

Building Blocks of Judicial Reasoning: Defining and Extracting Judicial Interpretative Formulas with LLMs

Rachele Mignone^{1,2,*}, Alessia Fidelangeli², Piera Santin³ and Luigi Di Caro¹

¹University of Turin, Italy

²CIRSFID, Alma AI, University of Bologna

³European University Institute

Abstract

The Court of Justice of the European Union (CJEU) frequently employs specific interpretative statements from previous rulings as building blocks to construct legal arguments. We define these recurring formulaic statements as “Judicial Interpretative Formulas” (JIFs) and present a novel computational approach for their automated extraction. Our methodology combines legal domain expertise with Large Language Models (LLMs), using an iterative process where expert error analysis refines both the formal definition of JIFs and extraction performance. We evaluated three LLMs (Claude 3.7 Sonnet, DeepSeek-R1, and Gemini 1.5 Pro) on 106 CJEU Value Added Tax judgments from 2006-2024. Claude 3.7 Sonnet with few-shot prompting achieved optimal performance with an F1-score of 0.932, precision of 0.919, and recall of 0.944. This approach serves a dual purpose: it provides empirical evidence of the widespread nature of JIFs and facilitates the creation of an EU case law dataset. This resource is expected to be used by legal scholars for conducting analyses on evolving trends in judicial reasoning across Europe, while simultaneously offering a practical reference for judges to ensure the consistent application of EU law.

Keywords

Large Language Models, Judicial Interpretative Formulas, Court of Justice of the EU, VAT law, Automated Extraction, Legal Concept Definition

1. Introduction

The Court of Justice of the European Union (CJEU) is mandated to ensure the uniform interpretation and application of EU law, primarily through the preliminary ruling procedure.

A distinctive and evolving characteristic of the CJEU’s judicial practice is the frequent verbatim reproduction or subtle rephrasing of interpretative statements from past judgments. This systematic reuse of language, often termed the “copy-pasting” [1] or “LEGO” technique [2], means the significance of a precedent often resides in these specific, recurring paragraphs rather than the full judgment text. We formalize this recognized drafting style by introducing the term Judicial Interpretative Formulas (JIFs) for these recurring building blocks of legal reasoning.

While the legal literature acknowledges this drafting style, there has been a lack of a clear, formal definition and a scalable, computational method for their systematic identification. Our work addresses this gap by proposing a human-centric methodology for the bulk extraction of JIFs from CJEU judgments.

Our interdisciplinary approach leverages Large Language Models (LLMs) to automatically extract JIFs using a binary text classification task. This methodology involves an iterative process that tightly couples legal domain expertise with Natural Language Processing (NLP): domain experts perform qualitative error analysis on the LLM output, allowing for a precise refinement of the initial JIF definition.

This research provides a scalable method to extract JIFs, empirically confirming their widespread existence, and yields a dataset that can be used by scholars seeking to analyze the evolution of judicial reasoning and by practitioners aiming to ensure the consistent application of EU law.

CLAIRvoyantS: Conventicle on Artificial Intelligence Regulation and Safety @ JURIX 2025

*Corresponding author.

✉ rachele.mignone@unito.it (R. Mignone)

🆔 0009-0009-2699-8730 (R. Mignone); 0000-0002-8281-4176 (A. Fidelangeli); 0000-0002-0734-9657 (P. Santin);

0000-0002-7570-637X (L. D. Caro)



© 2025 This work is licensed under a “CC BY 4.0” license.

After discussing the motivation (Section ??), the related work (Section 3), and the dataset (Section 4), this paper details our interdisciplinary method (Section 5), including how we formally defined and automatically extracted JIFs with the use of LLMs. We then present a quantitative and qualitative evaluation (Section 6) before concluding and outlining future work (Section 7).

The code, full prompts, extracted data, and metrics are publicly available on Github¹.

2. Motivation

The preliminary ruling procedure is central to the CJEU’s role, guiding national courts on interpreting EU law to ensure consistency. While formally binding only in the originating case, these rulings have a broader influence, shaping subsequent judicial reasoning and serving as interpretative benchmarks for legislators, administrations, and courts across harmonized areas of law.

