galli	UNIBO	Review
7.0	25/7/2024	Hristo Konstantinov	APIS	Final version

2. Contributors

Partner	Name	Role	Contribution
APIS	Hristo Konstantinov, Dilyana Bozhanova, Lilia Kachoreva, Vasil Oreshenski	WP Coordinator	Coordination of the Deliverable activities
UNIBO	Federico Galli, Giulia Grundler	WP Participant	Contribution to the Deliverable activities
UNITO	Rachele Mignone, Ivan Spada	WP Participant	Contribution to the Deliverable activities

Disclaimer: On behalf of APIS EUROPE, the present deliverable has been drafted by Hristo Konstantinov, Vasil Oreshenski, Dilyana Bozhanova and Lilia Kachoreva.

3. Table of Contents

1. Document History 2
2. Contributors 3
3. Table of Contents 5
4. List of Figures 7
5. List of Acronyms 8
6. Executive Summary 9
7. Introduction 10
8. Back-end Infrastructure of the Pilot Tool 10
8.1. Architecture of POLINE Back-end 10
8.1.1. Overview 10
8.1.2. Data Processing Workflow 12
8.2. Annotation Tool and Corpus 13
8.3. NLP Tools for JPOL Extraction and Analysis 15
8.4. NLP Tools for Identification of Legal References 17
8.5. Methods and Tools for Ontology Development 18
8.6. Tools for Network Analysis 19
9. Study of AI-powered Legal Research Tools 20
9.1. Overview 20
9.2. Examples of AI-driven functionalities 23
9.2.1. Pre-trained legal chatbot / virtual assistant 23
9.2.2. AI-generated summaries and extraction of key information 25
9.2.3. Detection of gaps and risks. Suggestions for revisions. Outcome prediction 27
9.2.4. Automatic identification of similar reasoning. Citation analysis 28
9.3. Comparison and conclusions 30
10. Content and Functionality of the Pilot Tool 31

10.1. Description of the Legal Content 31
10.1.1. National Legislation 31
10.1.2. National Case Law 32
10.1.3. EU Legislation 33
10.1.4. Case law of the Court of Justice of the EU 33
10.2. Non-functional Requirements 34
10.3. End-user Functionality and Features 35
10.3.1. Legal Database Module 35
10.3.2. Link Visualisation Module 39
10.3.3. Customised Detection Module 39

4. List of Figures

Figure 8-1. *Architecture of POLINE Pilot Tool* 11
Figure 8-2. View of the annotation tool Gloss with annotated JPOLs in the text of a Bulgarian court decision 14
Figure 9-1. *Chat with CoCounsel - AI Assistant, developed by Casetext (source: https://casetext.com)* 23
Figure 9-2. *AI Legal Assistant, developed by Predictice (source: https://predictice.com/assistant)* 24
Figure 9-3. *Paxton AI Legal Assistant (source: https://help.paxton.ai/help)* 25
Figure 9-4. *AI-generated summary of plaintiff's complaint in an action brought in the Northen District Court of Illinois (source: https://lexmachina.com/legal-analytics/)* 26
Figure 9-5. *"Knowledge Search" functionality of the DeepJudge system (source: https://www.deepjudge.ai/product)* 27
Figure 9-6. *"Contract review" functionality of the Casetext system (source: https://casetext.com/)* 28
Figure 9-7. *Court decisions on the same legal ground found in the Doctrine system (source: https://www.doctrine.fr/fonctionnalites)* 29
Figure 9-8. *The Bloomberg Law Citation Map (source: https://www.deweybstrategic.com/2017/10/6072.html)* 30
Figure 10-1. *Example of a JPOL in a judgment of the CJEU* 38
Figure 10-2. *Example of a JPOL in a judgment of the Bulgarian Supreme Administrative Court* 38

5. List of Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
APIS	Apis Europa, beneficiary in POLINE project
BERT	Bidirectional Encoder Representations from Transformers
CJEU	Court of Justice of the European Union
ESG	Environmental, Social, and Governance
EU	European Union
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformer
HTML	Hypertext Markup Language
HTTPS	Hypertext Transfer Protocol Secure
JPOL(s)	Judicial Principle(s) of Law
JSON	JavaScript Object Notation
LLM(s)	Large Language Model(s)
ML	Machine Learning
NGO	Non-Governmental Organisation
NLP	Natural Language Processing
ORU	Örebro University, beneficiary in POLINE project
SFS	Svensk Författningssamling
SG	State Gazette
TCP/IP	Transmission Control Protocol / Internet Protocol
UNIBO	University of Bologna, coordinator of POLINE project
UNITO	University of Turin, beneficiary in POLINE project
VAT	Value Added Tax
WP	Work package

6. Executive Summary

This deliverable specifies the architecture, content and functionality of POLINE Pilot Tool. Section 8 defines the overall architecture of the Pilot Tool and, in particular, its back-end infrastructure, which includes various ML/NLP methods and tools for identification, analysis and processing of judicial principles of law and other legal data. Section 9 details the results of our analysis of state-of-the-art AI-powered legal research tools. Finally, Section 10 provides a description of the legal content and the end-user functionality of POLINE Pilot Tool and its three modules.

7. Introduction

POLINE project aims at developing an AI-powered pilot tool for the retrieval and analysis of judicial principles of law (JPOLs) in the case law of the CJEU and the supreme courts of Italy, Sweden and Bulgaria in VAT domain. To this end, the project partners apply advanced AI, ML and NLP methods and techniques for extracting, clustering and linking judicial principles of law. The results from the research and development project activities will be embedded in a modular online platform consisting of three modules:

- The *Legal Database Module* will provide access to a multilingual database and a search engine for the retrieval of JPOLs extracted from the CJEU and national case law in the VAT field.
- The *Link Visualisation Module* will allow a user-friendly visualisation of 1) similar principles of law both within the same legal system and between different legal systems, and 2) networks of decisions relevant to the judicial principle concerned based on the identification of explicit references or textual and semantic similarity.
- The *Customised Detection Module* will support legal practitioners in identifying and assessing JPOLs in a specific VAT case by allowing users to upload a decision or a tax audit act and receive a customised legal assessment on the relevant JPOLs in the document at hand.

The purpose of this document is to design a comprehensive technical documentation that specifies, on the one hand, the components of the back-end architecture of POLINE Pilot Tool and, on the other hand, the content and end-user functionality of tool and its three modules.

8. Back-end Infrastructure of the Pilot Tool

8.1. Architecture of POLINE Back-end

8.1.1. Overview

A classic three-layer architecture will be applied for the implementation of the POLINE Pilot Tool to achieve maximum separation of concerns, thus improving the maintainability, scalability and reliability of the system. The Presentation Layer aims to provide the end user with a convenient interface to benefit from the tool's functionalities through a web application. The Business Layer contains the core logic of the individual modules and the various ML/NLP tools used to implement all these functionalities. The Data Layer is designed

to retrieve, store and manage the data processed by the Pilot Tool at different stages of development and interaction with users (see Figure 8-1).

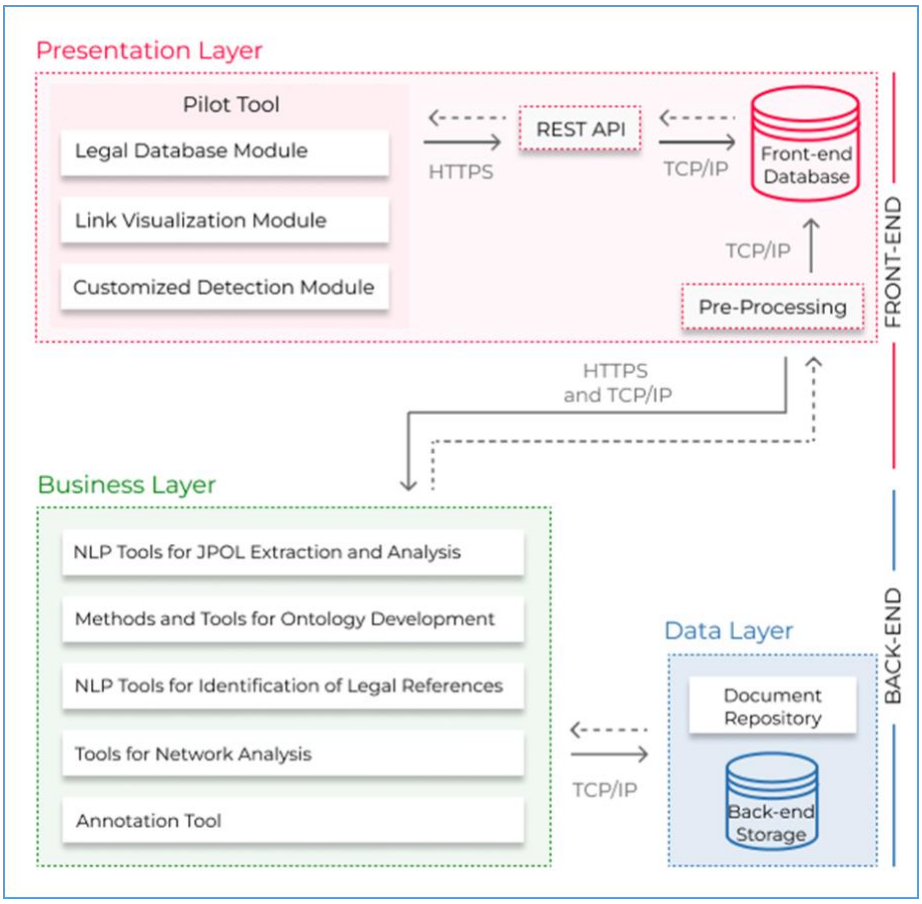


Figure 8-1. *Architecture of POLINE Pilot Tool*

A central document repository will be used to store the original texts of the court decisions in plain text format. Each team will have access to the repository. Furthermore, at various stages of development, each team will be responsible for maintaining sources and defining access and data format for the document to be processed. The architectural decision to use a central document repository in combination with multiple data sources after each processing stage provides teams with complete freedom to organise their internal workflows without affecting the operations of other teams.