From its inception, the CJEU adopted an abstract and general style to formulate interpretations that extend beyond the specific dispute. Over time, certain highly significant statements have achieved foundational status in EU case law, often designated as “formulas” and associated with the name of the originating case (e.g., the paradigmatic *Dassonville* formula in tax law).

Initially, judgments contained only a limited number of these formulaic expressions. However, the Court’s drafting style has evolved, leading to an increasing number of paragraphs being written in a formulaic manner. This proliferation caused a fundamental shift: these formulations ceased to function merely as occasional, groundbreaking assertions. Their importance became tied to the very manner of expression itself. When the Court articulated a proposition in a formulaic style, that statement acquired greater normative weight and a higher likelihood of being cited and reiterated in subsequent case law.

This pattern, although recognized in legal literature and often described as a “copy-pasting” or “LEGO” technique, lacks an official name. We propose the term Judicial Interpretative Formulas to address this phenomenon. This term recognizes the common scholarly usage of “formula” while emphasizing their judicial and “interpretative” role in shaping EU law. Our focus is specifically on these recurring, reused, or newly introduced statements that constitute the very fabric of the Court’s ongoing reasoning.

To empirically demonstrate the widespread presence and functional significance of JIFs, our work proposes a novel methodology: we employ LLMs for their automatic extraction. This computational approach establishes a foundation for their large-scale, systematic study.

The resulting dataset will enable legal scholars to conduct large-scale analyses on the evolution of judicial reasoning across the European Union, while also providing judges with an accessible knowledge base to ensure the consistent and informed application of EU law.

3. Related Work

Most studies on interpretative statements of courts concern common law legal systems, especially the United States, where the debate is dominated by the distinction between holding and dictum. The importance and intensity of this debate are related to the principle of *stare decisis*. In systems where higher court holdings are binding, correctly identifying the holding is crucial for determining the precedential effect of a decision. In contrast, since the CJEU does not operate under a strict doctrine of *stare decisis*, we concluded that this distinction was not suitable for our purposes. Narrowing the focus to the European context, we identified three main relevant strands of literature. First, the debate on the so-called competence creep of the European Union, which highlights the gradual and often opaque expansion of EU competences through CJEU case law [3, 4, 5]. Second, some literature refers to the drafting style of the CJEU described as “copy-pasting” or “LEGO” technique, to highlight that important passages of earlier judgments are reproduced *verbatim* or with minimal modification in subsequent decisions, regardless of their formal status [1, 2]. Third, and most relevant for our purposes, the European legal literature has used the concept of “formulas” to describe the practice of reproducing earlier binding interpretative statements in later judgments [6].

¹Github repository: github.com/RacheleMignone/JIF_extraction

Recent research on the automated extraction of interpretative statements has combined natural language processing, machine learning, and knowledge representation to address the complexity of legal texts. Two key approaches to extracting legal principles from court judgments stand out. [7] improved on older citation-based methods by using a combination of NLP and machine learning to identify *ratio decidendi* from paragraphs cited in future cases rather than from citing ones. This technique is known for its high accuracy. Separately, [8] developed an automated system that uses regular expressions and heuristics to map citation networks in CJEU rulings, creating hierarchical structures to study how legal principles evolve. Knowledge discovery techniques, such as association rule mining, have also been explored to uncover defeasible rules that represent the underlying reasoning in legal decisions, with formalization in systems like Defeasible Logic [9]. Deep learning models, including transformer-based architectures, have further advanced the field by enabling the summarization and structured extraction of legal holdings and principles, as demonstrated in recent work on Italian legal texts [10, 11].

While LLMs show remarkable capabilities in generalistic information extraction tasks [12, 13], recent research on LLMs in the legal field has highlighted the need for hybrid approaches with human oversight to address challenges such as the risk of hallucinations and explainability concerns [14, 15, 16, 17, 18]. Studies exploring enhanced reasoning approaches show promising results: [19] demonstrated that combining retrieval-based methods with Chain of Thought reasoning and few-shot prompting improved accuracy and F1 scores compared to zero-shot baselines, while [20] developed an error taxonomy framework showing that incorporating error feedback in prompting techniques increases LLM performance. [21] proposed SELF-ALIGN, a principle-driven approach using minimal human supervision that significantly outperformed state-of-the-art systems.