The primary method of communication between the layers and system components relies entirely on HTTPS and TCP/IP protocols. Additionally, at the layer level, each module/tool will be constructed as a web API or console application facilitating easy integration and reuse among different project teams. Detailed description of the ML and NLP tools to be developed in POLINE back-end is provided

in the following subsections, while specifications of the three modules of the front-end application are included in Section 10.3.

8.1.2. Data Processing Workflow

POLINE partners will implement a multi-phase data processing workflow. This workflow begins with the collection and annotation of the raw texts of the court decisions and proceeds with the application of ML and NLP methods and tools for JPOL extraction, identification and analysis of legal citations and ontology development. The final stage of this process involves content and knowledge representation within the Pilot Tool. The data processing operations will follow distinct, non-overlapping paths, which we outline below briefly.

Collection and annotation of court decisions

Each partner will be responsible for selecting relevant case law. The court decisions will be stored as text documents in the central repository. Each document will be named with a unique identifier, used throughout various development stages. The annotation tool Gloss, maintained by the coordinator UNIBO, will be used for manual annotation of JPOLs in four datasets: a multilingual dataset of CJEU case law and three datasets of case law of the Italian, Bulgarian and Swedish supreme courts.

Applying machine learning and NLP techniques

The annotated datasets created by the partners' legal experts will be used to train a machine learning system for the automated identification, clustering, and linking of judicial principles of law. ML and NLP techniques will be employed to automatically recognise:

1) JPOLs, both in the European and in the three national case law datasets, and
2) similarities and divergences between principles of law.

Finally, NLP and network analysis techniques will be used to detect explicit and implicit references in the texts of court decisions and mapping them with concepts of the VAT ontology, which will allow providing visual representations to end-users of maps of decisions containing the same/similar judicial principles or JPOLs and decisions based on the same legal concepts or citations to EU and/or national legislation and case law.

Front-end pre-processing and visual representation in the Pilot Tool

The developed ML/NLP tools will automatically process the documents from the repository. Each tool will generate a JSON file containing the processing results to be used in the front-end database. This unified database will contain the output of all processing tasks

performed by the different teams. The obtained data will be further pre-processed and formatted in a way that best matches the functionality that the Pilot Tool will provide to the end users.

The final service will be developed as a web application providing:

1) a search engine for the retrieval of JPOLs with an interface for simultaneous visualisation of EU and national JPOLs (Legal Database Module)
2) means for a user-friendly representation of networks of decisions and JPOLs (Link Visualisation Module)
3) interactive interface for automated identification and assessment of JPOLs in a specific legal text provided by the user, e.g. court decision or tax audit act (Customised Detection Module).

8.2. Annotation Tool and Corpus

The annotation task, part of WP2 activities, aims at developing four datasets with manual and automated annotations of judicial principles of law: a multilingual dataset of CJEU case law and three datasets of Italian, Bulgarian, and Swedish case law[1]. The work is based on tagging guidelines elaborated by the coordinator UNIBO after extensive theoretical and empirical legal analysis of the judicial principles of law in the national and CJEU's case law.

Manual annotation is done by legal experts at UNIBO, APIS, and ORU using the Gloss annotation tool. Gloss is an online tool that takes as input text files and gives as output JSON files. The annotation consists of (1) highlighting the relevant portion of a text, (2) labelling it as a judicial principle of law, and (3) selecting its attributes from a list. Thus, annotators are not required to handwrite any part of the tag and its attributes, which prevents a lot of mistakes (see Figure 8-2).

[1] Each of these data sets will contain decisions on VAT cases of the highest judicial authority in the respective country: Corte Suprema di Cassazione in Italy, Върховен административен съд in Bulgaria and Högsta förvaltningsdomstolen in Sweden.

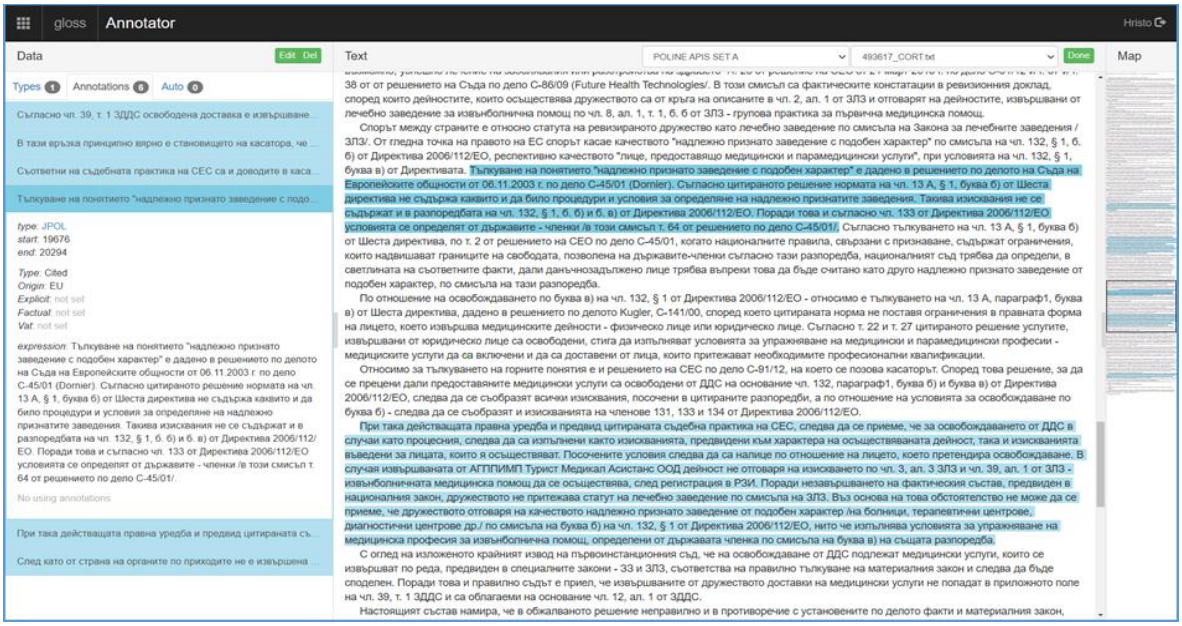


Figure 8-2. *View of the annotation tool Gloss with annotated JPOLs in the text of a Bulgarian court decision*

The following attributes are used to classify JPOLs:

- *New / Cited* – a JPOL is *New* if the Court provides it for the first time, and a JPOL is *Cited* when it contains a reference to a previous judgement in which that JPOL was already mentioned;
- *National / European / Both* – a JPOL is *National* when it refers to a previous national judgement, a *European* one, when it refers to a previous judgement of the CJEU, and *Both*, when the citing JPOL refers to both previous EU and national decisions;
- *VAT / Non-VAT* – a JPOL is *VAT* when its statement refers to legal issues in VAT domain, and a JPOL is *Non-VAT* when it refers to other topics;
- *Factual / Non-Factual* – *Factual* JPOLs contain an interpretation of the law as applied to the facts, while *Non-factual* JPOLs provide the interpretation of the law/principle in abstract terms, without references to the facts;
- *Explicit / Non-Explicit* – a JPOL is *Explicit* when it is identified as such or with a similar term by the court itself (this applies mainly to the Italian Court of Cassation, which must express the “principio di diritto” according to Italian law); otherwise it is a *Non-Explicit* JPOL.

The datasets of manually annotated judicial principles of law will be used to train a machine learning system for the automatic identification and annotation of judicial principles of law. The resulting merged datasets of manual and automated annotations of judicial principles of law will be used to train further the elaborated ML model for the automatic identification, clustering and linking of judicial principles of law.

The first release of the **annotation corpus** of four datasets and the **annotation guidelines** will be delivered at the end of the first project year, while the final corpus and guidelines – in project month 22.

8.3. NLP Tools for JPOL Extraction and Analysis

The extraction and analysis of Judicial Principles of Law (JPOL) from legal decisions involve advanced Natural Language Processing (NLP) techniques and machine learning algorithms. These methods facilitate the identification, annotation, and categorisation of legal principles, ensuring a robust and automated approach to legal text analysis. The integration of NLP tools and machine learning models significantly enhances the accuracy and efficiency of this process.

Supervised learning methods will be pivotal in extracting JPOL from legal texts. These methods involve training a machine learning model on a labelled dataset where the input features (text) are associated with known outputs (judicial principles). Common supervised learning algorithms used in this domain include:

1. **Support Vector Machines (SVM)**: Effective for high-dimensional spaces and widely used for text classification tasks.
2. **Random Forests**: An ensemble learning method that enhances classification accuracy by combining multiple decision trees.
3. **Neural Networks**: Especially effective for complex patterns in text data, with architectures like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) being particularly useful.
4. **Fine-tuning of BERT models**: a set of deep learning language models with the ability to consider context by analysing the relationships between words in a sentence bidirectionally.
    - **LegalBERT**: A variant of the BERT model fine-tuned on legal texts, which improves the understanding and generation of legal language across various jurisdictions.