Building on these works, our approach combines the abilities of LLMs with the reliability of human expertise to efficiently extract interpretative formulas from CJEU judgments.

4. Dataset

This study employs a manually curated dataset of CJEU judgments concerning VAT law, covering the period from 2006 to 2024.

Since the use of formulas is particularly characteristic of the European case law, this study takes the CJEU as its primary focus. This choice was also made because of the accessibility and structured format of CJEU judgments, which are parsed into discrete textual units suitable for automated processing. The judgments were retrieved from EUR-Lex², the official online repository of EU law.

Legal experts specializing in VAT law chose and gathered CJEU judgments focusing on two key subdomains: *Taxable Amount* and *Exemptions for the Public Interest*. Initially, 110 relevant CJEU judgments were identified (57 related to Taxable Amount and 53 related to Exemptions). However, four judgments concerning Taxable Amount were unavailable on EUR-Lex and subsequently excluded. The final dataset comprised 106 judgments, evenly distributed between the two subdomains, with 53 judgments in each.

To optimize subsequent steps of processing by LLMs, the judgments underwent preprocessing. Specifically, each judgment was truncated to its motivation section, the sole location where JIFs occur, as highlighted by the legal analysis described in section 5.1. This targeted approach reduced the input size while retaining the most relevant information. Further formatting was applied to eliminate extraneous character sequences, such as excessive newline characters, ensuring data cleanliness and efficient processing. This preprocessing step yielded an average document reduction of 15,317 characters or 2,384 words. The preprocessed text was segmented into paragraphs and stored in CSV format, preserving original paragraph numbering for indexing and retrieval purposes.

Although categorized into distinct subdomains, the judgments were treated as a unified dataset for analysis, given their common ground in the VAT domain.

²<https://eur-lex.europa.eu>

5. Methodology

This section outlines a cross-domain methodology that spans from the formal definition of JIFs in the legal domain to their automated extraction from CJEU judgments.

5.1. Judicial Interpretative Formulas Definition

Though the use of “formulas” has grown steadily and their presence has been noted in the literature (as shown in section 3), the features of these interpretative statements have not been systematically studied.

To address this gap, we conducted both a theoretical and empirical analysis of CJEU case law to provide a formal definition of JIFs and specify their features.

The theoretical analysis investigated how JIFs appear in CJEU case law. This involved studying the Court’s judgments, the structure and style of its reasoning, and its writing style, drawing inspiration from the work of [22]. This study also examined the role of precedent in the CJEU’s case law, the Court’s function in ensuring a uniform application of EU law, and the creative aspect of its judgments in promoting the harmonization of tax law.

The empirical analysis complemented the theoretical one through the systematic examination of selected CJEU judgments in the VAT field. The objective was to identify the occurrence and form of JIFs within this case law. Specific attention was given to (a) whether judgments were divided into paragraphs; (b) in which part of the decision JIFs appeared; (c) the frequency and consistency of cross-references to JIFs contained in earlier decisions; and (d) whether the statements addressed only VAT issues or extended to other matters. The analysis also considered how abstract the interpretative statements were, the extent to which they incorporated factual references, and the frequency of overruling or distinguishing prior formulations.

On this basis, we elaborated a definition of JIF as a segment of text extracted from the argumentative part of a judgment which sets out either (i) the interpretation of a rule or principle; (ii) the consequences stemming from its interpretation or application; (iii) the subsumption of a fact within a rule, or (iv) the qualification of a factual hypothesis as falling within a legal concept. The empirical findings also allowed us to specify further features of the JIFs. Frequently, a JIF is present at the conclusion of the reasoning concerning a preliminary question, where the Court decides whether a certain interpretation of a national rule is or is not precluded by European law. In many cases, these JIFs are very specific and tied to the factual circumstances of the case. JIFs are not confined to substantive VAT issues: they may address procedural questions, interpretative methods, or the consequences stemming from general legal principles. Their novelty also varies: some decisions introduce new interpretative statements, while others reproduce earlier ones, either through *verbatim* citations of precedent or through subtle reformulations. However, the presence of a citation does not automatically imply the existence of a JIF (e.g., as precedents may sometimes be invoked only for distinguishing purposes).