These algorithms require a significant amount of labelled data for training. For this reason, we will also explore the possibility of generative Large Language Models (LLMs) to support annotation activities. These models can process large volumes of legal texts to identify and extract judicial principles, significantly reducing the manual effort required. Additionally, they can assist annotators by suggesting potential annotations and verifying the consistency and accuracy of the manual annotations. Examples of these architectures are:

1. **GPT (Generative Pre-trained Transformer)**: Developed by OpenAI, GPT models are highly proficient in generating human-like text and understanding complex language structures.
2. **Gemini**: Google's multi-model LLM, known for its capabilities in natural language understanding and generation.
3. **Llama**: A robust LLM designed by Meta AI for diverse NLP tasks, including text classification and language generation. This is made available to the research community under a non-commercial license.

The goal is to implement a **multi-stage pipeline**. The initial stage will focus on designing a prompt for the LLMs to build a semi-automatically tagged training set, based on the manually annotated documents. Then, the second stage will involve the implementation of supervised learning techniques, using the annotated datasets to teach the models to accurately identify and extract the legal principles and their attributes.

1. **Prompt design and training set annotation**: A first prompt, P1, is developed partially based on the annotation guidelines and used to annotate a novel set of documents. The experts perform a qualitative analysis of the result and iteratively refine the prompt.

   Once the experts believe they have reached a satisfactory prompt P2, it is used to annotate the validation set. Through the comparison between the human labels and the LLM labels, the experts can evaluate the current prompt and possibly further refine it. This step can be performed multiple times until a satisfactory prompt P3 is obtained. It is fundamental never to involve the test set documents in this process.

   Once a satisfactory prompt is obtained, it is used to annotate all the documents that will be part of the training set. Human experts then revise the annotations (alone or in groups) through sampling, pattern matching and so on. The final result constitutes the annotated training set. Since the prompt has been designed with complete independence with respect to the test set, it is possible to fairly evaluate the LLM

using the test set, and also compare it to other NLP methods.

2. **Models' implementation**: The focus will be on employing supervised learning methods and fine-tuning pre-trained transformer-based models like BERT and LegalBERT, adapting them to the specific task of JPOL extraction, allowing them to understand the specific nuances of legal language and VAT-related concepts. These models will be validated based on the manually annotated test set.

Depending on the composition of the data, we will consider whether to employ data augmentation techniques and/or data cartography techniques with curriculum learning, to expand the dataset and balance the classes, thus improving the model robustness.

- **Data augmentation** refers to techniques used to increase the size and diversity of a training dataset by applying various transformations to existing data, such as paraphrasing, synonym replacement, random insertion or deletion of words in text data, and so on. To this end, we will leverage methods developed in previous research projects and state-of-the-art LLMs.
- **Data Cartography** is a method for understanding and visualising the training data by mapping instances based on specific characteristics, such as confidence scores and variance. This mapping helps identify which instances are easy or hard for the model to learn and can be used to train the model with the Curriculum Learning technique. This technique involves training the model on simpler examples first and gradually increasing the difficulty level. This approach mimics the human learning process and helps stabilise the training process, possibly leading to better model performance.

Another important aspect to consider is that the models will be able to process all the POLINE languages: Italian, Bulgarian, Swedish and English, thanks to multilingual NLP.

The final output of the models will be in JSON format.

8.4. NLP Tools for Identification of Legal References

While previous approaches for Italian legal documents have successfully utilised regex-based tools like Linkoln[2], these methods are not directly transferable to a multilingual context. Regex-based tools rely on pre-defined patterns that are highly language-specific. Developing and maintaining separate regex patterns for each language would be overly complex and resource-intensive.

[2] Linkoln, accessed 15.07.2024. URL: http://ittig.github.io/Linkoln/

Therefore, this task proposes a novel approach leveraging the capabilities of Large Language Models (LLMs) for multilingual legal citation extraction. LLMs, trained on massive datasets of text and code, have demonstrated a remarkable ability to process natural language, including complex linguistic structures and nuances across different languages. This makes them particularly well-suited for the task of identifying legal citations within a multilingual corpus.

This task will involve evaluating the performance of various open-source LLMs specifically, their ability to extract legal citations. This evaluation will be conducted through two primary methods:

Manual Annotation: A team of legal experts will meticulously annotate a representative set of multilingual legal documents, creating a "gold standard" dataset. This dataset will serve as the benchmark against which the accuracy and precision of the LLM-extracted citations will be measured.

Comparison with Regex-Based Model: For Italian documents, the output of the LLMs will be directly compared against the results generated by existing regex-based models (e.g., Linkoln). This comparative analysis will provide concrete evidence of the potential advancements offered by LLMs in terms of accuracy, efficiency, and adaptability to multilingual contexts.

To maximise the effectiveness of LLMs for this task, careful prompt engineering will be crucial. This will involve crafting specific instructions and providing relevant examples to guide the LLM towards identifying and extracting legal citations accurately. For instance, prompts might include instructions like "Identify and extract all legal citations from the following text, including case names, court identifiers, and year of decision," along with examples of correctly formatted citations in each target language. Different prompt engineering strategies will be explored and evaluated to determine the optimal approach for maximising extraction accuracy and consistency across languages.

This dual evaluation approach, combining expert annotation with a direct comparison against established methods, will provide a comprehensive assessment of the viability and efficacy of employing open-source LLMs for the task of multilingual legal citation extraction.

8.5. Methods and Tools for Ontology Development

This section describes the development of a comprehensive legal ontology, focusing on Principles of Law (POL) within the European legal framework, particularly VAT directives.

This ontology will be built through a synergistic approach, leveraging the efficiency of automated information extraction with the precision and depth of expert legal knowledge.

The initial stage involves automatically populating the ontology using NLP and web scraping techniques on legal documents from EUR-Lex. This process begins with keyword extraction, addressing the complexities of varied formats and separators. NLP will be further utilised to identify and link crucial legal entities such as specific articles from Directive 77/388/EEC and Directive 2006/112/EC, and legal concepts to their corresponding entries within the ontology.

The system will go beyond mere entity recognition to extract the relationships between these elements. This includes identifying which articles belong to which directives, their relationship to specific legal concepts, and how different legal concepts are interconnected. To ensure consistency and interoperability with existing legal knowledge resources, the extracted information will be meticulously aligned with the EUR-Lex taxonomy.

While this automated process lays a strong foundation, expert legal knowledge is essential for refining and enriching the ontology. Legal experts will play a vital role in defining and validating the identified legal concepts, ensuring accuracy and clarity within the ontology. They will also validate the automatically extracted relationships, correcting any inconsistencies. A key contribution of the legal experts will be the integration of JPOLs derived from legal precedents and case law analysis. This addition will provide a practical, applied layer to the framework established through automated extraction. Finally, experts will refine the hierarchical structure of the ontology, defining parent-child relationships between legal concepts to ensure logical consistency and enable automated reasoning for complex legal queries.

8.6. Tools for Network Analysis

This section details the application of network analysis techniques to an ontology of legal concepts, specifically focusing on JPOLs. By representing JPOLs within an ontology, we can use network analysis to uncover hidden relationships and similarities between these legal principles.

Our analysis will utilise NetworkX, a Python library specifically designed for studying complex networks. NetworkX provides a wide array of algorithms and metrics crucial for our investigation, including those outlined below.

Identifying Important Principles: Centrality measures help us identify which JPOL hold significant positions within the network. Degree centrality highlights frequently connected

principles, while betweenness centrality reveals principles acting as bridges between different areas of law. Closeness centrality points to principles easily connected to all others, and eigenvector centrality identifies influential principles connected to other influential principles.

Highlighting Similarities: To determine the similarity between different JPOL, we will employ various similarity measures. Jaccard similarity compares the shared connections between principles, while cosine similarity analyses the attributes of each principle to determine their resemblance. Pearson correlation coefficient will be used to measure the linear correlation between different principles attributes.

Detecting Clusters of Principles: To identify groups of related JPOL, we will utilise community detection methods. Modularity will quantify the strength of the network's division into communities, and the Louvain method, an efficient algorithm for large networks, will be employed to detect these communities.

Additional Insights: Beyond the core analysis, we will leverage additional measures for a more comprehensive understanding. The shortest path length will reveal the proximity of different legal principles within the network. PageRank, originally developed by Google, will help identify the most influential principles based on their connectivity. Finally, structural equivalence will highlight principles with similar roles and importance by analysing their shared connections.

By applying this suite of network analysis tools, we aim to expose the intricate relationships and structures within the legal domain. This will not only deepen our understanding of how different Principles of Law interact but also pave the way for more intelligent legal information retrieval systems and more efficient legal research practices.

9. Study of AI-powered Legal Research Tools

9.1. Overview

In the last 3-4 years, the situation on the market of legal technology applications has been quite dynamic. New start-ups offering AI applications for legal research emerged,[3] some

[3] Europe Grows Its Own AI Legal Tech (accessed on 15.07.2024); Eye On AI: Legal Tech Funding Getting AI Upgrade (accessed on 16.07.2024).

significant mergers and acquisitions have occurred in the sector,[4] while at the same time some of the legal tech pioneers have closed their businesses.[5]

An important action that will certainly affect all AI systems specifically placed on the EU market is the final adoption in 2024 by the European Parliament and the Council of the AI Act.[6] The Regulation enters into force on 1.08.2024 and will be directly applicable in EU Member States in stages over a two-year period. On the one hand, the AI Act aims to promote the beneficial use of AI and give the EU a competitive advantage in the development and provision of AI systems. On the other hand, it sets high requirements and standards that AI systems put into service must meet, completely restricting some use cases of artificial intelligence that could endanger citizens' rights, for example, biometric categorisation systems serving social control practices. The Regulation established a common supervisory authority at the European level, the European Artificial Intelligence Board.

In the meantime, businesses have increasingly started to implement AI-based technologies such as natural language processing and large language models. Generative AI has become the most in-demand technology since, in November 2022, OpenAI launched a free preview of ChatGPT, its AI chatbot based on the large language model GPT-3.5.[7] The release of ChatGPT and other similar AI Copilots sparked the interest of the general public thanks to the ability of AI to communicate with humans and to do so in more and more languages.[8] Many companies around the world soon recognised AI technologies as an opportunity to improve their products and services and began using generic LLMs as the basis for training their own customised domain-specific models.[9]

Legal information providers have not fallen behind this trend and many have deployed AI functionalities such as a virtual assistant capable of conversing in natural language, generating summaries of legal documents, identifying risks and predicting the outcome of

[4] Thomson Reuters Completes Acquisition of Casetext, Inc. (accessed on 9.07.2024).
[5] Judicial Analytics Provider Gavelytics to Close Company on June 30 (accessed on 9.07.2024); Litigation Prediction Pioneer, CourtQuant, To Close (accessed on 10.07.2024).
[6] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024).
[7] https://openai.com/index/chatgpt/.
[8] In the context of multilingualism, it is worth mentioning the free of charge release in March 2024 of the first large language model in Bulgarian language – BG GPT, developed by the Institute for Computer Science, Artificial Intelligence and Technology (INSAIT), located in Sofia, Bulgaria.
[9] Generic LLMs vs. Domain-Specific LLMs: What's the Difference? (accessed on 17.07.2024).

court cases.[10] Moreover, when it comes to legal practitioners, a growing percentage of lawyers declare that they use or plan to incorporate AI-powered tools into their daily work. In 2023, Wolters Kluwer published their report "Future Ready Lawyer 2023",[11] which analyses the results of a survey conducted among 700 legal professionals in law firms and corporate legal departments across the U.S. and nine European countries. According to that survey, 73% of respondents declare that they expect to integrate generative AI into their legal work in the next 12 months. Those excited about the technology highlight how it can boost efficiency and produce insights that are more comprehensive. However, the report also highlights some concerns about generative AI, most commonly related to "lack of authority, inconsistency, lack of explainability, and potential bias".

Even in the judiciary, which is traditionally more conservative, some recent examples of the introduction of AI-driven tools designed to facilitate the work of judges can also be pointed out.[12]

The ethical and human rights aspects of AI, and the application of AI in the legal domain, are being explored in more depth by legal academics and are increasingly finding a place in university curricula.[13] In the NGO sector, the emergence of specialised organisations that implement projects for responsible AI to benefit people, communities and society is noticeable.[14]

The POLINE project, funded by the European Commission's Justice Programme, also recognises the above values and in particular sets itself the practical goal of developing a Pilot Tool, which provides a robust and trustworthy use case of AI technologies for justice. After conducting preliminary research on navigating the landscape, below we present some of the most useful AI-powered functionalities that were identified and point to specific examples of such functionalities that are available in various legal information systems launched in Europe and North America.

[10] 25 Best AI Tools for Lawyers in 2024: Transform Your Practice (accessed on 9.07.2024).
[11] https://www.wolterskluwer.com/en/know/future-ready-lawyer-2023#key-findings (accessed on 15.07.2024).
[12] For instance, in 2022, the Supreme Judicial Council of the Republic of Bulgaria implemented a project co-financed by the EU: "Development and implementation in the judicial system of software for accelerated creation and reproduction of acts and other documents by dictation and automatic conversion of voice-to-text and accompanying processing systems".
[13] See "AI for law and governance" Scientific Unit of the University of Bologna, AI and Law Summer School at the European University Institute, "The 10 Best AI Training Courses for Lawyers".
[14] See "Responsible Ai UK": https://rai.ac.uk/.

9.2. Examples of AI-driven functionalities

9.2.1. Pre-trained legal chatbot / virtual assistant

As already mentioned, pre-trained legal chatbots are no longer rare and are embedded as core functionality in many of the operating legal information systems. The benefit for legal professionals who use them is, on the one hand, that they can communicate by asking questions and receiving answers in natural language, as if they were talking to a virtual colleague. On the other hand, the assistant does not just give an answer to the question, but also applies relevant legal arguments that find support in the applicable legislation and case-law. Such is the **CoCounsel** chatbot, a Casetext development that is now part of the Thomson Reuters portfolio (see Figure 9-1).

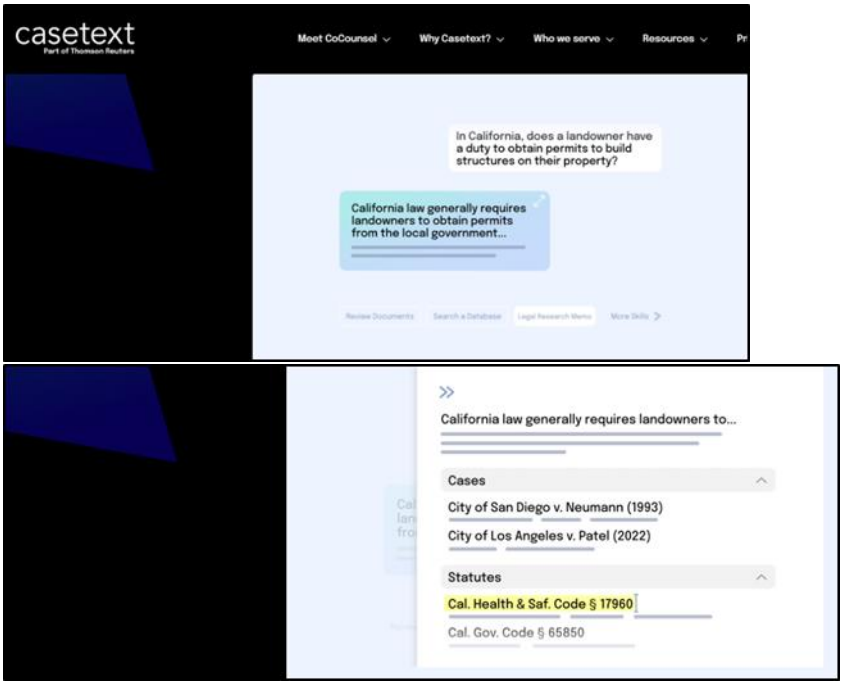


Figure 9-1. *Chat with CoCounsel - AI Assistant, developed by Casetext (source: https://casetext.com)*

The system responds to the prompt by offering a list of relevant precedent case law on the question at hand, as well as the applicable law, in this case the specific provisions of California Health and Safety Code.

An article published on the company's blog explains more about the technology behind CoCounsel, pointing out that the good results are due to the combination of artificial

intelligence with human insight.[15] The basis was the "brain" (the large language model GPT-4) and the "memory" (Casetext's own legal database, a comprehensive corpus of state and federal case law, statutes, regulations, codes, and rules). Based on this, specialised machine learning was carried out, involving a team of experienced lawyers and prompt engineers, so that the quality and reliability of the answers given by the system could be guaranteed and ensure that it would be safe for professional use.

In Europe, an example of a legal chatbot can be seen in the French legal information platform **Predictice**. Predictice's interface offers user-friendly predefined prompts that suggest what the system can help with – "Write a legal note", "Write a clause", "Find case law", "Translate this paragraph", etc. (see Figure 9-2).

Figure 9-2. *AI Legal Assistant, developed by Predictice (source: https://predictice.com/assistant)*

[15] What it takes to build an AI legal assistant lawyers can rely on: https://casetext.com/blog/building-an-ai-legal-assistant-lawyers-can-trust/ (accessed on 17.07.2024).

In another system that supports a legal chatbot, **Paxton Legal AI**, we can see a means of assuring the accuracy of the information provided. The virtual assistant's statement in response to a specific legal question refers to the relevant legal source, highlighting the exact location in the text of that source of the information sought. In this way, the user can ensure that the information provided by the chatbot is not a "hallucination" (see Figure 9-3).