5.2. Prompt Validation and Model Selection

To identify the optimal combination of LLMs and model-agnostic prompt for extracting JIFs from CJEU judgments, a systematic model selection process was implemented. An LLM set consisting of Claude 3.7 Sonnet³, DeepSeek-R1⁴, and Gemini 1.5 Pro⁵ was evaluated against a manually annotated dataset comprising 11 CJEU judgments, randomly selected from the initial dataset.

To facilitate efficient computational access and integrate diverse LLMs within a unified environment, Quora’s Poe⁶ was employed, employing its infrastructure for the execution of various LLM queries. The use of this tool significantly enhanced the experimental workflow by centralizing access to multiple

³model snapshot: 2025-02-19. Temperature: 0.5, top-p: 0.7, top-k: 35

⁴model snapshot: 2025-05-28. Temperature: 0.35, top-p: 0.7, top-k: 35

⁵model snapshot: gemini-1.5-pro-002. Temperature: 0.20, top-p: 0.7, top-k: 35

⁶<https://poe.com/>

		Precision	Recall	F1-score
Zero-Shot	Gemini 1.5 pro	0.644	0.835	0.727
	Deepseek R1	0.687	0.978	0.807
	Claude 3.7 Sonnet	0.688	0.914	0.785
Few-Shot	Gemini 1.5 pro	0.683	0.907	0.779
	Deepseek R1	0.733	0.858	0.791
	Claude 3.7 Sonnet	0.708	0.978	0.821
Chain-of-Thoughts	Gemini 1.5 pro	0.652	0.764	0.704
	Deepseek R1	0.716	0.978	0.827
	Claude 3.7 Sonnet	0.678	0.964	0.796
Few-Shot Chain-of-Thoughts	Gemini 1.5 pro	0.682	0.711	0.696
	Deepseek R1	0.708	0.993	0.827
	Claude 3.7 Sonnet	0.687	0.978	0.807

Table 1
Model selection evaluation table

		Precision	Recall	F1-score
Zero-Shot	Gemini 1.5 pro	0.806	0.815	0.810
	Deepseek R1	0.864	0.961	0.910
	Claude 3.7 Sonnet	0.887	0.922	0.904
Few-Shot	Gemini 1.5 pro	0.844	0.877	0.86
	Deepseek R1	0.891	0.817	0.852
	Claude 3.7 Sonnet	0.901	0.972	0.935
Chain-of-Thoughts	Gemini 1.5 pro	0.829	0.76	0.793
	Deepseek R1	0.895	0.955	0.924
	Claude 3.7 Sonnet	0.874	0.972	0.920
Few-Shot Chain-of-Thoughts	Gemini 1.5 pro	0.865	0.707	0.778
	Deepseek R1	0.887	0.972	0.928
	Claude 3.7 Sonnet	0.884	0.983	0.931

Table 2
Updated model selection evaluation table

models, thereby reducing overhead associated with managing disparate API endpoints and improving the overall efficiency of computational tasks.

Each model was tested across four distinct prompting strategies to assess performance in terms of precision, recall, and F1-score in identifying interpretative formulas. The evaluation was conducted automatically using predefined metrics, ensuring an objective comparison of model efficacy. The results of this process, aimed to determine the most effective model-prompt pairing for legal text extraction, are visible in Table 1.

Following the quantitative analysis, domain experts conducted a qualitative assessment of the extracted JIFs. This crucial step revealed the subjective nature of the task, as what were initially considered false positives were frequently confirmed to be valid JIFs upon closer inspection. This iterative process of concept refinement, whereby the formal definition of JIFs was updated based on expert feedback and subsequent re-evaluation, resulted in an updated gold standard. The new performance metrics are detailed in Table 2.

5.3. Extraction

Based on the systematic model selection process, the extraction of JIFs was performed using the Claude 3.7 Sonnet LLM. This model was chosen because, when paired with a few-shot prompting strategy, it demonstrated the highest performance in precision, recall, and F1-score for identifying JIFs.

The task was presented to the LLM as a binary classification task: given a CJEU input judgment, the model was prompted to analyze each of its paragraphs and determine whether it presented the requisites to be considered a JIF. To enable the LLM to identify JIFs with high precision, we employed a few-shot prompting strategy. The prompt (Appendix A) itself was carefully constructed to serve a dual purpose: to formally define the concept of a JIF and to provide clear examples for the model to learn from. It began with a bullet-point list outlining the specific characteristics of a JIF, derived from our theoretical and empirical legal analysis. This was immediately followed by a series of three examples of confirmed JIFs, each accompanied by a brief explanation detailing the specific legal reasoning that made it a valid JIF. To ensure consistency and create a machine-readable output, the extraction process was implemented with a structured JSON output request. The structured output for each potential JIF included the paragraph number, a boolean field indicating if the paragraph was a JIF, and a detailed analysis that allows legal experts to assess why a paragraph labeled as JIF is classified as such.

For example, a successful extraction would produce an output with the following structure, as shown in the case of *A & G Fahrschul-Akademie GmbH v Finanzamt Wolfenbüttel* (Case C-449/17), in which the LLM correctly identifies paragraph 26 as a JIF.

Paragraph 26: “26 Consequently, the concept of ‘school or university education’ for the purposes of the VAT system refers generally to an integrated system for the transfer of knowledge and skills covering a wide and diversified set of subjects, and to the furthering and development of that knowledge and those skills by the pupils and students in the course of their progress and their specialisation in the various constituent stages of that system.”

```
"26": {
  "analysis": {
    "reasoning": "This paragraph is a JIF as it interprets the concept
of 'school or university education' by applying it to driving
tuition, stating that such specialized tuition does not amount to
the transfer of knowledge and skills covering a wide and diversified
set of subjects characteristic of school or university education.",
    "meets_JIF_criteria": true
  },
  "JIF": true
},...
```

While the models were instructed to provide a structured JSON output, some post-processing was required to ensure the data was consistently machine-readable. This was a critical step in handling minor deviations and ensuring the integrity of the dataset for subsequent analysis. The necessary corrections included:

- Removing any additional model output exceeding the intended JSON output.
- Sanitizing special characters that could compromise the parsing process.
- Standardizing the capitalization of boolean fields, as “true” and “false” were sometimes returned in varying styles.
- Fixing issues with inner quotations, a common feature of legal drafting that could disrupt the JSON structure.

This post-processing step ensured that the output was uniformly formatted, reliable, and ready for further quantitative and qualitative evaluation.

6. Evaluation and Results

The evaluation process of the results consisted of two phases. The first was quantitative, assessing the results on a small set of rulings. The second was qualitative, providing an error analysis to determine the strengths and weaknesses of automated versus manual extraction.

6.1. Quantitative Evaluation

The *a posteriori* quantitative evaluation of the task was performed by legal experts who were presented with 185 JIFs that had been extracted from a set of 10 randomly selected test judgments. This sample of 185 JIFs represented a subset of the 1,493 total alleged JIFs resulting from the entire extraction process. The experts were asked to determine whether each identified paragraph was correctly classified as a JIF. At the same time, the experts had access to the original judgment text and were tasked to also identify false negatives, that is JIFs that were missed by the model.

The manual evaluation of the model's performance on the test set yielded an average precision of 0.9189 and an average recall of 0.9444. The average F1 score was **0.9315**. This performance indicates the model has a satisfactory ability to correctly identify JIFs.

6.2. Error Analysis

The second phase of the evaluation consisted of a detailed error analysis performed by legal experts on the totality of the output instances. The errors took two main forms: false positives, where the system wrongly classified a statement as a JIF, and false negatives, where statements to be classified as JIFs were not extracted. By carefully reviewing these errors, the experts were able to classify them into distinct categories.