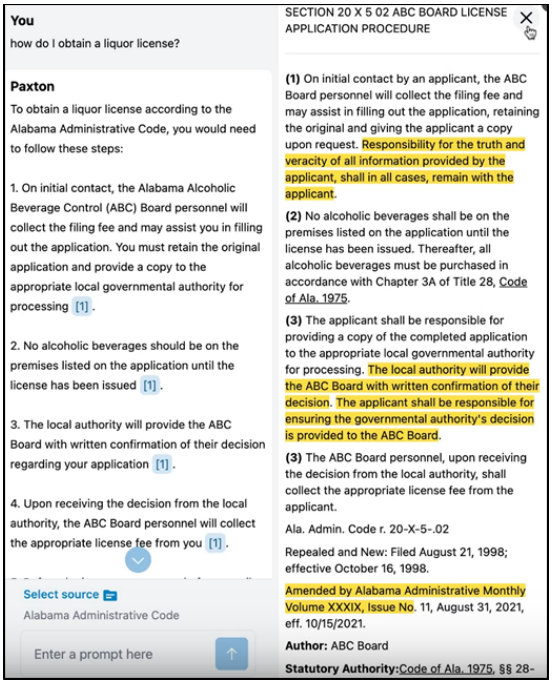


Figure 9-3. *Paxton AI Legal Assistant (source: https://help.paxton.ai/help)*

9.2.2. AI-generated summaries and extraction of key information

The AI capabilities are successfully leveraged to generate a summary of a legal document like a court decision, an appeal, a pleading, a contract, etc. This is appreciated by legal practitioners as it saves their working time. Dealing with long legal documents in less time is also facilitated by using AI functionalities to extract keywords, key provisions and key legal findings. Figure 9-4 shows the text of an AI-generated complaint summary created by Lex Machina's **Legal Analytics** app. The system is capable not just of extracting individual sentences from the text, but also of composing coherent natural language text itself, synthesising the most essential points of the case.

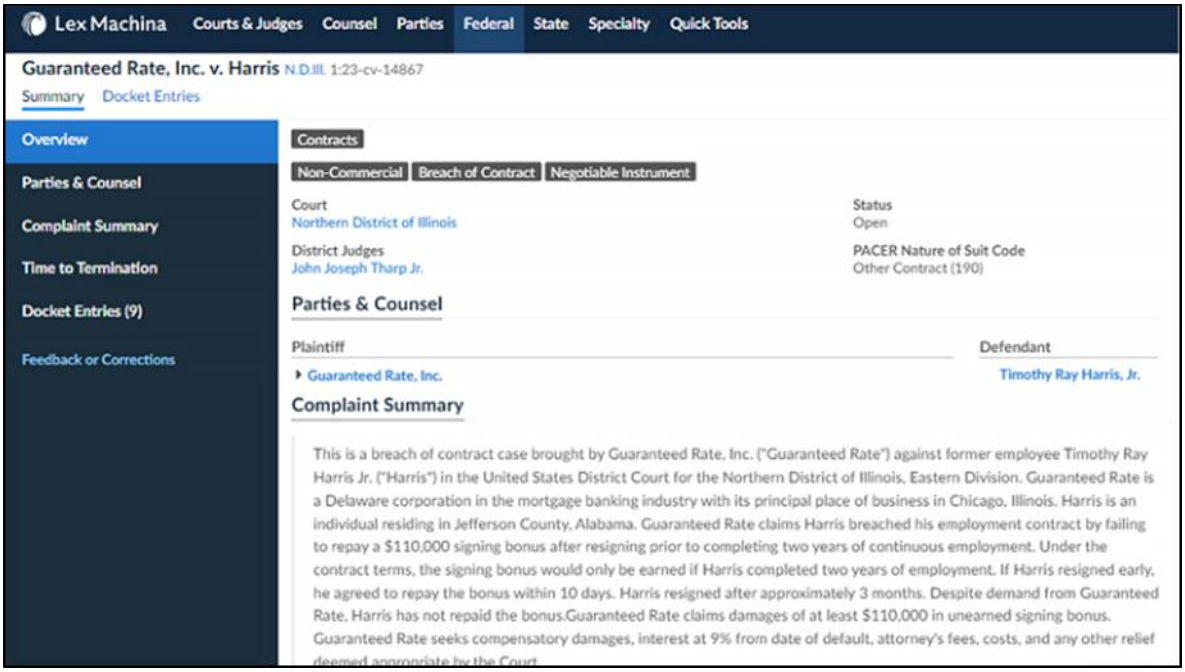


Figure 9-4. *AI-generated summary of plaintiff's complaint in an action brought in the Northen District Court of Illinois (source: https://lexmachina.com/legal-analytics/)*

Lawyers value as useful systems that provide a function to upload an external document, which the system can analyse and automatically extract relevant key information. Such functionality is used for the initial review of a new act of a judicial or administrative body, an opposing party's position, a draft contract, and much more. For example, from the description on the home page of the **Diligen** contract review system, we learn that it offers:

- Extraction of key information using natural language processing.
- Automatically generation of contract summaries in Word or Excel.
- Automatically extraction of key terms and key provisions from uploaded contracts to review contracts easily.

AI-powered extraction and highlighting of key information are also offered by **Westlaw Edge, CaseMark, Predictice, Harvey AI, DeepJudge**. What is interesting about the latter two systems is that they can search for an answer to the query and extract insights simultaneously from publicly available databases and from the user's own private databases. For this purpose, the integration of the tool with client management systems and document repositories is provided, which is also ensured by the relevant cybersecurity and confidentiality measures. Figure 9-5 shows that the **DeepJudge** system found and highlighted the requested information in one publicly available document (a court decision) and in one private document (an email).

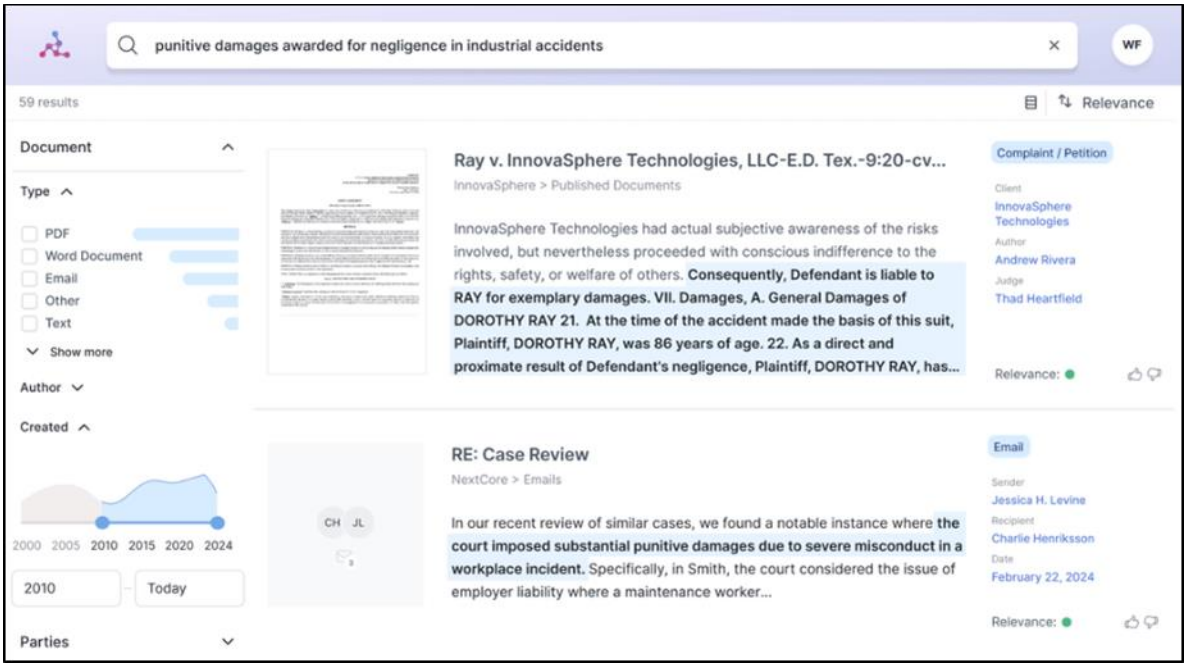


Figure 9-5. *"Knowledge Search" functionality of the DeepJudge system (source: https://www.deepjudge.ai/product)*

In any case, however, the companies providing such information services must be able to guarantee that with such integration the internal company information of their clients remains internal and will not be used without permission for subsequent training of AI algorithms.

9.2.3. Detection of gaps and risks. Suggestions for revisions. Outcome prediction

The power of AI tools also extends to detecting gaps, risks or errors and suggesting solutions to identified problems.

This can be used to check drafted depositions, terms of service or policies, to analyse legal trends, assess compliance with regulatory regimes such as GDPR, ESG, anti-money laundering measures, etc. Machine learning, based on a large number of cases, also enables the prediction of the outcome of a case (upheld/rejected claim) or the granting of an application before an administrative authority, e.g. percentage probability of granting a patent, license, etc.

Many applications offering such functionalities can be listed here:

- **Legal Analytics** – predictive analytics, insights regarding caseloads, timing, probable

case resolutions, specific findings, and damages awarded.

- **Legal Robot** – spotting potential pitfalls and suggesting better terms.
- **Harvey AI** – analysing legal trends, suggesting risk-mitigation strategies.
- **Blue J L&E** – predicting litigation outcomes and assessing risk, explaining each prediction in plain language.
- **OneLaw.ai** – intelligent clause suggestion capability.
- **LawGeex** – catching inconsistencies and suggesting revisions.
- **CoCounsel** – automatic review of contracts, identification of relevant clauses, conflicts and risks, recommendation of revisions.