One type of error involved the LLMs misidentifying the source of a passage. For instance, though it happened rarely, the system would mistakenly include the arguments of the parties or observations of a national court that requested a preliminary ruling, which are often reported at the beginning of the CJEU's reasoning. In case C-240/05, paragraph 45, for instance, the system incorrectly selected a lengthy passage reflecting a Member State's position, even though such a passage only reproduces external submissions and does not constitute an authoritative formulation from the Court.

Similarly, when the CJEU decisions are preceded by an opinion of an Advocate General (AG), the LLMs sometimes disregarded passages where the Court explicitly declared it shared the AG's interpretative reasoning, probably treating them as external commentaries rather than authoritative JIFs.

Another issue was the misinterpretation of a statement's purpose. Sometimes, LLMs misidentified passages as JIFs when they lacked new interpretative reasoning, a mistake that arose in two common scenarios: when the Court restated its own prior points in the same decision or when it reproduced the wording of legislation. For instance, in case C-589/12, paragraph 36, the Court reiterated a limitation from a previous paragraph, while, in C-174/14, paragraph 40, the judgment recalled the content of Article 9 of Directive 2006/112. Although these passages offered no new interpretative value, their formal structure, which resembles that of interpretative statements, likely caused the LLMs to misclassify them.

A further difficulty arose in relation to citations of previous decisions. While the citation of prior case law often constitutes a JIF, this is not always the case. For example, when the Court refers to previous decisions merely to "distinguish" the current case from an earlier one, the statement is not a JIF. In some cases, LLMs failed to make this distinction, as seen in case C-612/20, paragraph 30, where a passage was incorrectly classified as a JIF even though the citation served only to highlight a difference from earlier case law.

Finally, in some cases, the LLMs struggled to correctly classify passages based on the scope and factual context of the legal reasoning. For example, the system sometimes incorrectly classified passages describing the specific factual circumstances of a case as JIFs. In case C-717/19, paragraph 42, a statement tied to transactions in pharmaceutical supply chains was extracted as a JIF even though it did not establish a broader interpretative principle. In contrast, the system sometimes failed to qualify as a JIF

those passages in which the CJEU delineates the boundary between its role of interpretation and the national court's task of applying that interpretation to a specific case. For instance, in case C-588/10, paragraph 36, the LLM failed to qualify a passage as a JIF, perhaps because the Court's statement was framed in terms of a factual verification to be carried out by the national court rather than as a direct interpretative rule.

The error analysis highlights the primary challenge for LLMs in identifying JIFs: discerning the purpose and contextual value of a statement beyond its surface-level linguistic structure. The models sometimes misclassify passages that, while superficially resembling a JIF, are either reproductions of external content, restatements of existing law, or fact-bound applications of a broader principle. Conversely, they can miss crucial JIFs when the interpretative value is framed less as a direct rule and more as a clarification of legal responsibility or an adoption of external reasoning.

While the detailed error analysis reveals room for improvement, its findings, when viewed alongside the positive results from the quantitative analysis, confirm the feasibility of automated extraction of JIFs. The presence of these specific error categories provides a clear roadmap for future refinements to enhance accuracy.

7. Conclusions

This study has successfully developed and implemented a human-centric LLM approach for the automated extraction of JIFs from EU case law. We established a formal definition of JIFs through a combination of theoretical and empirical analysis, which served as the foundation for an automated extraction task using LLMs. Our methodology proposes a novel, iterative process driven by detailed error analysis. This qualitative evaluation allowed legal experts to refine the initial JIF definition by assessing the model's output. The updated definition and gold standard significantly improved the performance metrics of the LLMs, with the Claude 3.7 Sonnet model achieving the highest F1-score of 0.9315 in a few-shot prompting strategy. The success of this approach demonstrates the potential of using AI not only for the extraction of JIFs but also as an integral tool for conceptualizing and formalizing abstract legal concepts.