Figure 9-6 shows the result of CoCounsel's automatic contract review – the system warns of a conflicting clause in the contract and suggests revisions to the text of that clause:

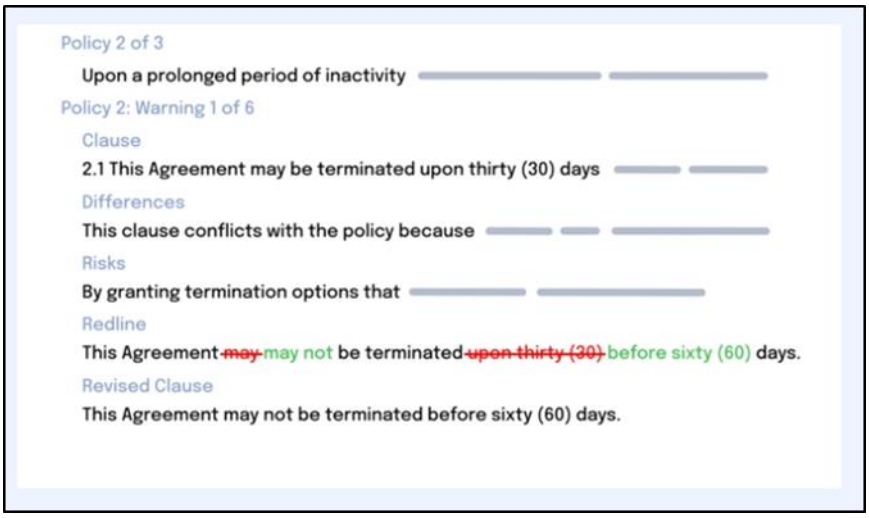
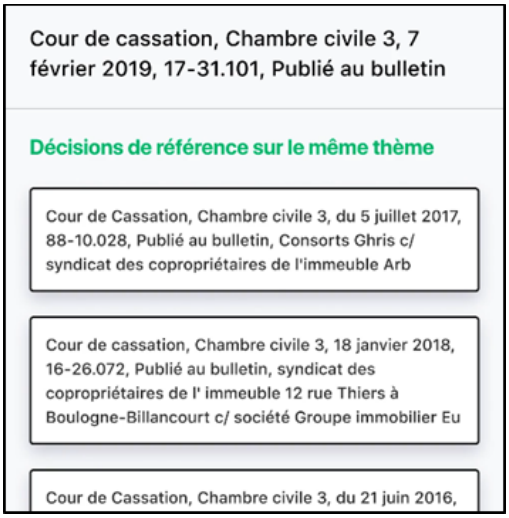


Figure 9-6. *"Contract review" functionality of the Casetext system (source: https://casetext.com/)*

9.2.4. Automatic identification of similar reasoning. Citation analysis

Finally, we will mention one feature that helps legal practitioners when they have to deal with a large volume of case law in a given legal branch, searching for precedents and similar cases. For such tasks, AI capabilities can be used to automatically recommend similar decisions, to classify case law according to certain criteria, as well as to visualise the relationships between cited decisions. Such functionalities accelerate case law research and allow lawyers toquickly findy arguments to support their claims.



For example, the French expert legal system **Doctrine** includes automatic recommendations of similar court decisions when reviewing any court decision available in the system. With the help of the tool, users can also perform a search for similar court decisions using the **same legal grounds** (see Figure 9-7).

Figure 9-7. *Court decisions on the same legal ground found in the Doctrine system (source: https://www.doctrine.fr/fonctionnalites)*

In this context, in an article presenting the **Paxton AI Citator**[16], the creators of this AI tool explain that it automatically "identifies and groups cases based on **similar legal principles** and does so even when cases do not directly cite each other". In addition, they claim: "Moreover, our Citator goes beyond traditional citation indicators by offering detailed explanations for each citation. This feature enhances the understanding of a case's relevance and legal standing, making the results easier to validate and increasing transparency within legal research processes".

We can see an example of visualization of a network of related cases also in another legal information system – **Bloomberg Law**.[17] Their Citation Map provides a visualisation of most cited cases, relationships among key cases, and changes over time for the point of law at issue (see Figure 9-8).

[16] Introducing the Paxton AI Citator: Setting New Benchmarks in Legal Research (accessed on 15.07.2024).
[17] Bloomberg Law Launches AI Enabled Research Features: Points of Law and Citation Maps (accessed on 16.07.2024).

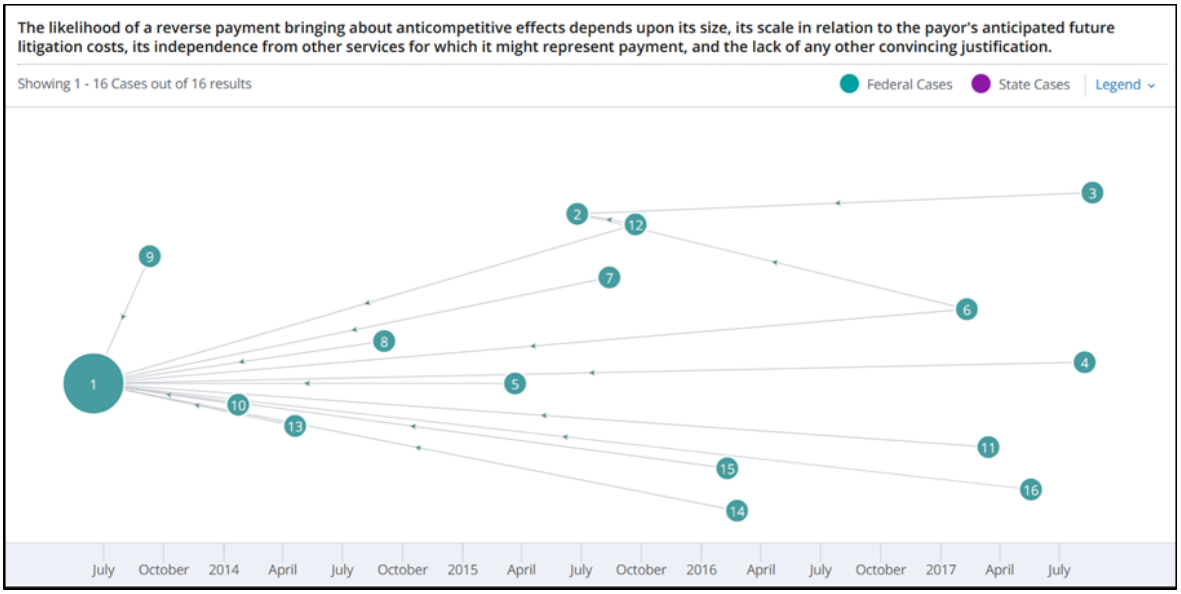


Figure 9-8. *The Bloomberg Law Citation Map (source: https://www.deweybstrategic.com/2017/10/6072.html)*

9.3. Comparison and conclusions

The above list of functionalities is not exhaustive. It is expected that new beneficial uses of AI methods in legal research systems are yet to be discovered. At this stage of art, we see that AI legal assistants can already converse in natural language, reason by citing provisions of current legislation and case law, summarise large legal texts, detect non-compliance with regulations, predict the outcome of cases, and discover similar cases. Thanks to their capacity to extract legal knowledge from the ocean of legal information, these new virtual legal assistants can not only save lawyers a lot of work time, but also increase their qualitative efficiency, helping to reveal undiscovered relationships between different sources of law and offering customised solutions. A major challenge for legal system providers in recent years will be to maximise the utility of LLMs by not only providing an answer in the form of human-like text, but also ensuring its reliability and validity. Last but not least, when developing legal systems based on AI, regulations related to the protection of personal and corporate data, copyrights, ethical and technical requirements, and standards for putting AI systems on the market should be strictly observed.

If we were to compare the POLINE Pilot Tool with other legal systems on the market, we would see that the functionalities planned for the POLINE Tool come to some extent close to the extraction of summary, keywords and key provisions and the highlighting of key

passages of text. However, the POLINE Pilot Tool offers legal professionals a new way to work with case law, by extracting judicial principles of law. The POLINE tool also differs from the others, because it is still largely experimental in nature. A limited area of law will be explored during the project and the datasets will include a relatively small number of court decisions. The subject scope of the POLINE ontology will be narrower to make it easier to test and validate. However, the ML-based technology we intend to use is scalable and eventually it could be applied to other legal domains.

10. Content and Functionality of the Pilot Tool

10.1. Description of the Legal Content

This section offers a concise overview of the primary legal content categories within the POLINE Pilot Tool and their corresponding sources. While judicial principles of law (JPOLs) serve as the cornerstone of the tool and its unique value proposition, they do not constitute a primary legal source, insofar as they are essentially extracts from court decisions. A comprehensive explanation of JPOL elements and the end-user functionalities of the Pilot Tool related to JPOLs can be found in Section 10.3.

10.1.1. National Legislation

The Legal Database Module will provide access to the latest consolidated versions of the national legislative instruments of Italy, Bulgaria and Sweden in the VAT legal domain. In addition to the consolidated texts currently in force, the tool's database will also include consolidated versions of repealed legislative instruments if the national case law related to the application of such acts is still considered an important source of inspiration for legal professionals. More specifically, this is the case with the Swedish VAT Act of 1994, which was only recently repealed in 2023.

In Italy, the main legal instrument governing VAT taxation is Decree No. 633 of the President of the Republic as of 26 October 1972 (OG n. 292 of 11-11-1972 - Ordinary Suppl. 1),[18] with all subsequent amendments, addendums and appendixes.

[18] Decreto del presidente della repubblica 26 ottobre 1972, n. 633 - Istituzione e disciplina dell'imposta sul valore aggiunto (GU n.292 del 11-11-1972 - Suppl. Ordinario n. 1): https://www.normattiva.it/eli/id/1972/11/11/072U0633/CONSOLIDATED

In Bulgaria, the main pieces of legislation in the area of VAT taxation are the Value Added Tax Act (promulgated, SG No. 63/4.08.2006, effective 1.01.2007)[19] and the Regulations for Application of the Value Added Tax Act (promulgated, SG No. 76/15.09.2006, effective 01.01.2007).[20]

In Sweden, the main legal instrument in the VAT field is the new Value Added Tax Act[21] (SFS No 2023:200, promulgated on 27.04.2023). Since the case law on the repealed Value Added Tax Act[22] (SFS 1994:200, promulgated on 30.03.1994; repealed on 01.07.2023) plays still an important role, the consolidated text of the repealed instrument will be also included in the database. In addition, we will consider the possibility of including some of the so-called 'preparatory works' in the database. These are documents that provide justification and context for legislation. While not legally binding, they play a crucial role in interpreting the law, often containing detailed explanations that enhance understanding of the legislation.

Along with the substantive legal framework, the Italian and Bulgarian codes of administrative and tax procedures will be included, which regulate the litigation proceedings in the respective matter.

10.1.2. National Case Law

The Pilot Tool will contain separate datasets of VAT-related case law of Italy, Bulgaria and Sweden. For the needs of corpus annotation, decisions of national courts will be pre-selected by legal experts of POLINE partners following common selection criteria. Firstly, the VAT cases will cover two key topics: exemptions in public interest and consideration. Secondly, the datasets will include only decisions of courts of the last instance, which only decide matters on points of law, i.e. judgments of the supreme courts of Italy, Bulgaria and Sweden. Decisions of lower instance courts will neither be selected for annotation, nor included in the system.

One part of the selected judgements from the national datasets will be tagged manually using the web-based Gloss annotation tool (see Section 8.2) with the aim to identify, label and set attributes to the judicial principles of law (JPOLs). These manually processed documents will be used for the initial experiments with ML algorithms. Another part of the selected judgements will be used as a test set for the semi-automatic annotation with LLMs. Both manually and automatically annotated decisions will be included in the final annotation corpus and the Legal Database Module.

---

[19] Закон за данък върху добавената стойност (ДВ, бр. 63 от 04.08.2006 г.): https://dv.parliament.bg/DVWeb/showMaterialDV.jsp?idMat=2310. This link leads to the basic act, i.e. to the initial version published in the Bulgarian State Gazette. Since in Bulgaria there is no official source of consolidated legislation, the Poline Pilot Tool will include the unofficial consolidated versions of the Bulgarian legal instruments available in the legal database of the partner APIS.

[20] Правилник за прилагане на Закона за данък върху добавената стойност (ДВ, бр. 76 от 15.09.2006 г.): https://dv.parliament.bg/DVWeb/showMaterialDV.jsp?idMat=2549

[21] Mervärdesskattelag (2023:200): https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/mervardesskattelag-2023200_sfs-2023-200/

[22] Mervärdesskattelag (1994:200): https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/mervardesskattelag-1994200_sfs-1994-200/

10.1.3. EU Legislation

Despite the fact that decisions of national courts in the chosen legal domain refer predominantly to national legislative instruments, the POLINE Pilot Tool will include all basic acts of EU legislation in the VAT domain. The list primarily includes directives that have been transposed into the national legal orders of Italy, Bulgaria and Sweden and, therefore, are rarely cited by national courts, which, in general, refer to the national provisions transposing the respective EU provisions.

The main instrument of EU legislation is the **VAT Directive** (2006/112/EC). Other important instruments of EU legislation in the VAT domain are:

- Directive 2008/9/EC (VAT Refund – EU business)
- Directive 86/560/EEC (VAT Refund – non-EU business)
- Directive 2009/132/EC (VAT-free importation)
- Directive 2006/79/EC (private consignments)
- Directive 2007/74/EC (travellers' allowances).

Binding implementing measures to ensure uniform application of the VAT Directive can be found in the **VAT Implementing Regulation** (Council Regulation (EU) No 282/2011). Those measures are directly applicable without transposition into national law.

In addition, Member States, to prevent fraud or simplify procedures for collecting VAT, may be authorised to derogate from the VAT Directive.[23] The European Commission, for reasons of transparency and legal certainty, maintains a list of derogations in force.[24]

10.1.4. Case law of the Court of Justice of the EU

The Pilot Tool will contain judgements and orders of the Court of Justice of the European Union (CJEU) in the VAT domain. In the same way as with national case law datasets, the manually and automatically annotated CJEU judgements will cover the two above-mentioned VAT topics: exemptions in public interest and consideration. Thus, the system will provide users with a more comprehensive understanding of the extracted JPOLs in the two selected VAT-related topics both at the EU and national levels.

Although the Curia Portal[25] is the primary source of CJEU cases, the POLINE partners will rely on the EUR-Lex Portal[26] to gather EU case law related to the chosen legal domains. The main reason for this preference is the fact that the information is provided in a well-structured XML format as open data for reuse for commercial or non-commercial purposes. The judicial documents will be available in

[23] For more information about the derogation options see here: https://ec.europa.eu/taxation_customs/derogations_en.
[24] See here the latest list of derogations: https://ec.europa.eu/taxation_customs/system/files/2021-05/vat_index_derogations_en.pdf
[25] https://curia.europa.eu/
[26] https://eur-lex.europa.eu/homepage.html

full text, accompanied in addition by a rich set of metadata (the so-called 'bibliographic references' in EUR-Lex).

The CJEU cases will be sourced from the following classifier of the 'Directory of case-law' classification scheme of EUR-Lex: 4.10.02.01 Value-added tax.

10.2. Non-functional Requirements

Non-functional requirements define the technical and business constraints for the operation of a software system. These are quality properties of the system that sit outside of its specific features and functionality.

Availability and reliability

POLINE Pilot Tool shall be developed as a web application providing 24/7/365 service. The tool will be freely available to its users without any login requirements. Although critical bugs or failures are extremely unlikely, uptime availability of the service is estimated at 99.5% on a yearly basis.

Performance

POLINE Pilot Tool combines, on the one hand, functional elements that are typical for legal databases and, on the other hand, advanced AI and legal analytics features. With regard to the former, there is a generally accepted standard for the response time of searching operations in legal databases, which is less than 0.1 seconds. As for specific AI and legal analytics features, such as searching for similar JPOLs by using the functionality of the Link Visualisation Module or providing an assessment of the relevant JPOLs to the user's case through the Customised Detection Module, we will strive to achieve a response time of less than 3 seconds.

Portability and compatibility

As a web application, POLINE Pilot Tool will be able to run from a variety of platforms and devices having an installed browser supporting the HTML5 features. However, the use of the service by smartphones would not be appropriate in view of its specifics as a legal information retrieval system.

The tool will be compatible with the five most popular browser engines: Google Chrome, Mozilla Firefox, Safari, Microsoft Edge and Opera.

The minimum hardware and software requirements for using the POLINE Pilot Tool coincide with the minimum hardware and software requirements for installing the respective browser.

Security and compliance

There are no critical security or compliance issues related to the operation of the POLINE Pilot Tool. Since the web application will not support user authentication and user-related personalisation features, it will not store personal user data except a few strictly necessary cookies. Some concerns raise only the functionality of the Customised Detection Module because it requires an input of legal

text inserted by the user, which may contain personal and/or confidential information. Thus, communication with the tool's website will be secured via HTTPS (Hypertext Transfer Protocol Secure) in order to protect the integrity and confidentiality of data between the user's computer and the website. In addition, the data collected from the user will not be stored by the tool after closing or expiring the session.

In order to protect the service from hacker attacks, the tool's website will be put behind a web application firewall, such as Cloudflare[27] or similar.

Localisation

POLINE Pilot Tool is targeting users in Italy, Sweden and Bulgaria that are interested in the respective national legal content and its interconnection with the EU law. Thus, the tool will provide multilingual interface in IT, SE and BG language and appropriate means to automatically identify the visitor's country. Based on these, the system will apply the respective interface language. For visitors from other countries, the interface language will be set to English.

10.3. End-user Functionality and Features

POLINE Pilot Tool is an AI-powered online platform for the retrieval and analysis of EU and national judicial principles of law in the domain of VAT. The platform includes three modules, which functionalities are described in detail in the following sections.

10.3.1. Legal Database Module

The core mission of the Legal Database Module is to provide access to a multilingual database of judicial principles of law (JPOL) extracted from the CJEU and national case law in the VAT field. At the same time, the module will serve as a traditional legal database, ensuring search and reference capabilities in the full texts of the relevant legislative and judicial acts at the EU and national levels.

The home page of POLINE Pilot Tool will simultaneously play the role of a home screen for the Legal Database Module. It will consist of three panels:

- A horizontal panel at the top of the screen with a ***search box***, ***command buttons*** and/or a ***menu*** for managing individual functionalities (e.g. selecting the interface language).
- A vertical panel on the left side of the screen with ***filters*** for narrowing down search results based on specific criteria or preferences.
- A wide horizontal panel on the right side of the screen displaying a ***list of JPOLs*** obtained as a result of searching and/or filtering.