In the future, this research could be expanded in several ways. The detailed error analysis might be used to further refine the definition of a JIF. This refined methodology could then be extended and applied to conceptualize, define, and extract other legal concepts. Moreover, the data we have already extracted might serve as a "silver set" for machine learning experiments that require large corpora, which would otherwise be too costly to create manually.

Additionally, future work could involve exploring prompting techniques to classify the extracted JIFs according to the VAT taxonomy. The data may also be used for other investigations, such as citation extraction and citation network analysis. Additionally, there is potential to expand this work to a multilingual dataset, which would allow for the discovery of connections between legal interpretation across different legal jurisdictions.

While not scalable to very large datasets, due to the use of a proprietary LLM, the use of one LLM call per judgment was deemed to be the approach that could grant the best tradeoff between the number of LLM calls (one per judgment instead of one per paragraph) while ensuring that no JIFs were left out due to context window limitations. Nevertheless, possible future work includes the exploration of open-source models to streamline the task and make the methodology affordable on larger datasets.

Acknowledgments

This work was partially supported by the following project: POLINE - Principles Of Law In National and European VAT (European Union's Justice Programme, G.A. No.: 101087342).

References

- [1] F.-X. Millet, In the name of analogy: Judicial copy-pasting and competence creep in the connection data case law, *Common Market Law Review* 61 (2024).
- [2] M. Jacob, *Precedent application by the ECJ*, Cambridge University Press, 2014, p. 87–126.
- [3] S. Garben, Competence creep revisited, *JCMS: Journal of Common Market Studies* 57 (2017). doi:10.1111/jcms.12643.
- [4] M. A. Pollack, Creeping competence: The expanding agenda of the european community¹, *Journal of Public Policy* 14 (1994) 95–145.
- [5] S. Weatherill, Competence creep and competence control, *Yearbook of European law* 23 (2004) 1.
- [6] L. Azoulai, La fabrication de la jurisprudence communautaire, Dans la Fabrique du Droit Européen, Brussels: Bruylant (2009) 153–170.
- [7] J. Valvoda, O. Ray, From Case Law to Ratio Decidendi, 2018, pp. 20–34. doi:10.1007/978-3-319-93794-6_2.
- [8] G. Sartor, P. Santin, D. Audrito, E. Sulis, L. D. Caro, Automated extraction and representation of citation network: A cjeu case-study, in: *International Conference on Conceptual Modeling*, 2022. URL: <https://api.semanticscholar.org/CorpusID:255547650>.
- [9] G. Governatori, A. Stranieri, Towards the application of association rules for defeasible rule discovery (2001).
- [10] D. Licari, P. Bushipaka, G. Marino, G. Comandé, T. Cucinotta, Legal holding extraction from italian case documents using italian-legal-bert text summarization, *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law* (2023). doi:10.1145/3594536.3595177.
- [11] M. Molinari, M. Quaranta, I. A. Amantea, G. Governatori, Using chatgpt to extract principles of law for the sake of prediction: an exploration conducted on italian judgments concerning lgbt (qia+) rights (2022).
- [12] D. Xu, W. Chen, W. Peng, C. Zhang, T. Xu, X. Zhao, X. Wu, Y. Zheng, E. Chen, Large language models for generative information extraction: A survey, *Frontiers Comput. Sci.* 18 (2023) 186357. URL: <https://api.semanticscholar.org/CorpusID:266690657>.
- [13] Z. Zhang, W. You, T. Wu, X. Wang, J. Li, M. Zhang, A survey of generative information extraction, in: *International Conference on Computational Linguistics*, 2025. URL: <https://api.semanticscholar.org/CorpusID:275821238>.
- [14] B. Padiu, R. Iacob, T. Rebedea, M. Dascalu, To what extent have llms reshaped the legal domain so far? a scoping literature review, *Information* 15 (2024) 662. doi:10.3390/info15110662.
- [15] J. Lai, W. Gan, J. Wu, Z. Qi, P. S. Yu, Large language models in law: A survey, *ArXiv abs/2312.03718* (2023). URL: <https://api.semanticscholar.org/CorpusID:266054920>.
- [16] M. Siino, Exploring the use of llms in the italian legal domain: A survey on recent applications, *Computer Law Security Review* 58 (2025). doi:10.1016/j.clsr.2025.106164.
- [17] Z. Sun, A short survey of viewing large language models in legal aspect, 2023. doi:10.48550/arXiv.2303.09136.
- [18] D. Audrito, I. Spada, R. Mignone, E. Sulis, L. Di Caro, Towards semi-automating european legislative harmonisation analysis: A harmonised glossary for llm-based legal concept detection, *Computer Law & Security Review* 58 (2025) 106171.
- [19] S. A. Reji, R. Sheik, S. A. A. R. M, S. Jaya Nirmala, Enhancing llm performance on legal textual entailment with few-shot cot-based rag, in: *2024 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, 2024, pp. 1–6. doi:10.1109/SPICES62143.2024.10779705.
- [20] V. Mishra, B. Pathiraja, M. Parmar, S. Chidananda, J. Srinivasa, G. Liu, A. Payani, C. Baral, Investigating the shortcomings of llms in step-by-step legal reasoning, in: *North American Chapter of the Association for Computational Linguistics*, 2025. URL: <https://api.semanticscholar.org/CorpusID:276250542>.
- [21] Z. Sun, Y. Shen, Q. Zhou, H. Zhang, Z. Chen, D. D. Cox, Y. Yang, C. Gan, Principle-driven self-alignment of language models from scratch with minimal human supervision, *ArXiv abs/2305.03047*