[27] https://www.cloudflare.com

On the top panel, there will also be command buttons, through which users can call the other two modules of the Pilot Tool and open the web pages providing access to their functionalities (see Section 10.3.2 and Section 10.3.3).

The Legal Database Module will offer the following basic features to its users:

- **Multilingual user interface.** The user interface will be available in four EU languages: English, Italian, Swedish and Bulgarian. Besides menu items, forms, windows, controls and other graphical user interface elements, the module will also provide language versions of the legal documents in all the above-listed languages if such versions are available (in practice, this only applies to EU legislative instruments and judgments of the CJEU).
- **Navigation and browsing**. The module will offer appropriate means, such as the VAT Ontology, which can be used as a thematic classifier of the legal content, and filters, to quickly and easily navigate and browse relevant JPOLs and legal documents. The list of JPOLs will allow users to view and read the first two or three lines of the text of each JPOL, so they can quickly understand the legal issue(s) to which the respective principle of law refers to. A special button on the right side of the text of the JPOL will allow users to expand and collapse the full text of that principle of law without the need to open it as a separate document.
- **Searching and filtering**. The module will allow its users to enter a search query in the search box and filter the search results. The main filter that will allow users to refine their searches by legal issue of interest is based on the classification of JPOLs according to concepts of the VAT ontology. In addition, the following filters based on JPOL's attributes will be also used to narrow down the search results in a list of JPOLs:
    - *Jurisdiction* – European Union / Italy / Sweden / Bulgaria
    - *JPOL type* – New / Cited
    - *JPOL origin* – EU / National / EU & National
    - *Designation* – Explicit / Implicit
    - *Interpretation type* – Factual / Non-factual
    - *VAT relevance* – VAT-related / Other.
- **Working with JPOLs and legal documents**. Users will benefit from the various features for viewing, browsing, printing and exporting the text of JPOLs and legal documents. The Legal Database Module will offer effective means to extract knowledge from legal data by exploring legal citations, references and other relationships between JPOLs and/or documents.

The "judicial principle of law" (JPOL) is the main structural unit in the Pilot Tool and its Legal Database Module. In POLINE project, it is understood as a part of the text of a judicial decision providing for the interpretation of a rule or a legal principle or their application to a concrete case.

The identification of the JPOLs is subject to the manual and automated annotation activities performed by the project partners within the framework of WP2 and WP3 tasks. The automated extraction relies on machine learning and NLP methods and algorithms (see Section 8.3), which will also be used to facilitate link detection and visualisation between different principles of law, and their conceptual and logical connection with the ontology of VAT concepts.

Similar to legal documents, each JPOL can be viewed in a list, opened, indexed and searched by the search engine, and interlinked with other JPOLs and documents. It will consist of the following structural elements:

- **Title** – the JPOLs' titles are formed by the titles of the court decisions from which they were extracted and the sequential number of the respective JPOL in the text of the decision, i.e. they adhere to the following format: *[title of the decision]–[seq. number of the JPOL].*
- **Text** – the full text of the JPOL as extracted from the text of the decision. The minimum length of the text of a JPOL is one sentence. Usually, the text of JPOLs consists of one or more whole paragraphs. When a paragraph is numbered, the number must be included in the JPOL's text.
- **Attributes and other metadata** – they provide information about JPOL's assignment to different classification criteria, its association with concepts of the VAT ontology and references to cited EU or national case law and/or legislation. The attributes used to classify JPOLs were already specified in Section 8.2 above.

Figure 10.1 and Figure 10.2 illustrate the JPOL structure by providing examples of a JPOL originating from a judgment of the CJEU and a JPOL from a judgment of the Bulgarian Supreme Administrative Court:

Judgment of the Court (Eighth Chamber) of 9 February 2023, C-482/21 – JPOL1

Ontology:

Taxable amount – Reduction – Principle of fiscal neutrality – Principle of effectiveness

Judicial Principle of Law (JPOL):

32 It should be recalled, first, that Article 90(1) of the VAT Directive ↗ refers to cases of cancellation, termination, rescission, total or partial non-payment or price reduction after the time when the transaction which gave rise to the payment of the VAT is carried out. That provision requires the Member States to reduce the taxable amount and, consequently, the amount of VAT payable by the taxable person whenever, after a transaction has been concluded, part or all of the consideration has not been received by the taxable person. That provision embodies one of the fundamental principles of that directive, the principle of fiscal neutrality, according to which the taxable amount is the consideration actually received and the corollary of which is that the tax authorities may not collect an amount of VAT exceeding the tax which the taxable person received (order of 3 March 2021, FGSZ, C-507/20, EU:C:2021:157, ↗ paragraph 18 ↗ and the case-law cited).

Full text of the decision

Attributes:

Cited – EU – VAT related – Implicit – Non-factual

Cited case law:

C-507/20:

- N 18

Cited legislation:

Directive 2006/112/EC:

- Article 90(1)

Figure 10-1. *Example of a JPOL in a judgment of the CJEU*

Решение № 5645 от 19.05.2015 г. на ВАС по адм. д. № 5446/2014 г., I о., докладчик съдията Мадлен Петрова – JPOL3

Ontology:

Exemptions – Exemptions for certain activities in the public interest – Exemption for hospital and medical care and closely related activities – Duly recognised establishment of a similar nature

Judicial Principle of Law (JPOL):

Тълкуване на понятието "надлежно признато заведение с подобен характер" е дадено в решението по делото на Съда на Европейските общности от 06.11.2003 г. по дело C-45/01 ↗ (Dornier). Съгласно цитираното решение нормата на чл. 13 A, § 1, буква б) от Шеста директива ↗ не съдържа каквито и да било процедури и условия за определяне на надлежно признатите заведения. Такива изисквания не се съдържат и в разпоредбата на чл. 132, § 1, б. б) и б. в) от Директива 2006/112/ЕО.↗ Поради това и съгласно чл. 133 от Директива 2006/112/ЕО ↗ условията се определят от държавите - членки /в този смисъл т. 64 от решението по дело C-45/01/. ↗

Full text of the decision

Attributes:

Cited – EU & National – VAT related – Implicit – Non-factual

Cited case law:

C-45/01:

- N 64

Cited legislation:

Sixth Council Directive 77/388/EEC:

- Article 13A(1)(b)

Directive 2006/112/EC:

- Article 132(1)(b)
- Article 132(1)(c)
- Article 133

Figure 10-2. *Example of a JPOL in a judgment of the Bulgarian Supreme Administrative Court*

The examples above also illustrate some of the features of the Pilot Tool that will be accessible at JPOL level, i.e. when the user opens the text of a JPOL. For each JPOL, the user will be able to open the full text of the court decision from which it originates and to examine the legal context in which it appears. The concepts of the VAT ontology associated with a particular JPOL, will be visualised as links that lead to a list of all JPOLs in the tool classified with the same ontology concept. Similarly, next to each identified legal citation in the text of a JPOL there will be a button (indicated with a blue arrow icon "↗" in the examples above) that leads to a list of all JPOLs in the tool, which are citing the same (provision of a) legal instrument or (paragraph of a) court decision.

10.3.2. Link Visualisation Module

This module will be used for linking relevant EU and national judicial principles of law.

The module has two functionalities. First, analysing textual and semantic similarities between judicial principles will allow the visualisation of similar principles both within the same legal system and between different legal systems. In particular, the module will allow the end-user to i) select one principle of law in one legal system, then ii) search for similar principles in the same or other legal systems, and iii) explore differences under a comparative visual setting.

Second, the module will include a user-friendly visualisation of networks of decisions relevant to the judicial principle concerned. These maps will be based on the identification of explicit references in the decisions and textual and semantic similarity and will include a web link to the case-law where the selected principle has been used.

10.3.3. Customised Detection Module

This interactive module supports users in identifying and assessing judicial principles of law on VAT in a specific decision.

In particular, the module will (i) allow the user to upload a decision or a tax audit act and receive a customised legal assessment on the relevant judicial principles of law in the document at hand. Moreover, the module will (ii) allow the evaluation of the correspondence of such principles with those existing at the national and EU levels or the selection of more relevant or more recent principles. Such a module will increase European taxpayers' awareness of the correct application of VAT law and support them in deciding whether it is making an appeal or claiming a violation of European law. Furthermore, it will empower citizens by facilitating their active role in the enforcement of VAT law and the pursuit of fair taxation.

As concerns the visualisations, the tool will identify and highlight automatically extracted JPOLs and related paragraphs and will provide an overview of selected attributes as per annotation guidelines (see D.2.2):

- *Type*, i.e. assessing whether the principle is new or cited;
- *Origin*, i.e. indicating if the JPOL pertains to a national or EU case law (or both);
- *Explicit/Implicit*, i.e. whether the JPOL is explicitly referred to in the decision as a "principle of law"[28];
- *Factual/Non-factual*, i.e. whether the JPOL refers to the interpretation of the law or to facts;
- *VAT-related*, i.e. whether the JPOL is specific to VAT or other legal aspects relevant to the decision.

To help users visualise those differences better, it will be possible to highlight and display different types of JPOLs in different colours within the decision text.

[28] For instance, this often occurs in the case law of the Italian Court of Cassation.