(2023). URL: <https://api.semanticscholar.org/CorpusID:258479665>.

[22] T. Lundmark, D. MacCormick, R. Summers, Interpreting precedents: A comparative study, *The American Journal of Comparative Law* 46 (1997) 211. doi:10.2307/841084.

A. Few-Shot Extraction Prompt

Find the Judicial Interpretative Formulas in the following judgment and output them in this format.

Each JIF corresponds to a paragraph.

Definition of a JIF:

A JIF is a statement within a CJEU decision where the Court interprets:

- A provision in a legislative text (e.g., European or national legislation).
- A legal concept or principle of law.
- An interpretative statement cited from another judicial decision.
- A statement made in an Opinion of an Advocate General of the Court of Justice of the EU.

Here are some JIF examples:

1. - JIF: As regards the refund of excess VAT under Article 183 of the VAT directive, it should be recalled that, as the Court has repeatedly held, the right to deduct provided for in Article 167 et seq. of that directive is an integral part of the VAT scheme and in principle may not be limited. In particular, the right to deduct is exercisable immediately in respect of all the taxes charged on transactions relating to inputs (see, inter alia, judgments of 21 June 2012, Mahagében and Dávid, C-80/11 and C-142/11, EU:C:2012:373, paragraph 38, and of 26 April 2018, Zabrus Siret, C-81/17, EU:C:2018:283, paragraph 33).
- reasoning: it refers to an interpretative statement made in previous judgments, but still relevant in the citing one
2. - JIF: However, as the Advocate General notes in point 79 of his Opinion, the amount paid in consideration for a transaction cannot constitute the consideration for another transaction, or even a payment on account in respect of the payment of the consideration for the other transaction.
- reasoning: the Court recognises the value of the interpretative statement made by the Advocate General and endorses it.
3. - JIF: In the light of the foregoing considerations, the answer to the second question referred is that Article 90(1) of Directive 2006/112 must be interpreted as meaning that, in the case of an agreement on payment in instalments, the fact that an instalment of the remuneration has not been paid before its term cannot be regarded as non-payment of the price, within the meaning of that provision, and, as a result, cannot lead to a reduction of the taxable amount.
- reasoning: It provides the interpretation of the whole content of art. 90 of the VAT Directive in the case of an agreement on payment in instalments.

****Output Format (JSON):****

```
“json
{
  [exact paragraph identifier]: {
    "analysis":{
      "reasoning": "text analysis and step-by-step full justification for JIF determination",
      "meets_JIF_criteria": [True | False]
    },
  },
}
```

```
"JIF": [True | False]
},...
"
```

Produce the output making sure that it is sanitized for JSON parsing.

judgment: