The volume offers an analysis of Judicial Interpretative Formulas, the recurrent interpretative statements that increasingly appear in EU and national case law, in the field of Value Added Tax (VAT). Combining legal theory, comparative analysis, and AI techniques, the book demonstrates how these formulas influence legal reasoning and foster dialogue between courts.
The work has been developed in the context of the EU-funded POLINE project and it introduces an interdisciplinary methodology for identifying, extracting, and organising Judicial Interpretative Formulas through large language models and natural language processing techniques.
The book offers a new perspective on the dialogue between legal systems regarding interpretation in the field of VAT. It may also be of interest to those who, through AI, wish to test the validity of a newly developed concept. In general, it offers interesting insights for anyone wishing to organise their knowledge of case law in complex and information-rich areas.

Piera Santin is a Research Associate in European tax law at the Robert Schuman Centre for Advanced Studies of the European University Institute. She has published extensively in the field of European tax law, with a particular focus on VAT, AI and tax law, and the use of quantitative methods in legal studies.

Alessia Fidelangeli is a Research Fellow at the University of Bologna (Department of Legal Studies) and Adjunct Professor in the Department of Business Economics. She holds a PhD in European tax law, with a dissertation concerning European VAT case law. She has published extensively in both European tax law and legal analytics applied to taxation.

Giuseppe Contissa is an Associate Professor in legal informatics and IT law at the University of Bologna. He has published extensively on artificial intelligence and law, as well as computable models of legal reasoning, including "Digital Technologies and the Law" (Torino, 2024).

Vito Santin, *Le ore e gli ori*,
tempera on paper

5 000002 737145
9 788813 397647
€ 39,00 I.V.A. INCLUSA

P. Santin, A. Fidelangeli, G. Contissa

A COMPUTATIONAL APPROACH TO VAT CASE LAW: AN ANALYSIS OF JUDICIAL INTERPRETATIVE FORMULAS

Wolters Kluwer

A computational approach to VAT case law: An analysis of Judicial Interpretative Formulas

edited by
Piera Santin, Alessia Fidelangeli, Giuseppe Contissa

Wolters Kluwer
CEDAM

P. Santin – A. Fidangeli – G. Contissa (edited by), *A computational approach to VAT case law: An analysis of Judicial Interpretative Formulas*

A computational approach to VAT case law: An analysis of Judicial Interpretative Formulas

edited by
Piera Santin, Alessia Fidelangeli, Giuseppe Contissa

Wolters Kluwer CEDAM

Funded by the European Union under the Justice Programme (JUST-2022-EJUSTICE), POLINE Principles Of Law In National and European VAT, Grant Agreement No. 101087342. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Printed by

GECA - Divisione Libri di CISCRA S.p.A., Via Belvedere, 42 - 20862 Arcore (MB)

Era necessario tornare a scegliere le parole, a scrutarle
per sentire se erano false o vere, se avevano o no vere radici in noi,
o se avevano soltanto le effimere radici della comune illusione.

Natalia Ginzburg, *Lessico Famigliare*

It was necessary for writers to go back and choose their words, scrutinise them
to see if they were false or real, if they had actual origins in our experience,
or if instead they only had the ephemeral origins of a shared illusion.

Natalia Ginzburg, *Family Lexicon*

Table of Contents

CHAPTER II

LEGAL INTERPRETATION IN ARITIFICIAL INTELLIGENCE AND LAW

Federico Galli – Giuseppe Contissa

CHAPTER III

THE ROLE OF EUROPEAN CASE LAW IN THE SYSTEM OF VALUE ADDED TAX

Eleonor Kristoffersson – Magnus Kristoffersson

CHAPTER IV

THE UNNAMED CONCEPT: PRESENT IN DISCOURSE, ABSENT IN NAME. JUDICIAL INTERPRETATIVE FORMULAS IN VAT CASE LAW

Piera Santin

Chapter V

"PRINCIPI DI DIRITTO" AND JUDICIAL INTERPRETATIVE FORMULAS: BETWEEN NATIONAL DOCTRINE AND EUROPEAN INTEGRATION

Alessia Fidelangeli – Andrea Mondini

CHAPTER VI

JUDICIAL INTERPRETATION AND LEGAL CERTAINTY: THE ROLE OF THE SUPREME ADMINISTRATIVE COURT IN BULGARIAN ADMINISTRATIVE JUSTICE

Dilyana Bozhanova – Lilia Kachoreva

CHAPTER VII

EU DIALOUGE AND VAT PRECEDENCE IN SWEDEN: OVERRULING AND DISTINGUSING IN THE SUPREME ADMINISTRATIVE COURT

Magnus Kristoffersson – Kevin Aiderfors – Eleonor Kristoffersson

Chapter XII

ACCESS TO KNOWLEDGE: STRUCTURE AND IMPORTANCE OF THE POLINE PILOT TOOL

Boycho Georgiev – Hristo Konstantinov – Vasil Oreshenski

INTRODUCTION

THE USE OF FORMULAS BY TAX COURTS: MYTH OR REALITY?

Piera Santin – Alessia Fidelangeli*

CONTENTS: 1. POLINE: From an idea, a project takes shape. – 1.1. A brief introduction. – 1.2. Extracting interpretation: The idea behind JIFs. – 1.3. The importance of classification: AI for a satisfactory retrieval system. – 2. The definition of Judicial Interpretative Formulas: Methodological remarks. – 2.1. The need to depart from common law concepts and approaches. – 2.2. From theory to practice: The need for a robust comparative framework before extraction. – 2.3. From JPOL to JIF: Towards a concept aligned with European and national frameworks. – 3. Conclusion and overview of the book structure.

1. POLINE: From an idea, a project takes shape

1.1. A brief introduction

This project pursues two interconnected objectives involving the use of Artificial Intelligence (AI) technologies: first, to empirically demonstrate the existence of the specific legal concept of *Judicial Interpretative Formulas* (see par. 2.3.) through their automatic extraction; and second, to organise legal knowledge in a way that enhances its accessibility, particularly for legal professionals with limited IT expertise.

The project was designed to be innovative in both its core

* Piera Santin is Research Associate at the Robert Schuman Centre for Advanced Studies of the European University Institute; Alessia Fidelangeli is Research Fellow at the University of Bologna.

domains: legal analysis and computer science. To achieve this, a robust interdisciplinary methodology was developed, which stands as one of the most significant results of our work.

From the outset, we made a deliberate choice to focus on Value Added Tax (VAT), as it represents one of the most harmonised areas of EU law. Over the years, case law from the Court of Justice of the European Union (CJEU) has played a pivotal role in fostering such harmonisation (see Ch. IV). VAT is also the field with the highest number of preliminary rulings, making it a particularly fertile ground for identifying interpretative patterns and verifying their consistency across jurisdictions.

Within the VAT domain, we further narrowed our focus to judgments concerning exemptions for public interest and the definition of taxable amount. This thematic constraint allowed us to manage computational costs effectively, while still ensuring that the methodology is scalable and applicable to broader areas of case law in future developments.

A brief account of the POLINE project's development is essential, as its evolution has directly influenced the project's design. When the proposal was first drafted, Large Language Models (LLMs) had not yet become prominent in public or academic discourse. At that time, the collaboration between legal scholars and data scientists focused on applying other Natural Language Processing (NLP) techniques to identify relevant portions of legal texts, primarily through manual annotation (see Ch. X).

We subsequently adopted both technologies for complementary purposes. LLMs were employed for the initial extraction of JIFs and for linking these with a structured legal ontology (see Ch. IX and XI). Later, NLP was used to explore an alternative method for JIF extraction from new decisions (see Ch. X), this was necessary to implement the automatic detection module of the platform (see Ch. XII).

Our second goal was to build a tool that improves the usability of AI-extracted legal knowledge. Legal research tools often yield unsatisfactory results, largely due to poor classification and retrieval systems. To address this, we created a structured taxonomy as the foundation for organising the automatically extracted information and employed AI techniques to classify each JIF accordingly (see Ch. XI).

1.2. Extracting interpretation: The idea behind JIFs

To clarify the focus of our knowledge extraction process, we must first introduce the concept of JIFs. The starting point for this work was the observation that the Court of Justice of the European Union (CJEU) frequently uses recurring interpretative statements, typically corresponding to a single paragraph, which are reproduced verbatim across multiple decisions. Over time, this practice has generated a corpus of interpretative rules that has significantly influenced national case law.

As discussed in Paragraph 2, although many legal scholars and practitioners intuitively recognise this phenomenon, it remains underexplored in the literature, in part because its existence is difficult to demonstrate through purely qualitative methods. The repetitive and formulaic nature of the phenomenon demands empirical validation.

The POLINE project was initiated to identify the features that characterise these formulas, with the aim of enabling their automatic detection through AI. This involved both theoretical analysis and empirical research, which informed the drafting of detailed annotation guidelines (see par. 2.2.). These guidelines were used for manual tagging of selected judgements (training the NLP models, see Ch. X), and for creating prompts used in LLM-based experiments (see Ch. IX).

Our hypothesis is that the relevance of JIFs extends beyond the EU legal system, impacting national legal systems, especially in harmonised fields of law as VAT. To test this, we included three national jurisdictions in our study (Bulgaria, Italy, and Sweden) following established methods of comparative legal analysis (see Ch. V, VI and VII). Legal-theoretical and empirical research was conducted across the EU and the three national legal systems, allowing us to examine whether similar formulaic interpretative practices could be identified in national case law.

For this comparative work, we adapted the original EU guidelines to fit each national context. This ensured both methodological consistency across datasets and sensitivity to the unique characteristics of each legal tradition.

1.3. The importance of classification: AI for a satisfactory retrieval system

The second major challenge addressed in this project concerns the accessibility of JIFs after their extraction. Beyond simply compiling these interpretative statements and enabling Boolean search functionalities, we placed particular emphasis on their systematic classification according to legal content and domain relevance.

Current legal databases suffer from two persistent limitations: they typically consist of full-text judgments or, less frequently, summaries of entire decisions, and they rarely isolate or highlight key interpretative passages. Moreover, even at the level of full judgments, classification systems tend to be inconsistent, leading to ineffective retrieval mechanisms. Legal professionals often face time-consuming searches that produce limited or imprecise results, especially when seeking specific legal reasoning or formulaic interpretations.

In addressing this gap, we observed that classification is not merely a technical convenience, but a structural necessity in enabling meaningful access to legal knowledge. During our analysis of CJEU case law (see Ch. IV), we noted a recurring practice: judges frequently reference precedents from domains outside the immediate subject matter of the case. This happens to address procedural issues, to invoke general principles (e.g. proportionality), or to bolster reasoning with analogical references. This interpretative practice suggests that access to JIFs, independently from the full text of judgments, could be highly beneficial for practitioners engaged in the application of EU law.

To make such access possible, we developed a legal ontology grounded in the EuroVoc classification system, as adopted by EUR-Lex, and enriched it with domain-specific keywords derived from the case law corpus. However, the massive and consistent classification of extracted JIFs (amounting to several thousand distinct entries) would not have been feasible without the application of LLMs.

Thanks to the inferential and semantic capabilities of LLMs, we were able to assign each JIF a precise label within the ontology, reflecting not only its linguistic content but also its legal substance. This process allowed us to maintain a high level of internal consistency, a key requirement for any

retrieval system intended to support structured and comparative legal analysis.

In practical terms, this means that users of the platform can conduct searches not only through Boolean expressions or keywords but also by navigating through an ontology-based classification interface. Each JIF is accessible through thematic categories, and users can move seamlessly from one formula to others within the same classification group. Additionally, the platform allows results to be filtered according to jurisdiction enabling both targeted research and broader comparative exploration in the EU, Bulgarian, Italian, and Swedish legal systems.

Ultimately, the integration of LLMs into the classification process represents an important aspect of the POLINE project. It transforms unstructured legal language into structured legal knowledge, making it possible to retrieve interpretative content with unprecedented precision and coherence. This advancement is crucial for our long-term goal: enhancing legal information systems in a way that supports transparency, consistency, and accessibility across legal systems.

2. The definition of Judicial Interpretative Formulas: Methodological remarks

The methodology underlying this project is complex, as POLINE included different activities: the comparative legal analysis of case law in the field of VAT; the extraction of JIFs; the creation of a multilingual ontology; the association of JIFs with that ontology; the extraction of citations to earlier case law contained in JIFs; network analysis between JIFs; and the examination of similarities between different JIFs. Each of these works followed its own methodology, which will be discussed in detail in the chapters devoted to each topic.

In this introductory chapter, our focus is on several methodological considerations relating to the definition of the concept of JIF (central to the project) rather than on the specific activities undertaken.

Drawing on our prior experience on AI and law, we identified three guiding priorities.

First, because the project concerned civil law jurisdictions, we aimed to develop a novel conceptual framework for identifying relevant interpretative statements. This framework

departed from approaches developed in common law systems, particularly those prevalent in the United States (see par. 2.1.), adapting them to the features of civil law legal systems.

Second, before starting the extraction through LLMs and NLP, experts in tax law and legal theory carried out an in-depth theoretical and empirical analysis concerning the features of national case law in the tax domain. This preliminary work required considerable time and amounted to the theoretical foundation for the subsequent stages of the project (see par. 2.2.).

Third, lawyers were involved in the preparation, execution, and validation of every phase of the project, and they revised the underlying assumptions guiding activities where necessary (see par. 2.3.).

2.1. The need to depart from common law concepts and approaches

A central challenge of the project concerned the definition of the object of our analysis, i.e. what, precisely, we sought to extract. Our experience in analysing both national and European case law made us certain that both jurisprudences make use of interpretative statements concerning the interpretation and application of VAT legislation. Extracting and classifying these statements could be very useful to interpreters for many purposes: granting legal certainty; investigating the coherence between national and European case law; examining the correctness of citations of previous case law; etc. The existence of these statements is mentioned here and there in literature (from the perspective of the doctrine of precedent; to criticise the creative role of case law; to show that the excessive workload of supreme courts). However, these statements are not analysed in themselves and, especially, are not given a name. This difficulty is reminiscent of a long-standing problem in American jurisprudence, famously captured in the observation that "dictum is one of the commonest yet least discussed of legal concepts. Every lawyer thinks he knows what it means, yet few lawyers think much more about it".

A bibliographic analysis made it clear that there were well-developed conceptual tools within the common law tradition for identifying and categorising judicial interpretative statements, but none of these frameworks could be transposed without

distortion into the civil law context in which our work was situated.

In the United States, the debate has long revolved around the distinction between *holding* and *dictum* [2]. A *holding* is generally defined as that part of a judicial opinion which is "necessary" to the outcome of the case, whereas *dictum* is any statement in a judicial opinion that does not form part of the holding [3]. Yet, as the literature repeatedly acknowledges, this distinction is far more elusive in practice than such concise definitions suggest. For nearly a century, American lawyers, judges, and scholars have debated where the boundary lies, and what practical consequences should follow from it [4]. The intensity of the *holding/dicta* debate in common law jurisdictions stems from the structural role of *stare decisis*: in a system where the holdings of higher courts bind future decisions, accurately identifying the "holding" is essential, as it determines the scope of a decision's precedential effect. Indeed, there is an existing literature focusing on the possibility of their automatic extraction through LLM [5].

By contrast, the interpretative statements we initially sought to identify are not embedded within a doctrine of binding precedent. The CJEU does not operate under a strict *stare decisis* regime, although it accords substantial weight to its earlier decisions, frequently citing them verbatim through what might be termed a "copy-paste" technique, and never expressly admitting to overruling itself [6]. National courts, too,

---

[2] A very famous work is *Dictum revisited*, in *Stanford Law Review*, 1952, p. 509 ff. See also M. ABRAMOWICZ, M. STEARNS, *Defining Dicta*, in *Stanford Law Review*, 2005, p. 953 ff.; L. ALEXANDER, *Constrained by Precedent*, in *Southern California Law Review*, 1989, p. 1 ff.; H.J. FRIENDLY, *In Praise of Erie - and of the New Federal Common Law*, in *New York University Law Review*, 1964, p. 383 ff.; A. GOODHART, *Determining the Ratio Decidendi of a Case*, in *Yale Law Journal*, 1930, p. 161 ff.

[3] K. GREENAWALT, *Reflections on Holding and Dictum*, in *Journal of Legal Education*, 1989, p. 431 ff.

[4] For a reconstruction of the debate, see J.M. STINSON, *Why Dicta Becomes Holding and Why it Matters*, in *Brooklyn Law Review*, 2010, p. 219 ff.

[5] C. ARVIN, *Identifying Legal Holdings with LLMs: A Systematic Study of Performance, Scale, and Memorization*, arXiv preprint – arXiv:2505.02172, 2025.

[6] F.X. MILLET, *In the name of analogy: Judicial copy-pasting and*

increasingly cite prior decisions, including those of the CJEU, especially in harmonised areas of law (see Ch. IV).

Nonetheless, the absence of a binding precedent rule in these jurisdictions means that transplanting [7] the common law *holding/dicta* dichotomy into our context would have been misguided. For this reason, we developed a new conceptual category, initially termed *Judicial Principles of Law* (JPOLs), inspired by the Italian doctrinal notion of *principio di diritto* as applied to the Italian Supreme Court (see Ch. V). As will be explained in the following section, we later replaced this terminology with *Judicial Interpretative Formulas* (JIFs), a conventional concept with a more explicitly European origin and a closer fit with the aims of our project.

Yet one feature is common to both the *holding/dicta* debate and the JPOL/JIF methodology and has been highlighted in American scholarship [8]. In earlier decades, legal research typically began with index searches, followed by reading entire cases to determine whether they addressed the relevant legal issue. Contemporary legal research, however, is mediated by electronic databases, which deliver search results by pinpointing keywords or phrases within opinions. This allows researchers to find an isolated sentence that appears to support their position without necessarily reading the remainder of the judgment to understand the significance of the quoted statement in relation to the decision as a whole.

The same issue arises in the context of JIFs: an isolated judicial statement may seem authoritative when read in isolation, but its actual relevance can only be determined in light of the decision. This is why, in the POLINE platform, for each JIF we always add the link to the whole decision.

2.2. From theory to practice: The need for a robust comparative framework before extraction

Too often in legal analytics projects, the practical phase,

---

competence creep in the connection data case law, in *Common Market Law Review*, 2024, p. 1289 ff.

[7] On legal transplants see A. WATSON, *Legal Transplants: An Approach to Comparative Law*, University of Georgia Press, 1974.

[8] T. FOWLER, *Holding, Dictum... Whatever*, in *North Carolina Central Law Journal*, 2003, p. 139 ff.

particularly the use of LLMs for data extraction, commences without a sufficiently precise definition of the object of study. In our case, a wide was essential to determine, from the very beginning, the precise target of extraction.

The theoretical analysis consisted in a detailed examination of the national legal systems of the participating partners. Each partner was tasked with preparing a study of their domestic tax law system, including a concise description of which courts have jurisdiction over the merits of tax cases, the identity of the court of last instance, whether that court is administrative or judicial in nature, and whether it may decide the merits of a case or is confined to reviewing questions of law. They were also required to describe the structure, content, and decision-making style of the tax judgments of their supreme courts [9].

Given the absence of a uniform concept of relevant interpretative statements in case law across jurisdictions, the core of this analysis was a report on the existence (or absence) of the concept of a "judicial principle of law" in each national system. This inquiry was inspired by the Italian concept of *principio di diritto*, a doctrinally notion linked to the nomophylactic function of the Italian Supreme Court (see Ch. V). Partners were asked to determine whether analogous concepts existed in their jurisdictions, whether public databases containing such principles (or the *rationes decidendi* of supreme courts) were available, and how such resources were used by tax judges and practitioners.

This work also addressed the ongoing debate over whether the CJEU functions more like a common law or a civil law court, particularly in terms of the weight accorded to previous decisions. Partners were therefore asked to describe the role of precedent in their national tax case law. They were also asked to report on any national debate concerning the creative role of judges in tax law, especially where judge-made law might raise constitutional concerns in light of the separation of powers or the principle, present in some systems, that tax legislation must have a statutory basis.

In addition, partners were asked to provide information on

[9] Taking inspiration, as to the strcture, from the famous work performed by N. MACCORMICK, N.S. SUMMERS, *Interpreting precedents. A comparative study*, Routledge, 1997.

the role of case law in the correct interpretation of national tax law and the role of CJEU and national courts in ensuring uniform application of EU law in the VAT field. Particular attention was paid to whether national supreme courts cite CJEU judgments directly and to the frequency and manner of such citations.

The outcomes of this research amount to the basis of the analysis presented in Ch.s IV, V, and VI.

The second stage was an empirical analysis in which partners examined judgments in a selected subdomain, identifying examples of JIFs and analysing structural features of decisions, such as whether they were divided into paragraphs and whether interpretative statements appeared within single or multiple paragraphs. The empirical analysis also assessed the frequency and uniformity of citations to precedent, both national and EU, and determined whether the identified JIFs concerned VAT specifically or addressed procedural or other issues. Further inquiries examined the level of abstraction of JIFs, the presence of factual references, and the frequency with which courts overruled or distinguished prior JIFs. Finally, partners reported on whether such statements were explicitly labelled as judicial principles within their domestic systems, albeit under different terminology.

Only after these theoretical and comparative analysis we defined the concept of what we called JIF. We defined a JIF as a portion of text, extracted from the argumentative part of a judgement which contains the:

- interpretation of a rule, of the portion of a rule, or of a general principle; or
- consequences stemming from the interpretation/application of a rule or a principle in a legal system; or
- subsumption of a fact within a rule; or
- qualification of a factual hypothesis as a concept contained within a rule.

This definition emerged from the collective observations of the partners and was formulated jointly by experts in tax law and legal theory. The definition was validated by all partners and subsequently presented to the supreme courts of the participating states during testing events. On the basis of these findings, guidelines for the manual annotation of JPOLs were drafted. Only then did we proceed to the technical phases of

the project: the extraction of JIFs, the construction of the ontology, the network analysis, and the related computational activities.

2.3. From JPOL to JIF: Towards a concept aligned with European and national frameworks

In the initial stages of the project, we employed the term *Judicial Principle of Law* to designate the category of interpretative statements we sought to identify. However, it soon became apparent that this terminology risked confusion. In particular, it could be mistaken for the broader legal notion of a "principle of law" in the sense of legal principles such as legal certainty, subsidiarity, or proportionality. This would have been conceptually problematic, as our focus was on interpretative statements emerging from judicial reasoning.

As the project progressed, the analysis of both national and European case law revealed an important characteristic of VAT-related case law: the recurrence of identical or near-identical textual passages across different decisions. This phenomenon is striking in EU case law, but it is also increasingly present in national decisions, particularly recent ones. National courts, especially in tax matters, frequently reproduce interpretative statements verbatim over time, and in many cases, they quote European case law directly and literally (see Ch. VIII).

The relevance of this observation became clearer considering scholarly debates on the so-called "competence creep" of the European Union [10]. This is the gradual and sometimes opaque expansion of EU competences which can be discerned in the case law of the CJEU. The VAT field provides a clear example: both European and national legal literature acknowledge the role of CJEU case law in driving harmonisation, and virtually all scholarly works on VAT devote attention to judicial decisions (see Ch. III). The huge volume of VAT-related case law has even prompted an

[10] S. Garben, *Competence Creep Revisited*, in *Journal of Common Market Studies*, 2017, p. 1 ff.; M.A. Pollack, *Creeping Competence: The Expanding Agenda of the European Community*, in *Journal of Public Policy*, 1994, p. 95 ff.; S. Weatherill, *Competence Creep and Competence Control*, in *Yearbook of European Law*, 2004, p. 1 ff.

institutional shift, with jurisdiction over certain VAT matters being transferred from the CJEU to the General Court [11].

Another well-documented feature of CJEU jurisprudence is its distinctive drafting style, frequently described as "copy-pasting" (see par. 2.2.). This practice, which emerged in the late 1970s as the length of the Court's judgments increased and a mass of prior case law became available for citation, involves reproducing substantial passages from earlier judgments, verbatim or with slight modifications, into later ones, irrespective of their precise legal status (e.g. *ratio decidendi* or *obiter dictum*) [12]. In some judgments, block quotations from earlier cases dominate the reasoning.

Legal scholarship has examined a specific aspect of this practice of reproducing literally earlier binding interpretative statements in subsequent case law using the concept of "formulas" [13]. This stylistic approach, now pervasive in CJEU decisions [14], is beginning to appear in national case law as well, particularly in courts facing heavy caseloads, such as the Italian Court of Cassation (see Ch. V).

In light of these considerations, we concluded that the concept of "formula" more accurately captured the phenomenon we intended to analyse. We therefore adopted the term *Judicial Interpretative Formula* in place of *Judicial Principle of Law*. In a harmonised field such as VAT, it is perhaps inevitable that concepts originating within the European legal system are better suited to the analytical task than those grounded exclusively in domestic traditions. Consistent with the existing literature, we chose to extract all

---

[11] Regulation (EU, Euratom) 2024/2019 of the European Parliament and of the Council of 11 April 2024 amending Protocol No 3 on the Statute of the Court of Justice of the European Union.

[12] F.X. MILLET, *In the name of analogy*, cit., p. 1291.

[13] L. AZOULAI, *La fabrication de la jurisprudence communautaire*, in P. MBONGO, A. VAUCHEZ (eds.), *Dans la fabrique du droit européen. Scènes, acteurs et publics de la Cour de justice des Communautés européennes*, Bruylant, 2009, p. 153 ff.

[14] On the linguistic–semiotic perspective regarding the drafting style of the Court of Justice and its impact on the development of EU case law, see K. MCAULIFFE, *Hybrid Texts and Uniform Law? The Multilingual Case Law of the Court of Justice of the European Union*, in *International Journal for the Semiotics of Law*, 2011, available at SSRN: https://ssrn.com/abstract=2208686 (last access: September 1st, 2025).

JIFs, including those not constituting the *ratio decidendi* of the case.

3. Conclusion and overview of the book's structure

The project shows how combining comparative legal studies and NLP techniques allows identification and extraction of JIFs. Even though often unnoticed by doctrine, they play a structural function in the transmission of legal meaning across courts and jurisdictions. Their recognition and classification, therefore, are not merely technical achievements but epistemological steps toward making the architecture of judicial interpretation empirically observable. The following chapters develop this premise, each addressing a specific dimension of the relationship between interpretation, technology, and the law.

Ch. I traces the historical trajectory of artificial intelligence in judicial activities through the lens of legal informatics. It reconstructs how successive computational approaches have interacted with legal reasoning. The chapter shows that far from being an external instrument, AI has evolved within the conceptual and procedural structures of the law itself.

Ch. II explores the modification of legal interpretation in the age of AI. It examines how interpretative reasoning continues to support legal practice as machine-assisted systems increasingly support decision-making.

Ch. III introduces the legal and institutional context of VAT law in the European Union, explaining how the CJEU has progressively shaped the system through its interpretative activity. The chapter provides the doctrinal foundation for understanding why VAT constitutes an ideal field for the study of JIFs.

Ch. IV offers a diachronic analysis of JIFs within VAT case law. It examines how the drafting style of the CJEU and the practice of citation and repetition have contributed to the emergence of interpretative formulas as building blocks of the *acquis communautaire*. The chapter also discusses the institutional implications of the 2024 reform transferring VAT preliminary rulings to the General Court, which may further consolidate the circulation of such formulas.

Ch. V focuses on the Italian legal system, analysing the

interaction between the newly defined JIFs and the traditional notion of *principio di diritto* elaborated by the Court of Cassation. It explores whether the extraction and classification of interpretative formulas may support greater coherence in national case law and contribute to the broader debate on the creative function of the judiciary.

Ch. VI examines the Bulgarian context, where the Supreme Administrative Court plays a crucial role in ensuring uniform interpretation of tax law. The chapter analyses interpretative judgments in VAT matters as expressions of judicial reasoning and legal certainty, while also addressing the national discussion on the role of judges in shaping administrative and fiscal law.

Ch. VII investigates the Swedish Supreme Administrative Court's approach to interpretative stability and doctrinal change in VAT jurisprudence. Drawing on a comprehensive corpus of judgments, it analyses the techniques of overruling and distinguishing and shows how interpretative authority is maintained in a system without formal *stare decisis*.

Ch. VIII explores the concept of judicial interaction and its analytical connection with JIFs. It demonstrates how the mapping of JIFs across jurisdictions show the patterns of dialogue and mutual influence among European courts. The chapter also identifies possible applications of JIF-based analysis for comparative law, judicial cooperation, and the study of transnational reasoning.

Ch. IX describes the creation of the multilingual dataset underpinning the entire project. Curated by experts in tax law, it compiles VAT-related decisions from the CJEU and national courts in Italy, Sweden, and Bulgaria, thus forming the empirical basis for the extraction and classification experiments.

Ch. X presents the methodological and technical foundations of the machine-learning framework developed for the automatic extraction of JIFs. It critically reviews earlier approaches to identifying legal principles and explains how new models based on large language technologies can balance automation with transparency and expert validation.

Ch. XI relies on that framework to construct a multilingual ontology of VAT law and to perform the classification and network analysis of extracted JIFs. By combining LLM-based inference with ontology engineering, the chapter illustrates

how interpretative formulas can be connected, compared, and studied as interlinked nodes in a semantic and citation network.

Ch. XII concludes the volume by presenting the POLINE Pilot Tool, the practical implementation of the project's research. The tool integrates NLP, ML, and ontology-based reasoning to provide a transparent, multilingual, and user-oriented platform for exploring interpretative reasoning in tax law. The chapter discusses its potential to improve consistency in judicial decision-making, foster transparency in the use of AI within courts, and promote more equitable access to legal information across jurisdictions.

This book brings together theoretical, comparative, and computational perspectives to examine how interpretation can be improved by tools developed using AI. It may contribute to a new branch of empirical legal research and to a deeper understanding of how legal meaning circulates within the European judicial space.

CHAPTER I

ARTIFICIAL INTELLIGENCE FOR TAX COURTS

Marco Billi – Federico Galli – Giuseppe Contissa*

CONTENTS: 1. Introduction. – 2. The Emergence of AI in Courts. – 3. Information Retrieval for Tax Case Law. – 4. Tax Expert Systems. – 5. Machine Learning for Tax Cases. – 6. Large Language Models and Tax Case Law. – 7. Conclusion.

1. Introduction

In this chapter, we explore the evolution of Artificial Intelligence technologies in court activities and, in particular, in judicial decision-making. We will do so through the lens of legal informatics, which is the interdisciplinary field that combines law, computer science, and information technology to enhance the understanding, analysis, and application of legal principles and practices. Our aim is twofold: first, to retrace the different contributions of legal informatics for building AI applications in the judiciary; second, to demonstrate that legal informatics is deeply rooted in legal principles, drawing on the content, processes, concepts, and theories of law.

We will trace the historical trajectory of AI in courts, analysing how different computational approaches have shaped, and been shaped, by legal reasoning. This exploration will reveal the ways in which the law has been modelled computationally, and how developers and researchers have

* Marco Billi is a Research Fellow at the University of Bologna; Federico Galli is Assistant Professor in Legal informatics and Computer Law at the University of Bologna; Giuseppe Contissa is Associate Professor in Legal informatics and Computer Law at the University of Bologna.

approached legal problems through the use of AI technologies in court settings. The journey through the history of AI in courts – from the digitisation of legal sources to the advent of generative AI – highlights the dynamic and reciprocal relationship between legal practice and technological innovation.

The perspective from which this analysis proceeds can be encapsulated in the insights of Gottfried Wilhelm Leibniz, a scholar who uniquely combined law, science, and philosophy. Leibniz observed that while mathematicians excelled in the art of reasoning about the necessary, jurists excelled in reasoning about the contingent, offering lessons on proofs, presumptions, and the interpretation of legal texts. His view underscores the bidirectional learning process between law and computing: the law adapts and learns from computational advancements, while computing draws valuable models and concepts from legal reasoning [2].

This reciprocal learning process is further emphasised by Thorne McCarty and Edwina Rissland, prominent figures in AI and law, who noted that the facets of human reasoning most studied by AI are also central to legal reasoning [3]. In law, computer scientists encounter diverse modes of reasoning – deductive, inductive, and analogical – applied to real or hypothetical cases, rules, and both structured and unstructured texts. This makes the legal domain a fertile ground for cognitive scientists and AI researchers, where the complexities of human and artificial intelligence converge and mutually inform each other.

The integration of AI into courts not only reflects this intersection but also expands the boundaries of both fields. As we examine four developments of AI in judicial settings, we will highlight how each phase – beginning with the digitisation of legal sources, progressing through rule-based and data-driven models, and culminating in the current era of generative AI and Large language models (LLMs) – has contributed to reshaping the way judges access, interpret, and apply the law.

---

[2] G. LEIBNIZ, *De Legum Interpretatione, Rationibus, Applicatione, Systemate*, in *Sämtliche Schriften Und Briefe*, 1923.

[3] L.T. MCCARTHY, E.L. RISSALAND, *An Artificial Intelligence Approach to Legal Reasoning*, in A. LIETH GARDNER (ed.), *An Artificial Intelligence Approach to Legal Reasoning*, MIT press, 1987.

2. The emergence of AI in courts

The emergence of AI in courts is part of a broader historical continuum in which law and technology have progressively intertwined, transforming how legal content is created, expressed, and applied. Throughout history, technological advancements have reshaped the mediums in which law resides, influencing the ways in which legal knowledge is preserved, transmitted, and enforced. From the earliest forms of unwritten law, which existed as social norms, customs, and spoken traditions, to the advent of handwriting and printing that allowed legal content to be inscribed and widely disseminated, each technological shift has expanded the reach and impact of the law.

Before computational technologies, legal content relied on human cognition – memory, language, and reasoning – for its interpretation and application. Handwriting first enabled the externalisation of legal texts onto physical artefacts, such as tablets, papyri, and parchment, which allowed legal rules, judgments, and contracts to be stored and transmitted across time and space [4]. Printing technology further revolutionised legal communication by making legal texts accessible to a broader audience, thus fostering the growth of complex legal systems and large social organisations. However, the content inscribed on these materials remained passive, requiring human engagement to interpret and enforce the law's mandates.

The rise of computational technologies introduced a fundamental change, transforming legal texts from static records into dynamic entities that could be automatically processed and applied. This shift marks the advent of computable law, where legal content not only resides in external forms but also interacts dynamically with computational systems. The ability to store, retrieve, and manipulate legal texts electronically opened new avenues for the automation of legal processes, enabling legal rules to be applied semi-automatically or even fully automatically in certain contexts. This transformation laid the groundwork for the integration of AI into courts,

[4] On the influence on printing and other information technologies on the law, see M. HILDEBRANDT, *Smart Technologies and the End (s) of Law: Novel Entanglements of Law and Technology*, Cheltenham, 2015.

fundamentally altering how judges and legal practitioners engage with the law.

The convergence of advanced computing power, vast datasets, instant communication, and novel AI technologies has catalysed the evolution of AI applications in judicial settings, giving rise to what we refer to in this Chapter as "four stages of AI in courts". Each stage represents a distinct phase in automating parts of judicial activities, reflecting different approaches to how legal knowledge is modelled, interpreted, and applied.

The first applications appeared with the digitisation of legal sources, making legal content more accessible and searchable, thus supporting judges in the initial stages of legal research. The second step introduced rule-based systems, where human-crafted logical models of legal reasoning were developed to automate aspects of judicial decision-making. The third step shifted towards data-driven models, employing machine learning techniques to analyse vast legal datasets and make predictions about case outcomes. The fourth step, characterised by the emergence of large language models, introduces new possibilities for the dynamic creation of legal texts, arguments, and judicial analyses, pushing the boundaries of how AI can assist in the judicial process.

In the following sections, we will explore these four steps in detail, examining how each phase has contributed to the evolving role of AI in courts. We will trace the historical trajectory of these technological developments and assess their impact on judicial decision-making. Through this analysis, we will gain a deeper understanding of the ongoing dialogue between law and computing and the ways in which AI continues to shape the judicial landscape.

In the early stages of exploring the potential of artificial intelligence within legal contexts, McCarty identified corporate tax law as a particularly suitable domain for experimentation [5]. He argued that this area of law, characterised by multiple layers of commercial abstraction, represents a highly formalised system largely detached from the practical realities of everyday life. Corporate tax law, he noted, relies heavily on concepts and

[5] L.T. MCCARTY, *Some Requirements for a Computer-Based Legal Consultant*, in *AAAI* 1, 1980, p. 298 ff.

constructs developed purely for legal purposes, rendering it more artificial in nature compared to areas such as civil or criminal law, which are deeply rooted in ordinary human experience, dealing with matters such as birth, marriage, or inheritance.

McCarty observed that this distinction bears directly on the challenges of applying AI in legal reasoning. Simpler legal problems, often encountered by first-year law students, tend to be the most difficult for AI systems precisely because they depend on a foundation of common human experience, an element inherently absent in computational systems. Conversely, the technical and highly formalised character of tax law makes it more amenable to algorithmic analysis. Its complexity, often daunting even for trained legal professionals, provides an environment in which AI can offer meaningful assistance by navigating intricate statutory and regulatory frameworks with precision and consistency.

3. Information retrieval for tax case law

The first application of AI in courts emerged in the 1960s and focused on making legal sources accessible in a digital format. This development, initially centred on legal information retrieval, has significantly transformed how courts and judges operate. Before this shift, courts relied heavily on physical legal texts and paper documents, which were cumbersome, time-consuming, and prone to inconsistencies in accessibility. The introduction of digitised legal sources revolutionised the legal landscape, particularly within judicial settings, by offering quicker, more reliable access to legal information.

During the early stages, legal information had to be manually digitised from paper sources through human encoding or scanning technologies. This process was often slow and resource-intensive, especially in court environments where timely access to legal precedents and statutes is crucial[6]. As technology progressed, more legal content became natively digital, allowing courts to streamline the

[6] S. SIMITIS, *Informationskrise Des Rechts Und Datenverarbeitung*, C.F. Müller, 1970.

initial acquisition and processing of legal documents. These documents are frequently enriched with metadata, structured according to standard formats, and integrated into searchable databases, facilitating efficient retrieval during court proceedings.

The application of information retrieval techniques in judicial contexts has enhanced the decision-making process by providing judges with rapid access to a comprehensive body of legal materials [7]. Search engines within court systems allowed various retrieval methods – Boolean, statistical, and conceptual – to deliver tailored results based on specific queries. Advanced functionalities like relevance ranking, citation analysis, and semi-automated summarisation enabled judges to quickly pinpoint the most pertinent legal texts, reducing time spent on manual research and allowing for more focused deliberations.

Compared with today's AI-powered legal search engines, computational techniques in this phase were highly controllable but limited in scope and adaptability. Early information retrieval systems primarily relied on deterministic algorithms that required explicit search queries, meaning judges and legal professionals needed to have a precise understanding of the terms and structure of the legal information they were seeking. While Boolean search methods allowed for some flexibility, they often resulted in either too many irrelevant results or missed pertinent information due to the rigid nature of the query structure.

Despite these limitations, it became apparent that the design and functionality of these systems could significantly influence judicial reasoning during case evaluations [8]. Low performance in information retrieval, such as in the case of missing relevant information (low recall) or retrieving excessively irrelevant information (low precision), can lead to the exclusion of important precedents or pertinent statutes, inadvertently

---

[7] See, e.g., the pioneering work by L.E. ALLEN, *Beyond Document Retrieval Toward Information Retrieval*, in *Minnesota Law Review*, 1962, p. 713; C.D. HAFNER, *Representation of Knowledge in a Legal Information Retrieval System*, in *Proceedings of the 3rd Annual ACM Conference on Research and Development in Information Retrieval*, 1980, p. 139 ff.

[8] D.P. DABNEY, *The Curse of Thamus: An Analysis of Full-Text Legal Document Retrieval*, in *Law Library Journal (LLJ)*, 1986, p. 5.

filtering out essential information needed for comprehensive legal analysis. Moreover, the ranking of search results can subtly affect judicial judgments, as documents that appear higher in the list are more likely to be examined closely, possibly prioritising certain interpretations over others. At the same time, the impact of these tools was largely confined to the preliminary stages of legal reasoning – accessing and understanding the law – and did not directly alter the core judicial functions of interpreting statutes, assessing evidence, or applying the law to specific cases.

4. Tax expert systems

The second development of AI systems in courts, which gained momentum in the late 1980s, centred on creating man-made models designed to automatically reproduce judicial reasoning. This objective is part of one of the most prominent goals of the AI & Law community to establish computable models of legal reasoning [9].

The most notable applications in this context have been rule-based systems. These systems consist of a rule base that holds legal rules, and an inference engine that applies these rules to the facts of a case to draw conclusions. This approach was initially demonstrated in significant projects, such as the formalisation of the British Nationality Act using logic programming [10], showcasing the potential of rule-based models in specific legal domains.

Rule-based systems have achieved discrete success in administrative decision-making, particularly in areas like social security and taxation. For example, among the most successful endeavours in building legal knowledge-based systems are the TAXMAN I and II system [11]. This pioneering system was developed to operate within the narrowly defined

---

[9] A. GARDNER, *An Artificial Intelligence Approach to Legal Reasoning*, Stanford, 1984; T. BENCH-CAPON, J. FORDER, *Knowledge Representation for Legal Applications*, in *Knowledge-Based Systems and Legal Applications*, Amsterdam, 1991.

[10] M. SERGOT et al., *The British Nationality Act as a Logic Program*, inn *Communications of the ACM* 29, 5, 1986, p. 370 ff.

[11] L.T. MCCARTY, *The TAXMAN Project: Towards a Cognitive Theory of*

domain of United States corporate tax law, specifically the reorganisation of corporations. Its primary function was to determine whether a given corporate reorganisation qualified for exemption from income tax under the Internal Revenue Code (IRC). The system achieved this by classifying each case according to the statutory categories of Type B, Type C, or Type D reorganisations, corresponding to sections 354, 355, and 356 of the IRC. An extended version of Taxman later broadened this scope to address the full tax implications for all parties involved in a reorganisation, as well as the treatment of corporate distributions beyond it.

The long-term aspiration behind Taxman was to capture and operationalise abstract tax concepts, such as the distinction between "form" and "substance", and to translate these into formalised, machine-readable logic. Although this ambition remained unrealised, the project marked a significant milestone in conceptualising how AI might engage with complex statutory reasoning.

Between the 1970s and the 1990s, several other AI systems in the domain of tax law emerged, generally employing rule-based reasoning methods, such as Tax Advisor [12] and Taxadvisor [13], the first being a Prolog based tax law representation. The sustained interest in tax law as a testing ground for AI reflected its distinctive characteristics: a domain governed by intricate, formal rules and definitions, yet relatively insulated from the ambiguities of everyday human experience, making it particularly amenable to computational modelling.

More recently, this tradition has been extended through the development of domain-specific languages (DSLs) designed to capture the semantics of legal norms, such as CATALA [14], which provides a structured and verifiable means of encoding legislative logic while maintaining close alignment with the

---

Legal Argument, in *Computer Science and Law: An Advanced Course*, 1980, p. 23 ff.

[12] D. SCHLOBOHM, *TA| a Prolog Program Which Analyzes Income Tax Issues Under Section 318 (a) of the Internal Revenue Code*, in *Computing Power and Legal Reasoning,* 765, 1985.

[13] R.H. MICHAELSEN, *An Expert System for Federal Tax Planning*, in *Expert Systems* 1, 1984, p. 14 ff.

[14] D. MERIGOUX et al., *Catala: A Programming Language for the Law* in *Proceedings of the ACM on Programming Languages* 5, 2021, pp. 1–29.

natural-language text of statutes. CATALA is a domain-specific programming language designed to formally encode and verify the logic of legal and regulatory texts. Developed by Merlin Carl, Denis Merigoux, and colleagues, CATALA is implemented on top of OCaml package. This foundation supports both human-readable and machine-verifiable representations of legal rules, enabling transparent computation in domains such as taxation, benefits eligibility, and regulatory compliance.

The development of domain-specific languages in the context of expert systems represents an evolution from traditional rule-based AI approaches toward more expressive, maintainable, and semantically faithful representations of expert knowledge. Early expert systems in law, medicine, and engineering relied heavily on ad hoc rule encodings within general-purpose programming environments, which made them difficult to scale and validate. DSLs, by contrast, are tailored to capture the unique syntax, semantics, and reasoning patterns of their target domains [15]. In expert systems, they facilitate clearer communication between domain experts and system developers, improving interpretability by being intelligible to legal practitioners.

These domains involve complex networks of rules, each with specific meanings, typically governing well-defined and uncontested cases. For example, rules determining eligibility for social security benefits may include criteria such as age, family status, income, and assets, often elaborated in additional rules. In such contexts, rule-based systems excel because they can meticulously apply all relevant rules to specific cases, provided these rules are accurately formulated within the system.

They have also been called "expert systems" because they seek to mimic the way in which a human expert in a certain field applies their skills to specific types of problems. These systems attempt to reflect the reasoning processes and decision-making patterns that an experienced specialist might use when confronted with problems in their domain. Unlike traditional software, which follows fixed algorithms, expert

[15] A. CHUN et al., *Domain-Specific Languages and Legal Applications*, in *The Journal of Robotics, Artificial Intelligence & Law* 7, 2024, p. 19 ff.

systems use knowledge representation and inference mechanisms to draw conclusions from stored information, much like a human professional might reason.

The professional responsible for translating the domain knowledge in a computable form is the knowledge engineer, who formalizes and encodes the expert's knowledge in a programming language. The task is both technical and interpretive. A knowledge engineer translates tacit, common-sense-based insights into clear rules or models that the expert system can use to reason effectively.

In contrast, in judicial proceedings, where facts, concepts, and rules are often contentious, rule-based systems encounter significant theoretical and practical challenges. Judicial decisions frequently involve nuances, contextual interpretations, and considerations of equity and justice that cannot be adequately captured by rigid rules. Laws and legal precedents are dynamic and can evolve, with new case-specific information potentially arising during proceedings that static rule-based systems cannot accommodate. Additionally, these systems struggle in situations where a degree of discretion is required to assess factual scenarios in light of legal principles and political objectives. Ultimately, the limitations of such systems reflect a form of "mechanical jurisprudence" [16], as they can only address cases through the application of predetermined rules.

Maintaining and updating expert systems present several ongoing challenges, both technical and conceptual. Once an expert system is deployed, its knowledge base and reasoning mechanisms must remain accurate and relevant to the evolving domain it represents. However, knowledge is rarely static, as new laws, interpretations, cases are constantly introduced. As a result, an expert system that performs well at one point in time can become outdated or even misleading if its knowledge

[16] The reference is to R. POUND, *Mechanical Jurisprudence*, in *Columbia Law Review*, 1908, p. 605 ff., where he argued that American common law or judge-made law had become sterile, unable to adapt to changing social and economic conditions, thus resulting in a closed system of many archaic rules that judges and lawyers deducted from general "conceptions" and applied mechanically to the actual situations before them.

base is not regularly revised. This problem is often referred to as knowledge obsolescence [17].

Updating an expert system is not as straightforward as modifying a database entry or a piece of procedural code. Because the knowledge base is structured around logical rules and interdependent relationships, adding or changing one rule can produce unexpected effects elsewhere in the system. Ensuring consistency and coherence across the knowledge base therefore requires careful validation and testing after every modification [18]. The knowledge engineer has to identify outdated rules, propose revisions, and verify that updates reflect current practice [19].

In response to these challenges, the development of argumentation-based systems has gained traction [20]. Argumentation plays a central role in judicial proceedings, where opposing parties present conflicting arguments that judges must weigh and evaluate. Judicial opinions reflect disputes by evaluating all arguments to reach a decision. A sound judgment requires showing that the winning argument is stronger than the opposing one or that counterarguments are invalid. This highlights the legal challenge of defeasibility, where both supporting and opposing arguments must be considered in a case [21].

Argumentation-based AI systems have been designed to capture this dialectical process, mapping out the interactions between arguments, counterarguments, and supporting evidence, thereby aiding judges in understanding the complex web of legal reasoning involved in a case. For example, ArgtuProlog is an argumentation-based AI system that uses logic programming to formalize and evaluate arguments within a

---

[17] J. DURKIN, *Expert Systems: A View of the Field*, in *IEEE Intelligent Systems* 11, no. 02, 1996, p. 56 ff.

[18] S. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach (4 Edition)*, Pearson, 2021.

[19] E.A. FEIGENBAUM, *Knowledge Engineering*, in *Annals of the New York Academy of Sciences* 426, 1984, pp. 91–107.

[20] T. BENCH-CAPON, *Argument in Artificial Intelligence and Law*, in *Artificial Intelligence and Law*, 1997, pp. 249–261; T. BENCH-CAPON et al., *Argumentation in Legal Reasoning*, Springer, 2009.

[21] H. PRAKKEN, G. SARTOR, *Law and Logic: A Review from an Argumentation Perspective*, in *Artificial Intelligence*, 2015, pp. 214–245.

legal context[22]. It captures the dialectical process by representing legal rules, facts, and arguments in a structured format using Prolog, a logic programming language, and generates possible conclusions based on the input data, highlights conflicts between opposing arguments, and supports the user in weighing evidence.

While these systems have been successfully utilized to develop platforms within recent EU projects[23], despite their potential, argumentation-based systems have not yet been widely adopted in courts, largely due to the challenges of accurately representing the nuanced nature of legal arguments and the need for continual expert input to maintain their relevance. Additionally, these systems face significant limitations, such as the reliance on simplified examples that do not scale to complex, large-scale legal cases, a lack of empirical validation making it hard to assess real-world utility, and difficulties in accurately modelling legal standards of proof, which are often complex and context-specific[24].

In the context of judicial reasoning, case-based legal reasoning systems have also been developed, especially, but not exclusively, in common-law countries[25]. These applications are built on the premise that legal cases can be represented as logical rules and that judicial reasoning could be modelled as a systematic process of applying these precedents to facts. The aim was to create computable models

---

[22] R. CALEGARI et al., *Arg-tuProlog: A Modular Logic Argumentation Tool for PIL*, in *Legal Knowledge and Information Systems*, IOS Press, 2020.

[23] M. BILLI et al., *A Hybrid Approach for Accessible Rule-Based Reasoning Through Large Language Models*, in *18th International Workshop on Juris-Informatics*, 2024; M. BILLI, A. PARENTI, *Access to Justice Through AI*, in *Facilitating Judicial Cooperation in the EU*, Brill Nijhoff, 2025.

[24] K. ASHLEY, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge, 2017, p. 144.

[25] K. ASHLEY, *Case-Based Reasoning and Its Implications for Legal Expert Systems*, in *Artificial Intelligence and Law* 1, 1992, pp. 113–208; K. ATKINSON, T. BENCH-CAPON, *Legal Case-Based Reasoning as Practical Reasoning*, in *Artificial Intelligence and Law*, 2005, pp. 93–131. For the civil law system, K. ASHLEY, *Case-Based Models of Legal Reasoning in a Civil Law Context*, in *Invited Paper, International Congress of Comparative Cultures and Legal Systems of the Instituto de Investigaciones Jurıdicas, Universidad Nacional Autonoma de Mexico, Mexico City*, Citeseer, 2004.

of case law that could assist judges and courts in managing the complexities of judicial reasoning by applying rules derived from cases and facts in a structured manner.

Case-based legal reasoning systems work by drawing analogies between new cases and previously decided cases. When a new case arises, the system identifies relevant precedents and analyses the factors that influenced past decisions. These factors are then used to predict the likely outcome of the current case based on similarities and differences with the prior cases. By employing this method, judges can benefit from a systematic approach to considering how established legal principles apply to new factual situations, thereby aiding their decision-making process.

An example of a case-based system is the HYPO system used in trade secret infringement cases [26]. This system's knowledge base collects several precedents, each of which is described or annotated with its outcome and a set of factors that support or oppose that outcome. For instance, if the defendant was aware of the plaintiff's activities and certain conditions were met, this would bolster the conclusion of trade secret infringement. Conversely, other factors could support a no-infringement conclusion, such as the plaintiff's communication during negotiations or information obtainable through reverse engineering. With these factors, the system employs analogical reasoning to predict the possible outcome of a new case. Essentially, it compares the factors of new cases to precedents with similar characteristics.

The use of case-based systems in courts has faced significant challenges. One of the major issues is the need to create a precise and computable representation of the case base. This process demands extensive effort from legal experts and technologists to encode cases into formal structures that can be processed by the system. Moreover, consistently assigning all relevant factors to a large set of cases is a time-consuming and controversial process that may easily reflect the biases of the experts performing the task. The representation of cases may not capture all the nuances that

[26] E. RISSLAND, K. ASHLEY, *A Case-Based System for Trade Secrets Law*, in *Proceedings of the 1st International Conference on Artificial Intelligence and Law*, 1987, pp. 60–66.

influence decisions, leading to oversimplifications that can affect the quality of the analysis. To overcome these limitations, natural language processing (NLP) technologies are being explored to automatically assign factors to cases, thereby enhancing the efficiency of case-based systems [27].

Overall, while argumentation-based and case-based systems represented significant advances in how AI can be used within courts, their real-world applications were –and still are – limited, primarily due to their reliance on human-provided legal knowledge that must be meticulously encoded and regularly updated. Furthermore, these systems have proved less effective in cases involving complex or controversial legal issues, where human judgement and interpretive skills remain crucial.

5. Machine learning for tax cases

The third step of AI in courts has been characterised by a paradigm shift in AI research towards data-driven approaches and machine learning, thus marking a significant departure from earlier human-crafted models of legal reasoning [28]. In machine learning, the system builds its own knowledge model by applying learning algorithms to extensive datasets. As the system processes more data, it continuously refines and updates its model, enhancing its ability to classify, evaluate, and predict outcomes for new cases presented before it. This self-improving capability has made machine learning a transformative tool in many fields, including in the legal domain.

Machine learning approaches are generally divided into three categories: supervised learning, unsupervised learning,

[27] For instance, M. H FALAKMASIR, K. ASHLEY, *Utilizing Vector Space Models for Identifying Legal Factors from Text*, in *Legal Knowledge and Information Systems*, IOS Press, 2017. More recently, on using LLMs to automatically identify factors, M. GRAY et al., *Using LLMs to Discover Legal Factors*, 2024, arXiv preprint – arXiv:2410.07504, 2024.

[28] N. CRISTIANINI, *On the Current Paradigm in Artificial Intelligence*, in *AI Communications*, 2014, pp. 37–43.

and reinforcement learning [29]. In supervised learning, the system is trained on labelled data containing correct answers from previous cases, allowing it to predict outcomes for new cases by identifying patterns and similarities with past instances. Unsupervised learning, in contrast, does not rely on labelled data; instead, the system autonomously identifies structures and relationships within the data, classifying and organising the information it receives. Reinforcement learning involves a feedback loop where the system learns through rewards and penalties, adapting its behaviour to maximise desired outcomes, which is particularly useful in complex decision-making environments.

In the legal context, machine learning has led to the development of legal analytics, a research field employing advanced statistical and natural language processing techniques to extract information from extensive collections of legal texts [30]. These technologies allow automatic analysis of judicial decisions, statutes, contracts, and other legal documents, making the vast amount of legal data more accessible and interpretable for judges and legal professionals.

Legal analytics applications in courts can be broadly categorised into case-oriented and document-oriented approaches. Case-oriented approaches employ machine learning models developed from large datasets of past cases to predict various aspects of new cases, such as their likely duration, potential costs, and possible judicial outcomes. These models embed correlations between features of cases – such as the nature of the claim, legal arguments, procedural history, and possibly even the identities of judges and parties – and the decisions made.

Document-oriented systems focus on the analysis of individual legal texts, extracting key information such as named entities (places, persons, organisations), dates, claims, and complex events. They also support automated summarisation, which condenses lengthy legal documents into coherent, concise summaries of case facts, judicial reasoning,

---

[29] For an introduction to machine learning, E. ALPAYDIN, *Introduction to Machine Learning*, MIT press, 2020.

[30] For a systematic review of research achievements in the field of legal analytics, K. ASHLEY, *Artificial Intelligence and Legal Analytics*, cit.

and outcomes. This allows judges to quickly access the most pertinent information without reading through voluminous records of cases. Automated parsing of statutory texts converts natural language legal documents into machine-readable rules, enhancing the efficiency of legal research and decision-making.

In the legal domain, machine learning systems based on documents have predominantly relied on a supervised learning approach. This requires a critical annotation step, where legal experts meticulously label data with relevant legal information, such as identifying legal issues, categorising case outcomes, or marking key legal concepts within documents. Annotation ensures that the system learns to recognise complex legal language, principles, and the context in which they are used. This phase is crucial because the quality of annotations directly impacts the system's ability to interpret and apply legal knowledge accurately.

Annotation methodologies in legal machine learning vary depending on the specific task and desired level of granularity. Common approaches include manual annotation by trained legal professionals, semi-automated annotation using pre-trained models to suggest labels that human annotators then verify, and crowdsourcing in limited contexts where the legal expertise required is minimal. Structured annotation schemes, such as hierarchical taxonomies or ontology-based frameworks, are often employed to ensure consistency across annotators and facilitate downstream analysis [31].

To ensure reliability and reproducibility, many projects aim to create "golden datasets", carefully curated, fully verified collections of annotated legal documents that serve as benchmarks for model evaluation and training. Golden datasets are often developed through multi-stage review processes, where annotations are cross-checked by multiple experts to achieve high inter-annotator agreement [32]. They play an important role in benchmarking legal NLP models,

---

[31] K. ASHLEY, *Automatically Extracting Meaning from Legal Texts: Opportunities and Challenges*, in *Georgia State University Law Review*, 2018, p. 1117.

[32] I. CHALKIDIS et al., *LexGLUE: A Benchmark Dataset for Legal Language Understanding in English*, arXiv preprint – arXiv:2110.00976, 2021.

allowing researchers to assess the generalizability and robustness of different algorithms under consistent conditions. Examples include domain-specific corpora for contract analysis, case law summarisation, or statutory interpretation [33].

Despite their value, annotation in the legal domain faces several common challenges. Legal texts are inherently context-dependent and open to ambiguity, leading to difficulties in establishing clear labelling criteria. Differences in legal interpretation between jurisdictions or even among experts can result in inconsistent annotations [34]. Furthermore, the annotation process is time-consuming and resource-intensive, as it demands expert-level legal knowledge. Data privacy and confidentiality restrictions on sensitive legal documents can further limit the availability of training data.

For example, in legal argumentation mining, machine learning systems can be trained to automatically identify and classify arguments from cases by using annotated datasets where legal experts have labelled different parts of judicial decisions. These annotations should reflect how arguments are made in case decisions, for instance, the extent to which they incorporate factual aspects, the type of interpretative canons used, and how they connect to other arguments being made in the same case. Once the training is completed, legal expertise is also required to validate the accuracy of the model's classifications and to ensure that the AI's understanding aligns with complex legal reasoning. Legal experts must review the outputs generated by the machine learning system, assessing whether the identified arguments have been correctly categorised and if the interpretative methods and factual aspects have been accurately captured.

The uptake of machine learning has spurred a wide variety of legal analytics applications and services in the justice sector, especially in litigation support. One notable application of

---

[33] ILIAS CHALKIDIS et al., *LEGAL-BERT: The Muppets Straight Out of Law School*, in T. COHN ET AL (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020*, Association for Computational Linguistics, 2020, https://doi.org/10.18653/v1/2020.findings-emnlp.261.

[34] D.G. GORDON, T.D. BREAUX, *The Role of Legal Expertise in Interpretation of Legal Requirements and Definitions*, in *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, IEEE, 2014, p. 273–282.

machine learning in the legal domain is IBM's Watson [35], which is integrated into systems like ROSS Intelligence [36]. ROSS combines advanced text analytics with predictive functionalities, allowing users to query the system in plain English and receive relevant legal answers drawn from statutes, case law, and other sources. It continuously learns from user interactions, refining its responses, and enhancing its utility in legal research. Watson's capabilities extend to the automated analysis of legal briefs, where it can cross-reference cited cases, evaluate their precedent value, and identify similar language across a vast array of legal texts.

Other prominent applications include LexMachina – a LexisNexis product originally developed for IP law now also expanded to antitrust and commercial litigation – which uses machine learning to analyse legal data, providing insights into case outcomes, judge tendencies, and litigation trends [37]; Premonition [38], which analyses judicial tendencies and attorneys' performance and the UK-based Luminance [39], which employs machine learning for document review, enabling the detection of anomalies and automating legal annotations. These systems provide predictive insights that can assist judges by highlighting trends in case law, helping them understand how their decisions align with broader judicial patterns.

Machine learning applications in judicial contexts have also raised several questions regarding their privacy and discrimination, opacity, and the unreliability of predictions. For example, concerns have been extensively raised about the potential biases embedded within machine learning systems, particularly if trained on historical data that reflects societal inequalities [40]. The risk of perpetuating existing biases could

---

[35] Artificial Lawyer, *IBM Watson's Inhouse AI App Will Change the Legal World (If It Catches on...)*, 2017, https://www.artificiallawyer.com/2017/05/23/ibm-watsons-inhouse-ai-app-will-change-the-legal-world-if-it-catches-on/.

[36] https://blog.rossintelligence.com.

[37] https://lexmachina.com.

[38] https://premonition.ai.

[39] https://www.luminance.com.

[40] J. KLEINBERG et al., *Discrimination in the Age of Algorithms*, in *Journal of Legal Analysis*, 2018, pp. 113–174; M.A. MALEK, *Criminal Courts'*

lead to discriminatory and unjust outcomes, raising the need for rigorous evaluation and validation of machine learning models before they are deployed in judicial settings.

Moreover, the "black box" nature of many machine learning algorithms poses challenges for accountability and transparency [41]. Judges, litigants, and citizens at large may struggle to understand how predictions are generated, which can undermine trust in the justice system. The opacity of algorithmic prediction may restrict the fundamental rights to a fair trial by preventing judges from fully evaluating evidence, compromising equality of arms when parties lack access to algorithmic processes, and hindering effective remedies when citizens cannot challenge or appeal decisions they do not understand [42].

In response to these issues, recent research has focused on explainability and interpretability in algorithmic decision-making. Explainable AI research aims to make machine learning models more transparent by offering interpretable, human-understandable reasons for their outputs [43]. In judicial applications explainability is necessary for a system to be employed, as legal decisions (e.g. judgements) depend as much on the reasoning process as on the final outcome. Models capable of indicating which factors most influenced a given prediction can support judicial review and allow appeals from citizens and legal professionals alike. Nonetheless, there is currently no technical standard for what explainability entails, and expectations of what constitutes an adequate explanation often differ between legal and technical communities [44].

A further issue relates to the data demands of machine

Artificial Intelligence: The Way It Reinforces Bias and Discrimination, in *AI and Ethics*, 2022, pp. 233–245.

[41] F. PASQUALE, *The Black Box Society*, Harvard University Press, 2015; A. DEEKS, *The Judicial Demand for Explainable Artificial Intelligence*, in *Columbia Law Review*, 2019, pp. 1829–1850.

[42] See, e.g., G. CONTISSA, G. LASAGNI, *When It Is (Also) Algorithms and AI That Decide on Criminal Matters: In Search of an Effective Remedy*, in *European Journal of Crime, Criminal Law and Criminal Justice*, 2020, pp. 280–304.

[43] G. SARTOR et al., *Thirty Years of Artificial Intelligence and Law: The Second Decade*, in *Artificial Intelligence and Law*, 2022, pp. 521–557.

[44] F. DOSHI-VELEZ et al., *Accountability of AI Under the Law: The Role of Explanation*, arXiv preprint arXiv:1711.01134, 2017.

learning systems used in judicial contexts. Building accurate models requires extensive and carefully curated datasets. This is in contrast with judicial data, which is often recorded in non-machine readable formats. Furthermore, much of it contains sensitive or personally identifiable information that must be processed and anonymized before it can be used [45]. Access to legal data is also limited by intellectual property and data protection regulations. Such restrictions often prevent researchers and public institutions from developing or evaluating algorithmic tools on an equal footing with private companies, leading to higher costs and less accurate models [46].

An example of a European project that aims to address some of these challenges is the ADELE project [47]. The work began with the collection and preliminary analysis, carried out by legal domain experts, of case law to understand the structure of decisions, the language used, recurring argumentative patterns, citation practices, and the overall content of judgments. From this analysis, a set of annotation guidelines was developed. These guidelines specify how legal experts should annotate texts, defining what information to capture and providing a structure for recording the relationships between different elements of a decision and the arguments presented.

Using these guidelines, the ADELE team carried out corpus annotation in stages, adding machine-readable metadata to judgments. Experts in the relevant legal fields performed the annotation, and part of the process used a double-blind method in which two annotators independently tagged the same documents to ensure consistency. Inter-annotator agreement was measured to maintain the coherence and reliability of the datasets and to avoid treating the same legal information differently across documents [48].

---

[45] H. SURDEN, *Machine Learning and Law: An Overview*, in *Research Handbook on Big Data Law*, Edward Elgar Publishing, 2021, pp. 171–184.

[46] T. MARGONI, *Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?*, in *Machine Learning and EU Copyright Law: Who Owns AI*, 2018.

[47] https://site.unibo.it/adele/en.

[48] F. GALLI et al., *Analytics for Deciding Legal Cases: The ADELE Project*, in *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights*, Scuola Superiore della Magistratura, 2024.

Furthermore, another research current aimed at solving these issues leverages "neuro-symbolic methods" which integrate machine learning with symbolic reasoning techniques [49]. These systems represent legal knowledge in interpretable, symbolic formats, providing explainable insights that foster trust among judges and legal practitioners. Meanwhile, their adaptive learning capability enables AI tools to stay current with evolving legal standards, ensuring they remain relevant and accurate in their support of judicial decision-making.

6. Large Language Models and tax case law

The fourth step of AI in courts is marked by the rise of generative AI and Large Language Models. This evolution builds upon recent advances in machine learning, in particular in the field of Natural Language Processing.

Initially, NLP systems relied on syntactic patterns, focusing on the structure of text rather than its deeper semantic meaning, which made it challenging for computers to perform tasks that seem intuitive to humans [50]. However, this limitation was overcome with the introduction of pre-trained language models. These models are trained on large corpora of text and capture the syntactic and semantic relationships between words and phrases by learning from vast amounts of language data before being applied to specific tasks. One of the most famous examples is BERT (Bidirectional Encoder Representations from Transformers) [51], the model released by Google, based on an encoder-based architecture [52]. Specialised language models have also emerged, including in the legal

[49] G. PISANO et al., *Neuro-Symbolic Computation for XAI: Towards a Unified Model*, in *CEUR Workshop Processings*, 2020, pp. 101–117.

[50] Cfr. Par 3.

[51] J.D. DEVLIN, M.C. KENTON, L. KRISTINA TOUTANOVA, *BERT: Pre-training of deep bidirectional transformers for language understanding*, in *Proceedings of naacL-HLT* 1, 2019, p. 2.

[52] An encoder architecture converts each word in a sentence into a high-dimensional vector that captures both its meaning and context by analysing the surrounding words. This enables the model to "encode" the semantic and syntactic features of text, making it highly effective for language comprehension tasks.

context. Models like LegalBERT [53] are specifically trained on legal documents and allow systems to process legal texts with contextual accuracy.

Large Language Models represent an even more advanced category within language models. Their design builds on the 2017 Transformer architecture, introduced by Google researchers, which revolutionised NLP through the so-called "attention mechanism" [54]. This approach allows models to understand text contextually and sequentially, enabling them to perform sophisticated text generation tasks. For instance, OpenAI's Generative Pre-Training (GPT) models have shown that training a model on vast, unlabelled datasets allows it to not only classify text but also generate new, contextually relevant content.

LLMs have introduced a plethora of new opportunities in the judiciary by offering capabilities that extend beyond the traditional boundaries of legal research and analytics [55]. For example, LLMs can enhance legal research by quickly retrieving relevant case law and statutes, thus enabling judges to access pertinent information without sifting through extensive volumes of documents. LLMs have also been enhanced through the use of Retrieval Augmented Generation (RAG), which extends the capabilities of LLMs to specific domains, or an organization's internal knowledge base, all without the need to retrain the model. This introduces an additional step in the information retrieval component, which utilizes the user input to first pull information from a new data source. The user query and the relevant information are both given to the LLM, which then uses the new knowledge and its training data to create better responses.

LLMs and RAG approaches have been developed to condense lengthy judicial cases and legal briefs into concise summaries and help decision-makers grasp essential points swiftly. In this context, one of the first real-world applications of LLMs in the summarisation and legal search of court

---

[53] CHALKIDIS et al., *LEGAL-BERT*, cit.

[54] A VASWANI, *Attention Is All You Need*, in *Advances in Neural Information Processing Systems*, 2017.

[55] H. SURDEN, *ChatGPT, AI Large Language Models, and Law*, in *Fordham Law Review*, 2023, p. 1941.

decisions was provided by the Italian PRO.DI.GI.T Project [56], in which GPT-4 was used to create summaries of all tax decisions delivered by the Italian Tax Commission and available in a digital format.

Beyond research and summarisation, LLMs provide valuable support in legal drafting by offering templates for repetitive cases and streamlining the creation of legal documents, such as judgments and orders. Specific support can also be provided in generating hypotheticals and arguments, thus exploring different scenarios and evaluating the potential outcomes of their decisions. Additionally, LLMs offer flexibility by structuring documents in various formats (e.g., Markdown, JSON, or Excel) as required, enhancing adaptability and efficiency in the document preparation process [57].

The use of LLM-based systems in the judicial domain has led to the uptake of "legal prompt engineering" [58], which involves crafting precise and contextually appropriate prompts to guide LLM-based systems in generating legally accurate and relevant outputs. Legal prompt engineering goes beyond general prompting strategies [59]; it requires a deep understanding of legal language, concepts, and the nuances of legal contexts. For instance, a judge might need to quickly extract the most relevant information from a case to understand its core legal arguments and implications. This involves defining what "relevance" means in a legal context, which can include factors such as the main legal issues at stake, the applicable legal provisions, the key evidence presented, and the court's rationale in reaching its decision.

The deployment of large language models in courts introduces risks that echo and extend those associated with machine learning. For example, bias and opacity are significant challenges with LLMs in judicial settings. The

---

[56] T. DAL PONT et al., *Legal Summarisation Through LLMs: The PRODIGIT Project*, 2023, arXiv preprint – arXiv:2308.04416, 2023.

[57] J. LAI et al., *Large Language Models in Law: A Survey*, *Preprint Submitted to Elsevier*, 2023, pp. 5–19.

[58] F. YU et al., *Legal Prompting: Teaching a Language Model to Think Like a Lawyer*, arXiv Preprint – arXiv:2212.01326, 2022.

[59] P. BANSAL, *Prompt Engineering Importance and Applicability with Generative AI*, in *Journal of Computer and Communications*, 2024: pp. 14–23.

complexity of these models makes it difficult to trace the reasoning behind specific outputs, leading to a lack of transparency. Additionally, LLMs often reflect historical biases in their training data, risking reinforcement of systemic inequities in legal judgments.

Also, specific concerns related to the tendency of models to "hallucinate", that is, fabricating non-existent or erroneous information, as seen in a US case when an AI used by a lawyer cited non-existent precedents [60]. Specific types of "legal hallucination" [61] may emerge from the models, such as closed-domain errors where responses conflict with prompts (e.g., mischaracterising legal opinions), open-domain errors where outputs stray from or contradict training data and factual hallucinations where responses diverge from actual legal facts. Such inaccuracies, particularly problematic in complex legal tasks, raise risks in contexts like drafting or interpreting case law.

Some risks may also stem from technical vulnerabilities exploited by malicious actors. For example, by crafting specific prompts (known as "prompt injection"), the model may be led to provide erroneous answers or unauthorised information [62]. Another attack technique is data poisoning, where an attacker introduces harmful data into the training set, subtly altering the model's behaviour.

LLMs share several fundamental limitations with traditional machine learning systems, first and foremost the absence of transparent reasoning. The output of an LLM is generated through a sequence of statistical and probabilistic computations derived from large-scale training data, rather than from any interpretable reasoning process [63]. This lack of

---

[60] M. BOHANNON, *Lawyer Used ChatGPT in Court and Cited Fake Cases – a Judge Is Considering Sanctions*, in *Forbes*, 2023, https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/.

[61] M. DAHL et al., *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, in *Journal of Legal Analysis*, 2024, pp. 64–93.

[62] I. KILOVATY, *Hacking Generative AI*, in *Loyola of Los Angeles Law Review*, 2025, p. 58.

[63] M AMIRIZANIANI et al., *Can LLMs Reason Like Humans? Assessing Theory of Mind Reasoning in LLMs for Open-Ended Questions*, in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 34–44.

explainability poses difficulties in contexts such as law and policy, where accountability and justification of decisions are to be considered necessary in order for a decision to be valid, as seen in the previous chapter. Users, regulators, and affected individuals are often unable to reconstruct why a particular outcome was produced or verify its consistency with normative or factual expectations.

Another structural limitation concerns the data used in training. Proprietary datasets are, by definition, not open to the public and the research community, thus making it difficult if not impossible to assess biases, representativeness, and compliance with data protection rules. Open-source models improve transparency but often perform worse than closed models due to smaller datasets or limited fine-tuning resources. This trade-off between openness and performance creates uncertainty regarding the reliability of LLMs in high-stakes domains, especially when the origin and quality of training data remain undisclosed [64].

Fine-tuning offers a partial remedy by adapting general-purpose models to specialized tasks. However, this process entails substantial computational costs and requires high-quality domain-specific datasets, which are difficult to obtain in legally or ethically sensitive areas. Furthermore, fine-tuned models tend to lose generalization capacity, making them less adaptable to new problems or contexts outside their training scope [65].

Benchmarking LLMs adds another layer of complexity. Existing benchmarks often fail to capture the nuances of real-world applications, focusing instead on narrow linguistic or reasoning tasks. Performance metrics may therefore overstate the practical capabilities of these systems, obscuring weaknesses in consistency, factual accuracy, or compliance with legal reasoning standards. Developing more representative and context-sensitive benchmarks remains a

---

[64] E. OLLION et al., *The Dangers of Using Proprietary LLMs for Research*, in *Nature Machine Intelligence*, 2024, pp. 4–5.

[65] V. BALAVADHANI PARTHASARATHY et al., *The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities*, arXiv Preprint arXiv:2408.13296, 2024.

pressing research challenge for the responsible deployment of LLMs in legal and regulatory environments [66].

All these risks are heightened by the potential for over-reliance on AI-generated content, such as when drafting judgments. If judges begin to rely heavily on LLMs for drafting, they may inadvertently lower their level of scrutiny. Individual responsibility and active oversight are essential to ensure that AI contributions are carefully validated to uphold the integrity of legal standards and principles [67].

7. Conclusion

This chapter has examined the evolution of artificial intelligence in tax law through the perspective of legal informatics, outlining the major technological and conceptual shifts that have influenced judicial practice. The analysis began with the emergence of information retrieval systems, which digitised tax case law and improved judicial efficiency by facilitating systematic access to precedents and statutes. These developments transformed legal research, though their deterministic logic constrained interpretative flexibility. Their legacy endures in today's search engines and legal databases that are still the basis for much of legal work, for both lawyers and judges alike.

The second phase introduced rule-based reasoning and formal logic to capture the structure of legal norms. Projects such as TAXMAN I and II demonstrated how tax law could be modelled as a computable domain, while later developments, like domain-specific languages, refined the capacity to represent legislative logic with greater precision and transparency. This research current also shows the limit of this approach, as human interpretation remains necessary in applying law to complex, real-world cases.

The third phase, marked by machine learning and legal analytics, shifted the focus from explicit rules to data-driven

---

[66] T. R MCINTOSH et al., *Inadequacies of Large Language Model Benchmarks in the Era of Generative Artificial Intelligence*, in *IEEE Transactions on Artificial Intelligence*, IEEE, 2025.

[67] E. MIK, *Caveat Lector: Large Language Models in Legal Practice*, in *Rutgers Business Law Review*, 2023, p. 70.

discovery. By analysing vast corpora of judicial decisions and legal documents, these systems laid the groundwork for developing tools for summarisation, classification, and outcome prediction. Initiatives like IBM Watson, LexMachina, and ADELE exemplify this evolution. However, this approach also raised new ethical and legal concerns regarding bias, opacity, and explainability, issues that keep the research community engaged until today.

Finally, the rise of generative AI and LLMs, and due to their characteristics, such systems could have a significant impact in the legal sector. Tools like the PRODIGIT Project have showed the capacity of LLMs to summarise, draft, and interpret tax decisions, extending AI's role from information retrieval and prediction to content generation and reasoning support. Yet these systems bring new challenges, from hallucination and bias to explainability and data collection, that must be addressed to safeguard fairness and transparency in judicial decision-making.

Across these phases, the interaction between computation and law has remained reciprocal: technology has redefined how judges access and interpret legal knowledge, while legal reasoning continues to highlight the need for transparent decision.

Chapter II

LEGAL INTERPRETATION IN ARITIFICIAL INTELLIGENCE AND LAW

Federico Galli – Giuseppe Contissa*

Contents: 1. Introduction. – 2. Interpretation in law. – 3. Interpretation in AI & Law. – 4. Interpretation and legal knowledge engineering. 5. Interpretation and legal data annotations. – 6. Interpretation and prompt-engineering. – 7. The end of legal interpretation or a new beginning?

1. Introduction

Artificial intelligence (AI) is rapidly transforming the legal domain, bringing both great promise and new challenges. We stand at a moment where AI systems can pass a bar exam [1], yet lawyers have also been sanctioned for submitting AI-generated briefs filled with fictitious citations [2].

---

* Federico Galli is Assistant Professor in Legal informatics and Computer Law at the University of Bologna; Giuseppe Contissa is Associate Professor in Legal informatics and Computer Law at the University of Bologna.

[1] See J. Arredondo et al., *GPT Takes the Bar Exam: What AI's Performance Means for Legal Education and Practice*, in *SSRN Electronic Journal*, ahead of print, 2023, https://doi.org/10.2139/ssrn.4429717. See also D. Katz et al., *Gpt-4 Passes the Bar Exam*, in *Philosophical Transactions of the Royal Society A,* p. 382, no. 2270, 2024, 20230254, who argue that such results illustrate not only technical advances in AI but also the pressing need to rethink legal education, emphasizing skills beyond doctrinal recall, such as ethical reasoning, judgment, and professional responsibility.

[2] For instance, in the United States, in Mata v. Avianca, Inc. (S.D.N.Y. 2023), attorneys were sanctioned after submitting a brief containing non-existent case law generated by ChatGPT. Similarly, in Italy, the Court of Florence (Order of 14 March 2025) addressed a defence memorandum citing

These contrasting episodes highlight the complex interplay between advanced AI capabilities and the nuanced demands of legal interpretation.

Legal practice has long hinged on interpreting texts – statutes, case law, contracts – in context, with human judgment calibrating the balance between literal words and broader purposes. As AI tools are increasingly deployed in tasks like legal research, document analysis, and even decision support, questions arise about how these systems understand (or misunderstand) legal meaning. Indeed, while AI offers efficiency and objectivity, its integration into law raises pressing questions about bias, accountability, and the fundamental principles of justice.

This chapter examines the role of legal interpretation in AI and Law, aiming to elucidate how theories and practices of interpretation influence the design, training, and evaluation of AI systems for legal applications.

Section 2 provides an overview of what "legal interpretation" entails in jurisprudence, establishing why interpretation is a central, non-trivial aspect of legal reasoning.

Next, *Section 3* introduces the ambiguous role of legal interpretation in AI and law, i.e. whether interpretation is something that is surpassed by the use of AI in legal setting shifting toward mechanized attribution of meaning.

To take part in the debate, *Section 4-6* show three key junctions where legal interpretation is actually relevant in AI development: (a) Legal Knowledge Engineering – how interpretative choices inform the modelling of legal knowledge in expert systems and ontologies; (b) Legal Data Annotation – how interpretation guides the labelling of legal texts for machine learning and the challenges of ensuring consistent, expert-informed annotations; and (c) Prompt Engineering – how the way we frame questions or instructions to AI (especially large language models) can dictate which legal interpretation an AI adopts.

Finally, *Section 7* answers the question whether legal AI applications refute the role of legal interpretation. We posit

---

fabricated Court of Cassation precedents. The Court recognized the imprudence of such conduct but declined to impose sanctions, noting that the fictitious citations were not decisive to the dispute and that no bad faith or gross negligence was shown.

that AI does not represent the end of legal interpretation but inaugurates a new beginning, where interpretative reasoning itself becomes augmented in collaboration with machines but necessitates mechanisms of transparency and contestability.

2. Interpretation in law

Legal interpretation is the process by which legal professionals determine the meaning of authoritative texts (such as statutes, regulations, contracts, or constitutional provisions) and apply them to particular facts [3]. In legal practice, "interpretation" is invoked whenever the meaning of a text is not immediately clear or is disputed in a concrete situation.

Legal philosophers have long recognized that many legal rules are written in open-textured or ambiguous language, requiring judgment in order to bridge gaps, resolve ambiguities, or choose among competing plausible readings [4]. As a simple example, consider the famous example used by Herbert L.A. Hart of a statute that prohibits "vehicles" in a public park [5]: does this statute applies also bicycles? wheelchairs? toy scooters? The text alone may not settle the matter, and different interpreters might emphasize the ordinary meaning of "vehicle", the law's purpose (e.g. safety and tranquillity in parks), or other considerations to reach a conclusion.

Over centuries, jurists in different legal traditions have developed various theories and canons of interpretation.

For instance, in common law jurisdictions like the United States and U.K., debates often centre on textualism versus purposivism (or intentionalism) [6].

---

[3] D. Walton et al., *Statutory Interpretation: Pragmatics and Argumentation*, Cambridge, 2021, p. 22.

[4] Among many, see H.L.A. Hart, *The Concept of Law*, Oxford, 2012, R. Dworkin, *Taking Rights Seriously*, Harvard University Press, 1977, J. Raz, *Practical Reason and Norms*, 2nd ed., Oxford, 1990, and F. Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Oxford, 1991.

[5] H.L.A. Hart, *The Concept of Law*, cit., p. 128.

[6] See, e.g., A. Scalia, *A Matter of Interpretation: Federal Courts and the Law*, Princeton University Press, 1997, for a canonical defence of textualism;

Textualists argue that interpretation should focus on the ordinary meaning of the legal text's words at the time of enactment, giving primacy to the statute's language. Purposivists contend that the text must be read in light of the law's purpose or the problem it was intended to remedy, sometimes allowing the spirit of the law to override a literal reading. Courts also employ numerous canons of construction – for example, the plain meaning rule (follow the clear text unless ambiguity remains), *ejusdem generis* (interpret general terms in light of specific ones listed), or the mischief rule (interpret a statute to suppress the problem it aimed to address).

Civil law systems, such as those in continental Europe, similarly recognize a plurality of interpretations. A comparative study by MacCormick and Summers found that judges in various countries use a hierarchy of interpretive arguments [7]: starting with the literal (linguistic) meaning, then considering systematic/contextual consistency with the legal system, and finally teleological or intentional arguments about the law's purpose or the legislature's intent. These interpretive methods come with meta-rules [8] for when a court may depart from a plain meaning (e.g. in case of absurd results) and how to resolve conflicts between interpretive arguments.

In Italy, for example, the Civil Code explicitly sets out interpretive directives. Article 12 of the Preleggi (the preliminary provisions to the Civil Code) requires that statutes be interpreted according to their "literal meaning" but also in harmony with the legislator's intent and the overall coherence of the legal system. Italian scholars often distinguish between literal interpretation (focusing on the text), systematic interpretation (considering the provision within the structure of the legal order), and teleological interpretation (seeking the underlying purpose) [9]. In practice, Italian judges routinely combine these methods, balancing literal and purposive approaches depending on the case at hand.

Importantly, interpretation is ubiquitous in law not only in

H.M. HART, A.M. SACKS, *The Legal Process: Basic Problems in the Making and Application of Law,* Foundation Press, 1994, for a purposivist tradition.

[7] N.D. MACCORMICK, R.S. SUMMERS (eds.), *Interpreting Statutes: A Comparative Study*, Dartmouth, 1991.

[8] J. WRÓBLEWSKI, *The Judicial Application of Law*, Springer, 1992.

[9] Italian legal doctrine often refers to these canons as "interpretative arguments", see G. TARELLO, *L'interpretazione della legge*, Giuffrè, 1980.

appellate courts deciding constitutional questions, but also in everyday legal practice – a lawyer advising a client must interpret the relevant statutes and cases; a regulator drafting rules must interpret authorizing legislation; a contract drafter anticipates how language might be interpreted in the future, etc. In all cases, interpretive reasoning aims to make sense of legal language in context.

3. Interpretation in AI & law

Given the inherent complexity of the law, any AI system intended to operate on legal tasks must grapple with legal interpretation.

Yet, a recurrent claim in discussions on AI systems in the legal domain is that AI offers an objective mechanism for applying legal rules. According to this view, algorithms can take facts and rules as inputs and yield determinate outcomes without the need for human discretion. AI thus appears to refute the need for interpretation by presenting itself as a purely technical, calculable process.

This claim has strong rhetorical power. By presenting legal decision-making as calculable, [10] AI seems to strip away the uncertainty of human interpretation and the perceived arbitrariness of judicial discretion.

In predictive justice, for example, models that forecast case outcomes or sentencing patterns are portrayed as offering a standard of objectivity [11]. If a system can predict with high accuracy how a court will decide, it is said to be capturing "the law" as it truly is, thereby bypassing the need to engage in interpretive debates about statutes or precedents. Here, prediction itself is equated with legal truth.

---

[10] The idea of "legal calculability" was advanced by Weber in his analysis of rationalization and formal legal systems. M. Weber, *Economy and Society*, Harvard University Press, 2019, p. 76. On this point, in the Italian doctrinal debate, see also the volume A. Carleo (ed.), *Calcolabiltià Giuridica*, il Mulino, 2017, which collects the proceedings of a conference held in Rome in 2016 under the patronage of the Accademia Nazionale dei Lincei.

[11] See, for instance, N. Aletras et al., *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective* in *PeerJ Computer Science*, 2016, p. 93.

The same narrative is visible in private law applications, such as automated contract review. Vendors promote these tools as objective reviewers, flagging non-standard clauses or hidden risks according to uniform templates. The promise is that unlike human lawyers – who may disagree, interpret language differently, or bring subjective judgment – the machine will provide a consistent and neutral reading of the contract. Again, interpretation is portrayed as unnecessary, because the algorithm appears to treat legal language as stable and calculable.

This belief that law can be applied without interpretation is not new. It echoes the ambition of a "mechanical jurisprudence" [12], a tradition in which law was imagined as a closed logical system: judges (or now, machines) would merely deduce outcomes from clear rules. AI reinvigorates this old formalist dream with the aura of modern technology. The machine, more than the judge, is imagined as capable of applying rules without bias or discretion.

The claim that AI refutes interpretation is, at best, only partially correct. What AI actually does is shift interpretation from the visible domain of judicial reasoning to the hidden layers of algorithmic design, training, and deployment. Interpretation does not vanish; it becomes implicit, embedded in datasets, coding decisions, and model outputs that present themselves as neutral but are saturated with human assumptions.

The design choices made by engineers or researchers building legal AI systems often amount to choosing one interpretation of a law over another or simplifying nuanced legal concepts in ways that may not capture every scenario. Likewise, when preparing datasets to train or evaluate AI models, humans must interpret legal materials to provide correct labels or examples. Even when using AI models interactively – for instance, querying a legal question to an LLM – the way we pose the question and the context we provide will influence the interpretive angle the model takes.

---

[12] See R. POUND, *Mechanical Jurisprudence*, in *Columbia Law Review*, 1908, pp. 605–623; R. BORRUSO et al., *L'informatica Del Diritto*, Giuffrè, 2004. Borruso saw the possibility of resolving the problem of interpreting the law using information technology, summarising this idea in the motto *ubi regula, ibi computer*.

In sections that follow, we show how legal interpretation remains a crucial moment in the development of different legal AI systems:

- Legal Knowledge Engineering: How building rule-based AI systems or legal ontologies requires choices about meaning and scope of legal rules.
- Legal Data Annotation: How creating labelled datasets for predictive AI systems (e.g. for case outcome prediction, document classification, etc.) hinges on interpreters to apply definitions consistently, often requiring expert knowledge.
- Prompt Engineering: How the use of LLMs and other generative AI in law is affected by the way prompts are written, including specifying interpretive approaches or context, and how AI outputs may vary with different interpretive framings.

4. Interpretation and legal knowledge engineering

Legal knowledge engineering refers to the process of structuring and formalizing legal knowledge so that it can be processed by a computer – traditionally, this has involved encoding rules or building legal ontologies and knowledge bases.

One of the earliest and most famous examples was the formalization of the British Nationality Act 1981 as a logic program by Sergot et al. in 1986 [13]. This project demonstrated that a complex piece of legislation could be translated into executable logical rules, allowing a computer to infer, for example, whether a person is a British citizen given certain facts.

The approach taken was isomorphic to the statute's text: the knowledge engineers attempted to follow the structure and wording of the act as closely as possible, translating each provision quite literally into a logical proposition [14]. An

[13] M. J. Sergot et al., *The British Nationality Act as a Logic Program*, in *Communications of the ACM* 29, 1986, pp. 370–386.

[14] T. Bench-Capon, F.P. Coenen, *Isomorphism and Legal Knowledge Based Systems*, in *Artificial Intelligence and Law* 1, 1992, p. 75, which

isomorphic approach aims to preserve a one-to-one correspondence between the legal text and the program rules, aiming for transparency and ease of maintenance. If every section of a law directly maps to a rule in the knowledge base, it is easier for lawyers to verify the rules and update them when the law changes.

However, a fundamental challenge emerges: a strictly isomorphic encoding necessarily commits to one interpretation of each provision, usually the most straightforward or literal interpretation. In practice, many legal provisions are ambiguous or can be applied in different ways depending on context. Lawyers know that a single statute can often be given more than one meaning, yet a computer program, by design, will apply a fixed set of rules: as noted by Branting "a logical representation of a statutory rule is simply one interpretation among many possible interpretations" [15]. Thus, if an expert system encodes only what the drafters believe is the evident or straightforward meaning of each clause, it may perform adequately for routine cases but fail when a case requires a creative or less literal reading. Kowalski and Sergot, reflecting on these limitations, noted that their logic programs "can only indicate what follows from one precise interpretation of the law" and questioned "whether an expert lawyer would find such an ability useful" [16] beyond routine scenarios. In other words, a program that rigidly follows one interpretation might be of limited help to a savvy lawyer facing a hard case.

To address this, researchers in AI & Law proposed more sophisticated architectures that could handle multiple interpretations.

One strategy is to incorporate a layer of meta-knowledge or meta-rules that reason about the object-level rules. For example, instead of a single rule for a term like "vehicle" in the park, the system could include two or more rules reflecting different interpretations each represented together with an identifier of the interpretive canon or source justifying it.

---

holds that "even an inconveniently structured piece of legislation should have its structure respected".

[15] K. BRANTING, *Data-Centric and Logic-Based Models for Automated Legal Problem Solving*, in *Artificial Intelligence and Law*, 2017, pp. 5–27.

[16] R. KOWALSKI, M. SERGOT, *The Use of Logical Models in Legal Problem Solving*, in *Ratio Juris*, 1990, p. 212.

In this case, a rule might look like the following: "If an object has wheels and is mechanically powered, then classify it as a "vehicle" prohibited in the park (interpretive canon: literal meaning"). Under this reading, cars, motorcycles, and even electric scooters would fall within the prohibition. A different rule might instead be phrased as: "If an object is motorized and poses risks to safety or tranquillity, then classify it as a "vehicle" prohibited in the park" (interpretive canon: teleological, based on statute's intent to preserve peace and safety). On this interpretation, cars and motorcycles would be banned, but bicycles and wheelchairs would not.

By introducing such multi-interpretation frameworks, an expert system can simulate, to a degree, the reasoning a human expert might do in considering alternative arguments. For instance, if a fact pattern triggers two rules that represent conflicting interpretations of a statute, the meta-level reasoning can decide which interpretation prevails under the circumstances or alert the user to the divergence. Kowalski and Sergot suggested implementing mechanisms to manage these alternative logical models of the law, possibly by meta-level reasoning that keeps track of consistency and context for each interpretation [17].

Subsequent prototypes, such as Hamfelt's 1992 system [18], even attempted to encode principles of statutory interpretation themselves as meta-rules that could generate context-appropriate interpretations from a single abstract rule. Although fully realizing this vision is challenging (Hamfelt noted the difficulty of spelling out all the metarules that produce acceptable legal interpretations), the effort underlines how crucial interpretive theory is to knowledge engineering.

Another approach in legal knowledge engineering is the development of legal ontologies – structured semantic models of legal concepts and relationships [19]. Projects like the Functional Ontology of Law or LKIF Core attempted to

---

[17] R. Kowalski, M. Sergot, *The Use of Logical Models in Legal Problem Solving*, cit., p. 212.

[18] A. Hamfelt, *Metalogic Representation of Multilayered Knowledge*, PhD thesis, Uppsala University, 1994.

[19] On the use of computer ontologies for the legal domain, see G. Sartor et al., *Approaches to Legal Ontologies: Theories, Domains, Methodologies. Law*, Springer, 2011.

define common legal concepts (e.g., Person, LegalAct, Obligation) with formal properties and relations.

While ontologies are less procedural than expert system rules, creating them also involves interpretive choices: one must decide how to conceptually model doctrines and whether to adopt a broad or narrow definition of a concept. For example, should an ontology treat a "contract" as always involving exactly two parties (promisor and promisee), or can it be multilateral? The answer may depend on interpretive preference or jurisdictional definitions. These design decisions often draw on comparative law and legal theory to ensure the ontology is not encoding one idiosyncratic interpretation.

In practice, knowledge engineers collaborate with legal domain experts to validate that the representations accord with accepted interpretations. Even then, maintaining transparency about those choices is important so that users of the ontology or system know the assumptions. Tagging rules with their interpretive justifications, as mentioned, is one way to maintain transparency. Providing explanations is another: for instance, if an expert system concludes "Person X is entitled to citizenship", it should ideally explain why, e.g. "because under Interpretation A of the statute, X meets criteria Y and Z". Research on explanation in AI and Law has emphasized techniques like argument-based explanation that can trace the legal reasons and interpretive arguments behind a conclusion.

A concrete example highlighting the impact of interpretive choices is the use of AI in analysing regulations or codes for compliance. If a system is designed to flag non-compliance with, say, a financial regulation, it must encode the conditions of the regulation. A strict interpretation of a rule might flag more cases as violations (reducing false negatives but increasing false positives), whereas a lenient interpretation might do the opposite. Engineers must calibrate this, sometimes effectively deciding how "aggressive" the rule enforcement should be. This is both a technical and a legal interpretive decision. If the system's behaviour is contested (e.g., a company says the tool falsely labelled them non-compliant), it comes back to which interpretation was encoded. For accountability, one would need to review those design choices.

In summary, legal knowledge engineering demonstrates the inescapability of interpretation: representing law as code is not

like encoding the laws of physics. Our legal "laws" are products of language and social context, rife with open texture. The best AI systems in this area explicitly incorporate interpretive frameworks – allowing multiple paths and documenting which interpretive rule leads to which outcome – rather than hiding a single, rigid view of the law. This not only makes the systems more flexible and accurate, but also fairer and more transparent: users can see alternative outcomes under different interpretations, and the developers can be held accountable for why one interpretation was favoured in the coding. These principles carry over, albeit in different form, to the next topic: how we annotate legal data for AI.

5. Interpretation and legal data annotations

Many modern AI applications in law rely on machine learning, which in turn depends heavily on data [20].

Legal data annotation is the process of labelling legal texts or data to create datasets for training or evaluating AI models. Examples include annotating court decisions with outcome labels (win/lose), marking segments of an opinion as the holding versus obiter dicta, tagging contract clauses by type (e.g. indemnity clause, force majeure clause), or classifying statements in a brief as factual or legal arguments.

Unlike general-domain data labelling (e.g., tagging images of cats vs. dogs), legal annotation projects face unique challenges due to the specialized knowledge and interpretive nuance of legal documents [21].

To ensure high-quality data, researchers stress the importance of clear annotation guidelines that map legal knowledge into consistent labelling rules. Guidelines serve as the roadmap for maintaining uniformity in how annotators handle the subtle nuances of legal terminology and reasoning.

For example, a guideline might instruct: "If a sentence

[20] See K. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, 2017, p. 234 for the different tasks where machine learning is applied to legal data and texts.

[21] An overview of these challenges is provided in H. Darji et al., *Challenges and Considerations in Annotating Legal Data: A Comprehensive Overview*, arXiv Preprint arXiv:2407.17503, 2024.

contains a citation and states a general principle of law, tag it as a *holding*; if it is specific to the case's facts, tag it as *dicta*". Such rules require judgment calls – what if the sentence is phrased generally but is actually an application to facts? The guideline might then give examples and exceptions.

The process of developing these guidelines is itself an exercise in distilling legal interpretations: the project leaders (often legal scholars or experienced practitioners) have to articulate how to consistently interpret various textual signals as particular labels. Involvement of legal professionals is thus valuable in ensuring that the data not only remains contextually accurate but also adheres to prevailing legal standards and interpretations. In practice, this means that annotation teams often include or consult with attorneys to clarify tough cases.

For instance, if annotators must classify whether a set of facts meets the legal definition of "negligence" in a tort case, they must apply the interpretation of negligence elements (duty, breach, causation, damage) as defined by law. If the case is borderline (was there a duty or not?), even courts might split on it; the annotators have to use their best legal judgment or follow the majority rule in their jurisdiction.

There is a second role of interpretation in legal data annotation, when annotators apply annotation guidelines to legal documents.

Take the task of labelling a sentence in a judgment as the legal reasoning versus background fact. This can be tricky – it might require understanding the context of the whole case and the legal significance of that sentence. Non-experts could easily mislabel content because they would not grasp what the judge is actually doing in that line of text. This is the reason why legal annotation typically requires significant expertise to identify complex phenomena such as modes of judicial reasoning or controlling precedents. The annotators are not interchangeable crowd-workers, but domain experts, and even they might disagree due to interpretive ambiguities in the material.

In other words, this second level of interpretation consists in deciding on the extension of a label in practice – that is, whether a particular word, sentence, or passage actually falls within the annotation class defined in the guidelines. For example, one annotator might decide that a sentence introducing a principle with a citation belongs to the category "holding", while another might see it as part of the case's factual background. Likewise,

a clause in a contract may be tagged as "force majeure" by one annotator but as a general limitation of liability by another, depending on how they read its wording and context.

Even with detailed guidelines, ambiguity often remains: what looks like background narrative to one annotator may look like legal reasoning to another. The act of applying a label is therefore never a purely clerical step; it is an interpretive judgment about how abstract legal categories map onto the messy reality of legal texts. In this way, annotation is not simply data processing but an act of legal reasoning at the micro-level of each labelled instance.

From a fairness and accountability perspective, how interpretation is carried out both at the abstract (guidelines) and the concrete level (annotation) is critical. If the training data for a legal AI system encodes a skewed interpretation, the model will learn that bias. For instance, imagine a dataset for a case outcome prediction model where the annotators labelled all cases involving a certain controversial law in a particular way because they personally interpreted the law strictly. The model might learn that stance and consistently predict outcomes in line with that strict interpretation, even if courts are actually divided. This could unfairly disadvantage some litigants if the AI is used to assess chances of success. It is therefore important that annotation projects strive for objectivity and balance, perhaps by reflecting majority jurisprudence or by including features that let the model know about differing opinions.

In certain NLP tasks annotation guidelines may even accommodate different interpretations. For instance, if the goal is to measure linguistic phenomena in a large legal corpus for analytics, guidelines can be purely descriptive: they simply record features of the text such as part-of-speech tags, syntactic structures, or named entities. In these cases, multiple interpretations may be accepted and recorded, since the aim is to capture variation rather than enforce a single authoritative reading.

In the data annotation for legal machine learning, however, annotation guidelines are rarely descriptive[22]. They are

[22] On the distinction between descriptive and prescriptive guidelines, see P. Rottger et al., *Two Contrasting Data Annotation Paradigms for Subjective NLP Tasks*, in *Proceedings of the 2022 Conference of the North American*

prescriptive and normative in their content: they do not just capture what is on the page but stipulate how a legal concept should be recognized and applied. Deciding what counts as a "holding", a "cause of action", or a "risk clause" means fixing the meaning of those categories in practice. In this sense, annotation guidelines operate as micro-doctrines, instructing annotators to treat certain textual instances as falling under specific legal classifications, while excluding others.

This is the reason why it is important to keep track of disagreements between annotators [23]. Higher disagreement rates are common on hard items and often reflect genuine legal ambiguity. In such instances, projects sometimes employ adjudication by a senior legal expert to decide the correct label, or they allow multiple labels to co-exist, representing different plausible interpretations, if the ML approach can accommodate uncertainty.

Inter-annotator agreement (IAA) is usually measured to assess consistency [24]. Low IAA often signals that interpretive ambiguity exists and may require more refined guidelines or better training for annotators. But in some cases, low agreement is simply a reflection of the fact that reasonable jurists themselves disagree on the issue. Those are often the very areas where AI predictions or classifications should be used with caution, and ideally accompanied by explanations, confidence scores, or markers of interpretive uncertainty.

Legal data annotation is also relevant beyond supervised learning: it matters in creating benchmark datasets for evaluation. For example, in the realm of legal question answering, one might annotate the correct answer to a legal question and the supporting rationale. If those annotations favour one interpretive style (say, a very literal reading of

---

Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2022.

[23] D. BRAUN, *I Beg to Differ: How Disagreement Is Handled in the Annotation of Legal Machine Learning Data Sets*, in *Artificial Intelligence and Law*, 2024, pp. 839–862.

[24] One of the most used measures for calculating agreement between annotators is the Kappa coefficient (K), specifically Cohen's Kappa in the case of two annotators, or Fleiss' Kappa for more than two. This coefficient considers the agreement that could occur by pure chance, thus providing a more realistic measure of reliability. The value of K ranges from -1 to 1, where 1 indicates perfect agreement, 0 indicates agreement equivalent to chance, and negative values indicate systematic disagreement.

statutes), a system that uses a more purposive reasoning might be unfairly penalized in evaluation, or vice versa. Thus, the evaluation metrics for legal AI can inadvertently embed an interpretive bias. Recognizing this, some benchmarks include multiple "acceptable" answers or ask evaluators to rate the answers' justifications rather than absolute correctness.

In sum, legal interpretation pervades the annotation process. The quality and reliability of AI in law depend directly on the quality of the labelled data we use, which in turn hinges on humans applying legal reasoning consistently. Best practices emerging in this area include: involving domain experts and training them in annotation, developing rigorous guidelines with examples, conducting pilot annotations to refine interpretive rules, and measuring agreement and analysing disagreements. By doing so, we aim to create datasets that reflect sound legal interpretations, thereby training AI models that behave in ways aligned with expert understanding of the law.

6. Interpretation and prompt engineering

With the advent of powerful large language models like GPT-4, Claude, and others, a new paradigm has gained prominence in legal AI: using general-purpose AI models to perform legal tasks via appropriate prompts.

Prompt engineering is the art and science of crafting inputs to AI models to elicit desired outputs [25]. In the legal context, prompt engineering might involve, for example, asking an LLM to "summarize the attached case and highlight any interpretation of 'vehicle' in park regulations" or "draft a brief arguing that the Third Amendment applies to state governors". Because LLMs do not truly understand law but generate responses based on patterns in their training data, how one formulates the request can significantly influence the answer's quality and correctness.

The role of legal interpretation in prompt engineering emerges here in two ways. First, the prompt may explicitly

[25] L.S. Lo, *The Art and Science of Prompt Engineering: A New Literacy in the Information Age*, in *Internet Reference Services Quarterly*, 2023, pp. 203–210.

invoke particular interpretive frames to guide the model to provide the expected answer.

For instance, if LLMs are used to detect dicta in judicial opinions, the prompt might instruct the model to "identify statements that are not strictly necessary for the holding of the case but illustrate the judge's reasoning or provide background". Such a formulation presupposes a particular understanding of the doctrine of dicta and guides the model toward that interpretive stance.

This is not simply confined to abstract "legal theories"; it is also a matter of pragmatic decision. For example, when LLMs are prompted to summarise judgments, the prompt should specify which parts of the decision are to be included in the summary – facts, holdings, reasoning, or dissenting opinions – and whether quotations from the judgment should be reproduced verbatim or paraphrased. Each of these choices reflects interpretive and practical judgments about what counts as salient legal content and how it should be represented.

A parallel can be drawn between prompt engineering and annotation guidelines: both embed legal interpretation by establishing a framework for how the system should treat legal texts. However, while annotation guidelines fix interpretive categories in advance to be applied consistently across a dataset, prompt engineering operates directly and dynamically, shaping the model's interpretive stance in real time depending on the question posed and the context provided. Annotation constrains interpretation *ex ante*; prompting steers it *ex post*. Both, however, reveal that interaction with legal AI systems is never interpretation-free but always mediated by human decisions about meaning, scope, and relevance.

The second role of interpretation is more nuanced and subtle, and it represents a novel challenge for AI and Law studies. This aspect concerns the sensitivity of LLM outputs to training data and prompt phrasing: the model may implicitly adopt one interpretation or another depending on slight variations in the wording of texts used for generating an answer.

A striking example comes from experiments by Coan and Surden on AI and constitutional interpretation [26]. They posed

[26] A. COAN, H. SURDEN, *Artificial Intelligence and Constitutional Interpretation*, in *University of Colorado Law Review*, 2025, pp. 414–501.

the same legal question to two different LLMs – OpenAI's ChatGPT and Anthropic's Claude – asking whether the U.S. Third Amendment (which forbids quartering of soldiers in private homes in peacetime) would bar a state governor from quartering themselves in a private home without consent. One might expect a straightforward answer, but the two AI systems diverged: ChatGPT answered "no" reasoning that a governor is not a "soldier" under the Amendment, whereas Claude answered "yes" reasoning that the Amendment's purpose was to prevent exactly the type of government intrusion that forcing any official into a private home would entail.

In essence, while ChatGPT took a literal/textual interpretive stance (soldier is just a soldier, nothing else), Claude took a purposive/teleological stance, focusing on the Amendment's broader protective purpose. Notably, neither answer was 'wrong' in any straightforward sense – each reflects a defensible interpretation that a human jurist could also argue for. These are precisely the kinds of divergent interpretations that appear in real constitutional debates, yet here they arose from AI systems presumably without explicit instruction to adopt one theory or another.

This example underscores a few important points. First, LLMs do perform a form of interpretation when confronted with legal questions, but it is an implicit interpretation derived from patterns in their training data. One model might have seen more texts emphasizing literal readings, another more texts emphasizing purpose, or they might just probabilistically lean one way or the other. The key difference is that when AI systems make such interpretive choices, they do so invisibly through complex statistical computations that even AI experts do not fully understand.

Consequently, a judge or lawyer using the LLM might get an answer without realizing what tacit value judgments or assumptions the model has made. In the Third Amendment scenario, if a user only asked one AI model, they might have gotten a single answer and assumed "the AI says governors are not covered, so that's the objective truth", whereas another model would have given the opposite result. This highlights the transparency problem: AI doesn't announce, "I applied a purposive canon here", it just gives an output.

From a prompt engineering perspective, one partial remedy is to explicitly ask the model to explain its reasoning or consider

alternatives. For example, one could prompt: "Explain whether a state governor could be considered a 'soldier' for purposes of the Third Amendment. Provide arguments for a narrow (textual) interpretation and a broad (purposive) interpretation, then give your conclusion". This kind of instruction forces the model to engage with multiple viewpoints and reveal its reasoning process, making the interpretive step more explicit to the user.

Second, the sensitivity of LLMs to how questions are framed means legal practitioners must be cautious and thoughtful in prompt design. As one commentary noted, LLM outputs on legal issues are "highly sensitive to how questions are framed, and which interpretive approaches are specified". The phenomenon of "AI sycophancy"[27] is also relevant: models tend to mirror the user's implied preferences. If a prompt subtly suggests a desired answer (even unintentionally through wording), the model might slant its interpretation that way.

For instance, asking "Don't you agree that the Third Amendment's historical context would exclude governors?" is likely to yield a different answer than a neutral phrasing. Therefore, legal scholars and professionals using these tools are advised to craft neutral, balanced prompts – or deliberately test multiple prompts – to see if the answers remain consistent. Coan advises practices such as "testing questions through multiple framings to assess consistency, explicitly requesting competing perspectives, and asking LLMs to articulate their embedded assumptions".

These align with general prompt engineering tactics like chain-of-thought prompting, where you ask the model to reason step by step, or role prompting, where you ask it to act as a devil's advocate or as a judge employing a certain method of interpretation.

In effect, prompt engineering becomes a new form of legal skill[28]. Lawyers must learn to "speak" to AI in a way that leverages their legal knowledge.

For example, to use an AI for statutory interpretation, a

---

[27] M. SHARMA et al., *Towards Understanding Sycophancy in Language Models*, arXiv Preprint – arXiv:2310.13548, 2023.

[28] P. MÜLLER-PELTZER et al., *Legal Prompt Engineering and How It Will*

lawyer might include the statutory text in the prompt, summarize any relevant precedents, and then direct the model: "Analyse this problem using a textualist approach" versus "... using a purposive approach" and compare the results [29]. Early experiences show that including more context in prompts tends to produce answers that are better grounded. By adding details like jurisdiction, facts of the case, or definitions of terms, "the LLM can generate content that is not only accurate but also deeply informed by existing legal knowledge".

The practice of enhancing a prompt by incorporating additional, contextually relevant information that aids the AI model in generating more accurate and reliable output is defined as "prompt augmentation". It involves embedding supplementary data—either preceding or following the query—that the model can utilize during inference [30]. Prompt augmentation significantly improves the model's performance as it provides a richer contextual basis of information for its reasoning and prediction tasks.

The approach is akin to giving the model the pieces of the puzzle it might otherwise guess from general training data. For instance, if asking about a term's ordinary meaning at the time of a law's enactment, one could supply dictionary definitions from that era as part of the prompt. This form of retrieval-augmented prompting ensures that the AI isn't drawing purely from its possibly biased memory, but from authoritative sources supplied by the user.

There are already practical tools and efforts embracing prompt engineering for legal use. Some law firms have experimented with in-house "prompt libraries" – templates for tasks like contract review or brief drafting that their lawyers can fill in with specifics. Products like Thomson Reuters' CoCounsel (Casetext) and others provide interfaces where a lawyer can select an action (e.g., find contradictions between

---

Change the Way Lawyers Work, in K. Jacob et al. (eds.), *Liquid Legal–Sustaining the Rule of Law*, Springer, 2025.

[29] See Y. Arbel, D. A. Hoffman, *Generative Interpretation*, in *New York University Law Review*, 2024, discussing the perspective on "generative interpretation", i.e. the use of generative AI models to propose possible interpretation of legal contents.

[30] H. Surden, *ChatGPT, AI Large Language Models, and Law*, in *Fordham Law Review*, 2023, p. 1969.

these two contracts) which effectively translates into a carefully engineered prompt behind the scenes.

Despite these tools, recent evaluations counsel vigilance. An empirical study in 2025 tested several leading AI legal research assistants (from vendors like LexisNexis and Westlaw) that rely on LLMs with prompt-based querying. The providers claimed their systems would be "hallucination-free" [31] by using techniques like retrieval augmented generation, which adds relevant documents to the prompt to ground the answer. Indeed, these systems did reduce the rate of wildly made-up answers compared to a raw GPT-4. Yet, the study found each tool still produced hallucinated legal citations or false statements about 17% to 33% of the time.

This underscores that prompt engineering, while helpful, is not a panacea; the models have inherent limitations [32]. From an interpretive standpoint, it also means a user might get a confidently delivered answer with citations that appear genuine but are entirely fictitious – essentially an AI misinterpretation or misrepresentation of law. The burden remains on the human legal professional to verify AI outputs and ensure they make sense.

To use LLMs responsibly in legal settings, experts recommend a combination of AI literacy and institutional safeguards. Lawyers and judges should be trained to understand things like the randomness in LLM responses, the importance of prompt wording, and the fact that outputs can change with even minor rephrasing.

For instance, an attorney might learn that if an answer seems too pat, they should try rewording the question to see if the model changes its tune – if it does, that signals the issue is sensitive to interpretation. Judges might develop a practice of asking an AI for both sides of an argument explicitly, to avoid one-sided answers. On the institutional side, courts and firms

---

[31] "Hallucination", in the context of large language models, refers to the phenomenon whereby the model generates information that is factually incorrect or fabricated, yet appears plausible and coherent within the given context. See H. YE et al., *Cognitive Mirage: A Review of Hallucinations in Large Language Models*, arXiv Preprint – arXiv:2309.06794, 2023.

[32] See, for instance, the discussion in E. MIK, *Caveat Lector: Large Language Models in Legal Practice*, in *Rutgers Business Law Review*, 2023, p. 70.

are considering guidelines: e.g., requiring that any AI-drafted material be reviewed by a human, or even court rules mandating disclosure if a filing was prepared with AI assistance. The idea is not to ban the use of AI, but to ensure it is "used as a tool to augment human legal reasoning rather than replace it entirely".

In conclusion, prompt engineering in the legal domain is where the rubber meets the road in human-AI interaction for legal interpretation. It is an area where legal interpretive acumen and technical skill intersect. By carefully crafting prompts and critically evaluating answers, legal professionals can harness LLMs to explore interpretations and generate draft analyses, while still injecting the crucial oversight of human judgment. The interpretation is inescapable: if we don't guide the AI, it will implicitly choose an interpretive path on its own (drawn from data), which may be hidden or suboptimal. Thus, the lesson is to engage with these models thoughtfully – to prompt them in a way that surfaces the interpretive dimensions and to remain alert to the normative choices that even a statistical model's output can embody.

7. The end of legal interpretation or a new beginning?

The question is whether AI heralds the end of legal interpretation or a new beginning. On one hand, the rhetoric of automation and objectivity tempts us to believe that interpretation is being displaced by technical computation.

On the other hand, as this chapter has shown, interpretation does not vanish but rather reappears in new forms: embedded in annotation guidelines, encoded in rule-based systems, hidden in training data, and shaped through prompt design. What is truly new is not the disappearance of interpretation, but its relocation from the visible reasoning of judges and lawyers to the invisible architectures of algorithms.

Coan and Surden encapsulate this insight as the "law of conservation of judgment" [33]: the introduction of AI does not eliminate the need for human normative judgment in legal

[33] A. Coan, H. Surden, *Artificial Intelligence and Constitutional Interpretation*, cit., p. 494.

interpretation; it simply shifts where and how that judgment is exercised. For example, instead of a judge openly deciding between two interpretations of a statute, we might have an engineer deciding which interpretation to encode in a system, or an LLM implicitly deciding based on its training data – but the decision is still being made by someone, somewhere.

This shift does not end interpretation; it challenges us to recognize and make explicit the interpretive choices that are now distributed across human and machine agents [34]. In this sense, the rise of AI in law can be understood as a new beginning for legal interpretation – one that forces lawmakers, legal scholars and practitioners to confront, with renewed clarity, the foundational role of interpretation in shaping both human and artificial legal reasoning.

This shift carries significant implications for fairness, transparency, and accountability in legal AI. Fairness requires that AI systems do not systematically disadvantage any party or perspective. Yet if an AI unknowingly adopts one interpretive stance – for instance, consistently favouring a strict reading of asylum law to the detriment of applicants – it risks embedding structural bias. The remedy lies in awareness and balance: incorporating multiple viewpoints, testing systems across varied scenarios, and correcting for skewed training data.

Transparency is equally crucial. Legal AI must be able to explain or reveal its reasoning to be trustworthy. When an AI makes an interpretive leap "invisibly" it undercuts the ability of parties to understand or contest the result. This is why scholars advocate explainable AI in legal contexts – systems that highlight which facts and rules were most influential, or that can indicate: "I treated term X as including Y based on Z source". In knowledge-based systems, this might be achieved by tagging rules with sources and interpretive labels. In

[34] Cf. M. HILDEBRANDT, *Qualification and Quantification in Machine Learning. From Explanation to Explication*, in *Sociologica*, 2022, pp. 37–49 and M. HILDEBRANDT, *Boundary Work Between Computational 'Law' and 'Law-as-We-Know-It'*, in D. CURTIN, M. CATANZARITI (eds.), *Data at the Boundaries of European Law, Collected Courses of the Academy of European Law*, Giuffrè, 2023, arguing for a "new legal hermeneutics" which emphasizes the need to critically examine how legal meaning is mediated by data infrastructures and computational architectures.

machine learning systems, it might involve post-hoc explanations or interactive dialogues where the AI can be asked to justify its answers. Current research in explainable AI for law is experimenting with methods ranging from highlighting relevant passages in legal documents to generating natural language justifications aligned with legal principles.

Accountability likewise demands that responsibility for AI-assisted legal reasoning is not diffused or obscured. This could involve maintaining audit trails for automated decisions or enforcing regulations such as the EU AI Act[35], which classifies AI used in justice as "high-risk" and subjects it to strict requirements of transparency, human oversight, and accuracy. A practical safeguard is the human-in-the-loop principle: even if a court uses an AI tool to draft a summary or recommend a sentence, a human judge must review and ultimately accept, modify, or reject the output. As one judicial guideline notes, "ensuring that AI-generated recommendations remain transparent and subject to human oversight is crucial in maintaining judicial accountability".

This speaks to a broader theme: AI in law should augment human interpretative endeavours, not replace it. Properly used, AI can enhance consistency, for example by flagging when similar cases have been treated differently or when statutory language has been applied inconsistently across jurisdictions. It can improve efficiency by rapidly retrieving relevant authorities, summarising complex judgments, or detecting patterns in large volumes of legal texts that would otherwise remain hidden. It can also broaden access to justice by supporting affordable legal assistance for individuals and small organisations who might not otherwise have access to professional advice.

To maximize the benefits and mitigate the risks, interdisciplinary collaboration is key. Legal experts, AI

[35] 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, OJ L, 2024/1689, 12.7.2024.

engineers, and ethicists need to jointly develop standards and best practices. For instance, building evaluation frameworks for legal AI that go beyond aggregate accuracy, to measure whether the system's reasoning aligns with sound legal interpretation. Recent work on benchmarks often includes not just the correctness of answers but whether the justifications cite appropriate sources.

Likewise, there is a push for open datasets and challenges in areas like statutory reasoning, so that the research community can compare methods and identify where models fall short of human interpretive abilities. One emerging consensus is that purely statistical models will struggle with deeper legal reasoning unless they are integrated with legal domain knowledge – whether through hybrid systems that use logic and ML, or via training that includes legal reasoning tasks. The complexity of legal interpretation – factoring in evolving principles, societal values, and the uniqueness of fact contexts – is a known stumbling block for predictive models that just mimic patterns. Addressing this may require novel AI architectures that can handle not just language but also reasoning with rules and cases – a classical focus of AI & Law research.

Finally, education and literacy are critical. Tomorrow's lawyers and judges must be conversant not only with legal doctrine, but with the basics of how AI systems function and where they might go wrong. Law schools are beginning to include AI literacy, and bar associations are issuing ethics opinions on the use of AI (for example, warning attorneys that using an AI without understanding its limitations could violate duties of competence and deontology). Prompt engineering might even become a standard skill, akin to legal research skills. This chapter's discussions on knowledge engineering, annotation, and prompt design all illustrate that people remain at the heart of making AI in law effective and just.

We conclude that the role of legal interpretation in legal AI is both profound and unavoidable. Rather than viewing AI as a threat to the nuanced human practice of legal interpretation, we can view it as a catalyst that forces us to articulate and encode our interpretive principles more clearly than before. This process can lead to valuable insights and improvements in consistency. But we must proceed with humility: law is not merely data to be mined or code to be executed; it is a human

enterprise of reason-giving and value alignment. AI, no matter how impressive, is ultimately a tool that should serve that enterprise. By designing AI systems that respect and incorporate the richness of legal interpretation – and by ensuring human oversight and accountability – we can harness technology to enhance the rule of law, rather than inadvertently undermining it. The chapters ahead in this volume will no doubt continue this conversation, examining specific methodologies and case studies in greater detail, but the foundational understanding remains: legal interpretation is the lens through which AI in law must be focused to achieve fair, transparent, and accountable outcomes.

CHAPTER III

THE ROLE OF EUROPEAN CASE LAW IN THE SYSTEM OF VALUE ADDED TAX

Eleonor Kristoffersson – Magnus Kristoffersson*

CONTENTS: 1. Introduction. – 2. The EU VAT system. – 3. The role of the Court of Justice of the European Union. – 4. The relation between the CJEU and the national courts – 5. Recent and future development in the AI era – 6. Concluding remarks.

1. Introduction

This chapter explores the role of European case law in shaping the system of value added tax (VAT) within the European Union (EU). While several European countries lie outside the EU, the scope of this study is limited to EU Member States. This limitation is justified by the existence of a common judicial institution, the Court of Justice of the European Union (CJEU), which plays a central role in interpreting and developing VAT law across the Union.

Historically, the CJEU has been the primary source of case law in the EU VAT domain. However, beginning in 2024, the General Court has also been granted jurisdiction over VAT-related preliminary rulings. Although jurisprudence from the

* Eleonor Kristoffersson is Full Professor of Tax Law at the Örebro University; Magnus Kristoffersson is Associate Professor of Tax Law at the Örebro University. The authors have contributed equally to this work. Beside the EU funding, the working hours of Eleonor Kristoffersson were partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

General Court remains limited at this early stage, its decisions are subject to appeal before the CJEU, and its role is expected to grow in significance over time.

The chapter is structured as follows: Section 2 provides an overview of the EU VAT system and its legislative foundations. Section 3 examines the interpretative role of the CJEU, followed by Section 4, which analyses the relationship between the CJEU and national courts. Section 5 discusses recent and anticipated developments in the field, particularly in light of advances in artificial intelligence (AI). The chapter concludes with the authors' final reflections on the implications of European case law for the future of VAT.

2. The EU VAT system

A common VAT system was introduced in what was then the European Community (EC) and what is now the EU in 1967 through the First and Second VAT Directives [1]. The objective of the First VAT Directive was that all Member States should replace their various turnover taxes with VAT [2]. According to article 2 section 1 of the First VAT Directive, the common system of VAT involves the application to goods and services of a general tax on consumption exactly proportional to the price of the goods and services, whatever the number of transactions that take place in the production and distribution process before the stage at which tax is charged. Under Article 2 Section 2 of the First VAT Directive, VAT shall be chargeable on each transaction, calculated on the price of the goods and services at the rate applicable to such goods or services, after a deduction of the amount of VAT borne directly by the various cost components. VAT shall be

[1] First Council Directive of 11 April 1967 on the Harmonisation of Legislation of Member States concerning Turnover Taxes (67/227/EC) and Second Council Directive of 11 April 1967 on the Harmonisation of Legislation of Member States concerning Turnover Taxes. Structure and Procedures for Application of the Common System of Value Added Tax (67/228/EC). Section 2 in this chapter builds upon a historical review in E. KRISTOFFERSSON, I. JEONG, *A Comparison between EU and Korean VAT from a Legal Transplants Perspective*, in *Nordic Tax Journal*, 2024, https://doi.org/10.2478/ntaxj-2024-0001.

[2] Article 1 of the First VAT Directive.

applied up to and including the retail stage, but initially, there were particular possibilities for the Member States for a transition period to combine other retail taxes with VAT up to and including the wholesale trade stage [3].

The Second VAT Directive regulates the structure and the procedures for applying VAT [4]. Accordingly, the supply of goods and the provision of services within the territory of the country by a taxable person against payment as well as the importation of goods shall be subject to VAT [5]. The Second VAT Directive defines "the territory of the country" [6], "taxable person" [7], "supply of goods", "provision of services" [8], and "importation of goods" [9]. It also contains rules regarding the basis of assessment [10], tax rates [11], exemptions [12], deductions [13], and requirements of records of account [14]. Annex A of the Second VAT Directive contains explanations of the articles in the directive.

A decision from April 21,1970 provided that the budget of the European Community (EC) (now EU) should be financed entirely from the EC's own resources [15]. These resources include tax income accrued from VAT [16]. In order to increase the fairness between the member states, there was a need for a greater harmonization of the common VAT system. This was

---

[3] Article 2, Sections 3 and 4 of the First Council Directive of 11 April 1967 on the Harmonisation of Legislation of Member States concerning Turnover Taxes (67/227/EC) and Second Council Directive of 11 April 1967 on the Harmonisation of Legislation of Member States concerning Turnover Taxes. Structure and Procedures for Application of the Common System of Value Added Tax (67/228/EC).

[4] Article 1 of the Second VAT Directive.

[5] Article 2 of the Second VAT Directive.

[6] Article 3 of the Second VAT Directive.

[7] Article 4 of the Second VAT Directive.

[8] Articles 5 and 6 of the Second VAT Directive.

[9] Article 7 of the Second VAT Directive.

[10] Article 8 of the Second VAT Directive.

[11] Article 9 of the Second VAT Directive.

[12] Article 10 of the Second VAT Directive.

[13] Article 11 of the Second VAT Directive.

[14] Article 12 of the Second VAT Directive.

[15] OJ No L 94, 28. 4. 1970, 19.

[16] See M. LAMENSCH, *European Value Added Tax in the Digital Era: A Critical Analysis and Proposals for Reform*, IBFD Doctoral Series, 2015, p. 11.

one of the main reasons for the introduction of the Sixth VAT Directive from May 17, 1977[17]. Before the Sixth VAT Directive, the treatment of agriculture, the deduction of import taxes, the exemptions and the supply of services across borders were far from harmonized[18]. Many of the explanations in Annex A of the Second VAT Directive were incorporated in the different articles of the Sixth VAT Directive, which gave them a higher legal status. The Sixth VAT Directive contains much more detailed rules than the Second VAT Directive, creating the basis of today's EU VAT system[19]. In 2006, the approximately 40 long and complicated articles of the, by then many times amended, Sixth VAT Directive, were transferred into more than 400 shorter and much more accessible Articles of the current VAT Directive[20].

Besides the VAT Directive, the Implementing Regulation (EU) 202/2011 is an important source of EU VAT law[21]. Other than the VAT Directive, the Implementing Regulation is directly applicable, without implementation into national law. The objective of the Implementing Regulation is to ensure uniform application of the VAT system[22].

3. The role of the Court of Justice of the European Union

The CJEU shall ensure that in the interpretation and application of the treaties, the law is observed.[23] The role of the CJEU in the VAT system is to explain and interpret the rules of

---

[17] Sixth Council Directive 77/388/EEC of 17 May 1977 on the Harmonization of the Laws of the Member States Relating to Turnover Taxes - Common System of Value Added Tax: Uniform Basis of Assessment.

[18] Fehr, Rosenberg, and Wiegard, 13.

[19] European Commission. 1973. Explanatory Memorandum to the Proposal for a Sixth Council Directive on the Harmonization of Member States Concerning Turnover Taxes. Common System of Value Added Tax: Uniform Basis of Assessment, COM (73) 950, 20 June 1973. *Bulletin of the European Communities* 11/73, 7.

[20] Council Directive 2006/112/EC of 28 Nov. 2006 on the common system of value added tax (VAT Directive).

[21] Council Implementing Regulation (EU) 282/2011 of 15 March 2011 laying down implementing measures for Directive 2006/112/EC on the common system if value added tax (Implementation Regulation).

[22] Paragraph 4 of the preamble of the Implementation Regulation.

[23] Article 19 of the Treaty of the European Union (TEU).

the VAT Directive.[24] Most of the cases before the CJEU are preliminary rulings. Where a question regarding the validity and interpretation of acts of institutions of the EU, such as the VAT Directive, is raised before any court or tribunal of a Member State, the national courts may, if it considers that a decision on the question is necessary to enable it to give judgment, request the CJEU to give a ruling thereon[25]. Where such a question is raised in a case pending before a court against whose decisions there is no judicial remedy under national law, that court or tribunal shall bring the matter before the CJEU[26]. This is interpreted by the CJEU as meaning that a court of last instance is obliged to bring the matter before the CJEU, unless it has established that the question raised is irrelevant, or that the provision in question has already been interpreted by the CJEU or that the correct application is so obvious as to leave no scope for any reasonable doubt[27]. The CJEU commits to provide the national court with an answer which will be of use to it and enable it to determine the case before it[28].

The strict requirements of the CJEU on the national courts to ask for a preliminary ruling as well as the great number of VAT cases before national courts have led to a heavy burden on the CJEU[29]. Thus, as of October 2024, the General court has jurisdiction to hear and determine requests for a preliminary ruling under that come exclusively within the common system of value added tax[30]. The request can also include one or more of the areas of excise duties, the custom duties, the tariff

---

[24] A. VAN DOESUM, H. VAN KESTEREN, S. CORNELIE, F. NELLEN, *Fundamentals of EU VAT LAW*, 2025, Wolters Kluwer, par. 1.9.1.

[25] Article 276 of the Treaty of the Functioning of the European Union (TFEU).

[26] Article 276 TFEU.

[27] CJEU, 6 October 1982, *Cilfit*, C-283/8, ECLI:EU:C:1982:335. For criticism, see A. LIMANT, *Recent Developments in the Acte Clair Case Law of the EU Court of Justice: Towards a more Flexible Approach*, in *Journal of Common Market Studies*, 2016, pp. 1384–1397.

[28] CJEU, 25 July 2018, Dyson, C-632/16, ECLI:EU:C:2018:599. See also S. CAZES, *Positive Outcome for Businesses of Recent VAT Casescfrom the Court of Justice of the European Union*, in *Intertax*, 2020, p. 934.

[29] K. POSCH, *Conference Report, 'Court of Justice of the European Union: Recent, VAT Case Law 2024' In Which Direction Does the CJEU Go?*, in *EC Tax Review*, 2024, p. 139.

[30] Article 50b Sec. 1 of the Protocol (No 3) to the TFEU on the Statute of the Court of Justice of the European Union.

classification of goods under the Combined Nomenclature, compensation and assistance to passengers in the event of denied boarding or of delay or cancellation of transport service and the system for greenhouse gas emission allowance trading [31]. The request for preliminary rulings shall still be submitted to the CJEU. However, after verifying that the request for a preliminary ruling fall exclusively within one or more of the above-mentioned areas the CJEU shall transfer that request to the General Court. The CJEU retains jurisdiction if the questions go beyond VAT or the other abovementioned areas or relate to the interpretation of primary law, public international law, general principles of Union law or the Charter of Fundamental Rights of the European Union [32]. The rulings by the General Court can be appealed to the CJEU [33]. The effects of the 2024 reform still remain to be seen [34]. One positive effect could be, that the General Court develops into a EU tax court, with high specialization and competence within the field [35].

The other option of VAT cases to reach the CJEU is through the infringement procedure [36]. An infringement procedure is normally initiated by the commission [37], but can also be initiated by another member state [38]. The commission, in its capacity of being a watchdog over EU legislation, has a role in both cases. When the commission initiates an infringement procedure, it shall deliver a reasoned opinion on the matter after giving the member state the opportunity to submit its observations [39]. Only if the member state does not comply with the opinion, the commission may bring the matter before the CJEU [40]. In the rare cases when one member state initiates an infringement procedure against another member state it

---

[31] Article 50b Sec. 3 of the Protocol (No 3) to the TFEU.

[32] Article 50 b Sec. 2 of the Protocol (No 3) to the TFEU.

[33] Article 56 of the Protocol (No 3) to the TFEU.

[34] K. POSCH, *Conference Report, Court of Justice of the European Union: Recent, VAT Case Law 2025: Conference in Vienna. The General Court Steps into the Spotlight*, in *EC Tax Review,* 2025, p. 175.

[35] K. POSCH, *Conference Report, 'Court of Justice of the European Union: Recent, VAT Case Law* cit., p. 139.

[36] Articles 258 and 259 TFEU.

[37] Article 258 TFEU.

[38] Article 259 TFEU.

[39] Article 258 Sec. 1 TFEU.

[40] Article 258 Sec. 2 TFEU.

shall bring the matter to the commission [41]. Thereafter, the commission submits it reasoned opinion [42]. After that, as well as if the commission has not delivered an opinion within three months, the matter can be brought before the CJEU [43].

In both preliminary rulings and infringement procedures, the CJEU interprets the VAT directive. An infringement procedure results in a judgement against a member state, where it is declared that a member state has or has not failed to fulfil its obligations under the VAT Directive [44]. Such a judgement can concern both application and interpretation of the VAT Directive, but normally, there are different opinions about how the VAT Directive should be understood, which concerns interpretation rather than application on an actual case. Preliminary rulings always concern the interpretation of the VAT Directive. It is always up to the referring national court to apply the interpretation and decide upon the tax dispute. Consequently, the main role of the CJEU is to provide the member states and the courts with guidance on how the VAT Directive should be interpreted [45].

4. The relation between the CJEU and the national courts

In the beginning of the history of the common European VAT, the interaction between national courts and the CJEU was not taken for granted. The famous Swedish tax scholar *Gustav Lindencrona*, states in his doctoral thesis from 1972:

"Due to the structure of the articles and the fact that the directive was chosen as the instrument of harmonisation, no case law can ever develop that could shed light on the interpretation of the articles" [46].

---

[41] Article 259 Sec. 2 TFEU.

[42] Article 259 Sec. 3 TFEU.

[43] Article 259 Sec. 4 TFEU.

[44] See e.g. CJEU, 16 July 2009, *Commission/Italy*, C-244/08, ECLI:EU:C:2009:478.

[45] A. VAN DOESUM, H. VAN KESTEREN, S. CORNELIE, F. NELLEN, cit., 1.9.1.

[46] G. LINDENCRONA, *Skatter och kapitalflykt. Beskattningens inverkan på det finansiella kapitalets och dess ägares rörlighet över Sveriges gränser mot bakgrund av den pågående ekonomiska integrationen i Västeuropa*, Jurist- och samhällsvetareförbundets Förlag, Stockholm, 1972, pp. 33-34

Another example can be found in the UK. Although VAT was implemented in the UK in 1973, it took up until 1984 until a national court overruled national legislation with EU VAT provisions [47]. Today, there is an active interaction between the national courts and the CJEU in the field of VAT. There are more than 1.000 judgements from the CJEU in this field [48].

As has been stated above, a national court can refer a case for a preliminary ruling to the CJEU regarding the interpretation of the VAT Directive. However, the most significant impact of the CJEU's judgments lies in their function as precedent. Through the Treaty of the European Union (TEU), the member states have committed to be loyal to the EU. The member states shall take any appropriate measure, to ensure fulfilment of the obligations arising out of the treaties or resulting from the acts of the EU institutions [49]. The Member States shall furthermore facilitate the achievement of the EU's tasks and refrain from any measure which could jeopardize the attainment of the EU's objectives [50]. This includes that the member states should not only implement EU directives into national law, but also that the national courts should interpret it in accordance with the CJEU case law. Thus, loyalty to the EU overpasses the independence and the sole judiciary of the national courts [51]. Consequently, the precedence of the case law of the CJEU is supported by the loyalty principle in article 4 TEU.

Another aspect of the interaction between the CJEU and national courts concerns the rule of law. Under article 2 TEU the EU is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for

---

(the authors' own translation from Swedish to English. Original: "*På grund av artiklarnas konstruktion och det faktum, att direktivet valts som harmoniseringsinstrument, kan aldrig någon rättspraxis utvecklas, som skulle kunna kasta ljus över artiklarnas tolkning*").

[47] D. MILNE, *Taking VAT Cases to the ECJ*, in *EC Tax Journal*, 2002, vol. *6*, p. 35 and ff.

[48] See A. VAN DOESUM, H. VAN KESTEREN, S. CORNELIE, F. NELLEN, *Fundamentals of EU VAT LAW*, cit.

[49] Article 4 Sec. 3 TEU.

[50] Article 4 Sec. 3 TEU.

[51] P. ZINONOS, *Judicial Independence & National Judges in the Recent Case Law of the Court of Justice*, in *European Public Law*, 2019, p. 634.

human rights, including the rights of persons belonging to minorities. These values are, due to the same provision, common to the member states in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail. According to the preamble of the TEU, the cultural, religious and humanist inheritance of Europe have developed the universal values of the inviolable and inalienable rights of the human person, freedom, democracy, equality and the rule of law. The rule of law is thus two-sided, it is deeply rooted in the member states' values, and at the same time a principle that shall govern EU policy [52]. A cornerstone of the rule of law is independent courts. The rule of law in the EU context restricts the independence of the national courts, in so far as they must follow the preliminary rulings and the precedence of the CJEU.

In practice, the interaction between the CJEU and national courts causes the field of VAT to evolve rapidly. The role of the highest tax courts in the member states differs significantly in the VAT domain compared to non-EU-harmonized areas of taxation. Firstly, the CJEU is highly productive, issuing several preliminary rulings each month, all of which apply *ex tunc* [53]. This means that the CJEU's interpretations are considered to have been applicable from the moment the directive entered into force [54]. Secondly, national tax authorities and lower courts are guided not only by the jurisprudence of their own highest tax court, but also by the CJEU, whose case law holds superior authority. As a result, tax authorities and lower courts may begin to follow CJEU case law rather than the precedent set by their highest national tax court. This dynamic can compel the highest national tax court to clarify or even revise its position on the same issue multiple times.

The role of case law from the highest tax courts in one member state for tax authorities and national courts in another member state also warrants attention. Judgments from courts

---

[52] See e.g. Article 21 TEU.

[53] CJEU, 22 November 2017, *Cussens et al.*, C-251/16, ECLI:EU:C:2017:881.

[54] A. van Doesum, H. van Kesteren, S. Cornelie, F. Nellen, *Fundamentals of EU VAT LAW*, cit.

in other Member States do not formally enter the hierarchy of legal sources or interpretative authorities in another jurisdiction. Nevertheless, they offer valuable insights into how the VAT Directive may be understood, particularly when the highest judicial bodies of a country have addressed the issue. Given the harmonization of VAT within the EU and the obligation of higher courts to refer questions to the CJEU for preliminary rulings, the reasoning of other Member States' highest tax courts can be highly informative. Such case law may be used by tax authorities, courts, and taxpayers to strengthen legal arguments. In practice, it may be considered at least as valuable as legal doctrine or case law from lower national courts.

5. Recent and future development in the AI era

Artificial intelligence (AI) is commonly defined as the capability of a machine to imitate intelligent human behavior. This definition is intentionally neutral with respect to the underlying technology and does not make any ontological claims about whether machines are truly intelligent or capable of thought. It focuses instead on observable behavior and functional outcomes, rather than on philosophical or cognitive dimensions of intelligence [55].

Building on this definition, Generative AI (GAI) refers to systems that produce content such as text, images, audio, or code imitates human creativity and reasoning [56]. These outputs are generated autonomously by the machine, often in response to prompts, and are designed to resemble the kind of work a human might produce. GAI has gained significant attention in recent years due to its applications in fields ranging from education and law to art and software development.

In contrast, Artificial General Intelligence (AGI) represents a more ambitious concept: a form of machine intelligence that matches or exceeds human cognitive abilities across a wide

[55] T. MELHAM *Generative AI. An introduction*, in M. ZOU, C. PONCIBO, M. EBERS, R.CALO (eds.), *The Cambridge Handbook of Generative AI and the Law* Cambridge University Press, 2019, p. 1.

[56] Ibid.

range of tasks. AGI would not only imitate intelligent behavior but also possess the capacity to understand, learn, and adapt in ways comparable to human reasoning. Unlike narrow AI systems, which are designed for specific tasks, AGI would be capable of transferring knowledge between domains, solving novel problems, and exhibiting self-directed learning [57].

As of today, there is no empirical evidence that AGI has been achieved. However, many researchers and industry leaders remain optimistic about its future development [58]. This optimism is driven by rapid advances in machine learning, neural networks, and computational power, as well as by the growing sophistication of models like Large Language Models (LLMs).

Even without AGI, AI has already revolutionized the way to access, assess and systematically review case law from the CJEU and national courts. Using LLMs for retrieving, summarizing and analysing case law is an efficient way to deal with a large material in little time [59]. Given the large amount of VAT cases from both the CJEU and the national courts, LLMs are useful within this field. The POLINE project is based on supervised learning when analysing VAT case law from the CJEU and the national courts. The outcome of the project has been tested and adjusted manually, which has created a controlled learning environment. Starting in supervised learning and as a second step use the whole potential of LLM's, is a different approach to only uploading legal documents to an LLM [60]. The POLINE model has thus the potential to better use the combination of human knowledge and intelligence with artificial intelligence, especially if this takes place in a sand-box environment, although working with such model is more time consuming.

In the era of artificial intelligence, it can be expected that the quality of judicial reasoning both in the Member States

---

[57] Ibid.

[58] See T. MELHAM, *AI. An introduction*, cit. p. 2.

[59] R. LIEPINA, F. LAGIOIA, M. LIPPI, P. PALKA, H.W. MICKLITZ, G. SARTOR, *Automating Legal Tasks. LLMs, Legal Documents, and the AI ACT*, in M. ZOU, C. PONCIBO, M. EBERS, R.CALO (eds.), *The Cambridge Handbook of Generative AI and the Law* Cambridge University Press, 2019, pp. 408-412,

[60] See R. LIEPINA et. al. , op. ult. cit., regarding the similar tool CLAUDETTE.

and at the CJEU will improve. One key reason is the sheer volume of VAT-related judgments, which is too large for any individual to efficiently locate, review, and analyse within a reasonable timeframe. AI tools, particularly LLM's, offer powerful support in navigating this complexity. However, their use also introduces certain risks. There is a danger that legal professionals may become overly reliant on machine-generated analysis, potentially weakening their own critical reasoning skills. Additionally, AI systems are known to occasionally produce inaccurate or misleading outputs, a phenomenon referred to as hallucination. These risks underscore the importance of maintaining human oversight and combining technological efficiency with sound legal judgment.

6. Concluding remarks

The system of VAT within the EU is deeply shaped by the interplay between legislation and case law, particularly the jurisprudence of the CJEU. Over time, the CJEU has evolved from a distant interpreter of directives to a central actor in shaping the VAT system through its extensive and authoritative case law. The introduction of the General Court's jurisdiction over VAT-related preliminary rulings in 2024 marks a significant institutional development, potentially enhancing specialization and efficiency in EU tax adjudication.

The relationship between national courts and the CJEU is both dynamic and complex. While national courts retain their independence, they are bound by the loyalty principle and the rule of law to interpret VAT provisions in line with CJEU jurisprudence. This has led to a shift in judicial hierarchies, where national court especially lower courts and tax authorities may prioritize CJEU case law over domestic precedent. The result is a legal landscape in which VAT law is continuously refined, not only through legislation but also through judicial dialogue across member states.

Moreover, the emergence of AI, and particularly GAI, introduces new possibilities for legal research and analysis of VAT case law. AI tools now enable more efficient access to and interpretation of vast bodies of case law, including VAT decisions from both the CJEU and national courts. Projects like POLINE demonstrate how supervised learning can be

combined with LLM's to enhance legal understanding while maintaining human oversight. Although AGI remains a theoretical concept, current AI technologies already offer transformative potential for legal practice and scholarship.

In sum, European VAT law is not static but continuously evolving driven by judicial interpretation, institutional reform, and technological innovation. The interaction between national and European legal actors, supported by emerging digital tools, ensures that VAT remains a responsive and harmonized system within the broader framework of EU law.

CHAPTER IV

THE UNNAMED CONCEPT: PRESENT IN DISCOURSE, ABSENT IN NAME. JUDICIAL INTERPRETATIVE FORMULAS IN VAT CASE LAW

Piera Santin*

CONTENTS: 1. Introduction – 2. The CJEU and the architecture of interpretation in VAT – 3. The evolution of interpretation and the silent rise of JIFs. – 4. Judicial Interpretative Formulas and the mechanics of precedent. – 5. The era of repetition: copy-pasting and the standardisation of legal meaning. – 6. Concluding remarks.

1. Introduction

Since the introduction of the Sixth Directive in 1977 and, later, the consolidation achieved through Directive 2006/112/EC, the Court of Justice of the European Union (CJEU) has developed an interpretative corpus that is not merely complementary to the written provisions but constitutive of their meaning. In this context, it appears that the Court in its reasoning increasingly uses formulaic statement, both in the form of percedent cited from previous judgements and new interpretation introduced to answer a specific question. These kind of interprertative statements are present in CJEU's drafting style, but absent in name, which is one of the reasons why POLINE Project has started, as explainned in the Introduction. This chapter proposes to call these recurring reasoning patterns *Judicial Interpretative Formulas* (JIFs).

* Piera Santin is Research Associate at the Robert Schuman Centre for Advanced Studies of the European University Institute of Florence.

They are the units of EU judicial interpretation: formulaic even if generative, recurrent but adaptive, anonymous but foundational.

The notion of a JIF captures what has been described as "structured reasoning" [1], where legal justification rests not solely upon textual reference but upon recurrent argumentative *topoi* that convey legitimacy through consistency rather than authorship. Within the CJEU's VAT case law, these formulas operated as interpretative conventions that sustain the coherence of the tax system without the formal status of rules or principles.

The CJEU's interpretative creativity in VAT has been historically linked to the characteristics of the tax itself. Because of its technical and economic complexity, VAT law depends on consistent interpretation at the Union level, and the CJEU may fully express its role of interpreter of harmonisation. From *Tolsma* to *Halifax*, from *Apple and Pear Development Council* to *WebMind Licence* [2], the Court has shaped a dense semantic network where abstract notions such as "supply", "consideration", or "public body" are defined, or new concepts as "abuse of right" or "good faith" in VAT fraud are introduced. Each judgment, while purporting to apply the Directive, in fact contributes to the construction of an autonomous vocabulary of EU tax law—a vocabulary held together less by statutory definition than by interpretative formulas iterated through the repetition amongst many different judgments.

The anonymity of JIFs is part of their power. Unlike codified principles, which attract doctrinal commentary and legislative citation, these formulas circulate silently within the body of judgments. They are invoked through phrases such as "according to settled case law", "it is necessary to consider not only the wording but also the context and objectives of the provision", or "the requirement of strict interpretation must

---

[1] N. MACCORMICK, R. SUMMERS, *Interpreting Precedents: A Comparative Study*, Routledge, 1997, p. 164.

[2] CJEU, Judgements of 3 March 1994, *Tolsma*, C-16/93, EU:C:1994:80; 21 February 2006, *Halifax,* C-255/02, EU:C:2006:12; 7 May 1988, *Apple and Pear Development Council,* C-102/86, EU:C:1988:120; 17 December 2015, *Webmind Licence*, C-419/14, EU:C:2015:832.

not deprive the exemption of its intended effect". They appear as self-evident, neutral recitations, even though each repetition concurs in the creation of the strength and reliability of the concept itself.

The term "unnamed concept" thus refers to this paradoxical condition. The formulas are omnipresent: they open, structure, and close the argumentative sections of judgments. Hence, they remain mostly unnamed (some well-known formulas exist in CJEU case law, such as Dassonville Formula [3], but, interestingly, none of them pertains to the field of VAT) because the Court never acknowledges them as such; they are linguistic habits rather than proclaimed doctrines.

The following sections will trace the emergence, consolidation, and transformation of JIFs in VAT case law. The analysis begins by examining the CJEU's dafting features that make the recurrence of formulas both necessary and possible. It then explores how fundamental VAT concepts have been gradually shaped through interpretative formulas rather than legislative amendment. The third section addresses the role of precedent and citation practices in maintaining these formulas, and the progressive standardisation of reasoning through repetition. Finally, the fourth paragraph examines the institutional implications of the 2024 reform, transferring preliminary VAT rulings to the General Court. Together, these explorations reveal how interpretative formulas, though unnamed, have become structural elements of the *aquis communautaire* and the European interpretation in the field of tax law.

[3] CJEU, 11 July 1974, *Dassonville*, C-8/74, ECLI:EU:C:1974:82. See on its importance: R. Schütze, *Re-reading Dassonville: Meaning and understanding in the history of European law* in *European law journal*, 2018, 24(6), pp. 376-407; Y. Panagis, U. Sadl, F. Tarissan, *Giving every case its (legal) due the contribution of citation networks and text similarity techniques to legal studies of european union law* in *30th International Conference on Legal Knowledge and Information Systems (JURIX'17)*. Vol. 302. IOS Press, 2017; L. Azoulai, *The 'Retained Powers' Formula in the Case Law of the European Court of Justice: EU Law as Total Law?*, in *European Journal of Legal Studies*, 2011, 4, pp. 192–219.

2. The CJEU and the architecture of interpretation in VAT

From a structural standpoint, the CJEU's judgments display remarkable consistency. Each decision follows a highly stylised architecture: the procedural background, the legal context, the questions referred, the Court's reasoning, and the operative part. This format, which has remained largely unchanged since the 1960s, functions as a rhetorical device to mask interpretative creativity beneath procedural regularity. They are not merely linguistic conventions but markers of authority. When the Court writes that "it is settled case law that the terms used to specify exemptions must be interpreted strictly, yet consistently with the objectives of the Directive", it is performing a dual act: reaffirming its role as interpreter of EU law and inscribing the interpretative formula as a recurring template for future reasoning.

The Court's interpretative style is often described as *teleological*. This characterisation, although partly accurate, risks oversimplification. The CJEU's teleology is not a mere appeal to purpose but a structured reasoning practice that integrates text, context, and system. The canonical phrasing "it is necessary to consider not only the wording but also the context and the objectives of the provision", first articulated in *van Gend en Loos*[4] and subsequently reiterated across decades, is one of the primordial JIF of EU jurisprudence. Within VAT case law, this formula performs the critical task of mediating between linguistic ambiguity and functional coherence. It allows the Court to move from the textual indeterminacy of the Directive's provisions to a systemically consistent interpretation aligned with the broader objectives of the internal market.

This interpretative posture also reflects the Court's procedural environment. Preliminary rulings under Article 267 TFEU constitute the principal channel through which VAT interpretation evolves. The questions referred by national courts are rarely abstract; they emerge from concrete disputes. The CJEU must thus produce answers that are both legally precise and functionally transposable. JIFs serve this dual

[4] CJEU, 5 February 1963, *Van Gend en Loos,* C-26/62, EU:C:1963:1.

requirement by offering reusable argumentative patterns that can be applied across analogous contexts [5].

A distinctive feature of CJEU judgments is the absence of dissenting opinions. The Court delivers a single, unanimous judgment, unsigned by individual judges [6]. This practice reinforces the impersonal authority of its reasoning: interpretation is presented as collective truth, not as deliberative outcome. Within such a structure, JIFs become the main vehicle for continuity. Their repetition across judgments guarantees that the Court's voice remains consistent, even as compositions of chambers and national backgrounds of judges vary. The formula, rather than the individual jurist, becomes the subject of enunciation.

Scholars have noted that this stylistic uniformity, while enhancing the authority of EU law, also blurs the boundaries between adjudication and legislation [7]. In the VAT domain, this is particularly evident in cases where the Court has introduced brand new legal concepts. The notion of "abuse of right", introduced in *Halifax* (2006), was not derived from any textual provision of the Directive but from the Court's own understanding of the system's objectives. Yet the reasoning was framed through the standard JIF of functional coherence: "the prohibition of abusive practices is inherent in the system of VAT". The use of formulaic drafting has a twofold effect. On one hand, it tries to reduce the impact of the innovation, transforming the judicial law-making into an act of interpretative necessity. On the other side, the judgment prepares the spot for future repetitions, as the concept of *abuse of* right in the field of VAT is introduced in a formulation that seems to be ready to be cited as a precedent in the future. In this sense, JIFs function as what legal semioticians call "neutralising mechanisms": they present creative reasoning as mere reiteration of established logic.

The interplay between neutrality and creativity is also visible in the Court's treatment of fundamental VAT notions.

---

[5] C. PERELMAN, L. OLBRECHTS-TYTEC, *The New Rhetoric: A Treatise on Argumentation,* 1969, Notre Dame, p. 423.

[6] T. TRIDIMAS, *The General Principles of EU Law*, Oxford, 2006, p. 6.

[7] P. CRAIG, G. DE BÚRCA, *EU Law: Text, Cases and Materials*, 7th ed., 2020, Oxford, p. 111.

When defining "supply of services" or "consideration", the Court systematically employs the same sequence of interpretative steps: textual reference, contextual integration, purposive justification. Each step is accompanied by a formula that signals deference to the Directive's system while silently extending its meaning. For example, in *Tolsma*, the Court held that an activity constitutes a supply of services only where there is a judicial link between the service rendered and the consideration. The concept of "judicial link" thereafter became a fundamental part of that of "taxable transaction". Similarly, such a paragraph from *Tolsma* became a recurrent JIF, deployed across numerous subsequent judgments as a criterion of taxable transactions. Through such reiterations, the CJEU effectively codifies interpretative standards without codifying their definitions.

The rhetorical economy of CJEU reasoning thus depends on rhythm rather than novelty. Each judgment oscillates between the reiteration of familiar formulas and their contextual recalibration. The authority of the Court lies not in explicit doctrinal statements but in the predictability of its syntax. This predictability, however, should not be mistaken for rigidity. As MacCormick observed, "coherence in judicial reasoning is a dynamic equilibrium, not a static system"[8]. The JIFs embody this equilibrium: they preserve continuity while enabling incremental adaptation. Over time, they accumulate semantic sediment, gradually altering the meaning of the concepts they articulate.

In this architecture, the absence of formal precedent doctrine is not a weakness but a precondition for flexibility. The CJEU has repeatedly affirmed that its case law does not constitute binding precedent in the Anglo-American sense. Hence, it systematically cites its previous judgments as authoritative sources of interpretation[9]. Citations operate not as a rule of *stare decisis* but as a reaffirmation of formulaic reasoning. When the Court refers to earlier cases, it does not

---

[8] N. MACCORMICK, *Rhetoric and the Rule of Law: A Theory of Legal Reasoning*, 2005, Oxford, p. 84.

[9] A. G. TOTH, *The Authority of Judgments of the European Court of Justice: Binding Force and Legal Effects,* in *Yearbook of European Law*, 1984, pp. 1–34.

reproduce their factual matrices but extracts and re-applies their JIFs. In doing so, it creates a cumulative network of reasoning patterns, a *jurisprudence of formulas* rather than of holdings.

Finally, the structure of preliminary rulings itself amplifies the diffusion of these formulas. Because the CJEU's answers must be applicable across Member States, they are phrased at a level of abstraction sufficient to transcend national legal peculiarities. The resulting reasoning tends to favour conceptual generality over contextual specificity. JIFs thus serve as linguistic anchors that maintain coherence across translations, both linguistic and legal. They embody the Union's aspiration to uniformity through interpretative repetition.

3. The evolution of interpretation and the silent rise of JIFs

The evolution of VAT law within the European Union is inseparable from the interpretative work of the Court of Justice. While the legislative framework has remained relatively stable since the recasting of Directive 2006/112/EC, the Court's jurisprudence has continually transformed the conceptual content of that framework

The early VAT case law reveals how the Court's interpretative enterprise emerged as a response to the Directive's inherent indeterminacy. The Sixth Directive sought to harmonise the basic structure of VAT across Member States, but many of its key terms such as "supply of goods", "supply of services", and "consideration" were left undefined. In *Tolsma*, the Court established the now-classic formula that a service is supplied "for consideration" only where there is a direct link between the service provided and the value received. This formulation, seemingly factual, was in fact conceptual: it defined the notion of "economic activity" in the VAT sense, separating it from mere social or voluntary exchanges. The "direct link" became a JIF of enduring influence, subsequently invoked in cases concerning everything from complex financial transactions to public broadcasting services. Each reiteration reinforced the formula's authority, allowing it to function as a de facto definition without the need for codification.

A similar process occurred in relation to the notion of

"neutrality". Although the principle of fiscal neutrality appears nowhere in the Treaty or in the Directive's provisions, it has become one of the cornerstones of VAT jurisprudence [10]. The Court's invocation of neutrality follows a consistent formula: exemptions must be interpreted strictly because they derogate from the general rule of taxation, but such interpretation must not deprive the exemption of its intended effect, in order to preserve neutrality among taxable persons. This balancing formula, repeated across decades of judgments, has generated a complex equilibrium between restrictive and purposive interpretation. Neutrality thus operates not as a moral or economic ideal but as a judicial interpretative formula that mediates between textual fidelity and systemic coherence.

The construction of VAT concepts through JIFs is particularly evident in the field of exemptions. The CJEU has repeatedly stressed that exemptions represent derogations from the general principle of taxation and therefore must be construed narrowly. Yet, through its reasoning, the Court has progressively expanded the interpretative reach of these exemptions. For example, in *Commission v. France* (C-197/12), the Court held that Member States could not determine the existence of certain maritime exemptions solely by reference to objective criteria such as the tonnage of vessels; rather, they had to assess whether the vessel's use corresponded to the economic function envisaged by the Directive. The formula "it is for the national authorities to verify, taking account of all the relevant circumstances" has since become a recurring interpretative pattern, transferring the burden of factual assessment to national courts while preserving the Court's control over conceptual interpretation.

One of the most profound examples of silent conceptual construction is the above-mentioned concept of "abuse of

[10] A. MONDINI, *Il principio di neutralità dell'IVA tra mito e (perfettibile) realtà*, in A. DI PIETRO, T. TASSANI (eds.), *I principi europei del diritto tributario*, Padova, 2013, pp. 269-306; C. AMAND, *VAT neutrality: a principle of EU law or a principle of the VAT system?*, in *World Journal of VAT/GST Law*, 2013, pp.163-81; R. DE LA FERIA, *EU VAT Principles as Interpretative Aids to EU VAT Rules: The Inherent Paradox*, in M. LANG et al. (eds.), *Recent VAT Case Law of the CJEU*, Linde, 2016; P. DANIEL, *Relations between the principle of neutrality and elements of value added tax structure*, in *Financial Internet Quarterly*, 2021, pp. 56-63.

right" in VAT as articulated in *Halifax*. Here, the Court declared that "the prohibition of abusive practices is inherent in the system of VAT", thereby introducing a general anti-avoidance principle into EU tax law without legislative basis. The paragraph containing the first formulation of the abuse of right has since become a prototypical JIF. Its power lies in its ability to present innovation as continuity: by claiming that the prohibition of abuse was "inherent" rather than newly introduced, the Court transformed a creative act into a declaratory one. The doctrine's subsequent development in cases such as *Part Service* and *RBS Deutschland* confirmed that the JIF had taken root, enabling the Court to construct an entire field of anti-abuse jurisprudence without amending a single provision of the Directive.

The iterative nature of these formulas produces a cumulative effect. Over time, each repetition adds a layer of semantic nuance, slowly reshaping the conceptual landscape. The cumulative effect is visible in the evolution of the concept of "supply". Initially defined in transactional terms, "supply" has come to encompass composite and complex operations, including digital services and cross-border transactions, through the Court's systematic deployment of functional formulas such as "economic reality prevails over form" and "substance over appearance". These expressions, though borrowed from general principles of law, have acquired specific meaning in VAT law through repetition and adaptation. Their origin in the Court's case law rather than in the Directive's text underscores the creative potential of interpretative formulae as vehicles of legal evolution.

The Court's interpretative style also reveals a temporal dimension. Because preliminary rulings often concern transactions governed by earlier directives, the CJEU must interpret old provisions in the light of subsequent developments. This retrospective hermeneutics reinforces the continuity of JIFs: by asserting that the interpretation of a provision in the Sixth Directive remains applicable to its corresponding article in the Recast Directive, the Court ensures semantic stability across temporal layers. Thus, even when the legal text changes, the interpretative formula persists, anchoring the evolution of meaning in the Court's own jurisprudence.

The interplay between constancy and change becomes

particularly visible when one compares the Court's early and recent case law on the same topics. The concept of "consideration", for instance, has shifted from a narrow, contractual focus to a broader economic one. In *Apple and Pear Development Council*, the Court rejected the idea that statutory levies imposed on producers could constitute consideration, since there was no direct link between payment and service. Decades later, in cases involving complex corporate structures or digital platforms, the Court has invoked the same "direct link" formula to encompass multi-layered exchanges of value, extending the concept far beyond its original scope. The formula's constancy conceals the transformation: the words remain the same, but their referential field expands.

At the methodological level, the Court's evolution demonstrates an increasing reliance on what could be described as "formulaic precedent". While the CJEU maintains that its judgments are not formally binding, the recurrence of specific linguistic constructions functions as a mechanism of building the bindingness of precedent. Each formula acts as a citation anchor, linking new judgments to prior ones. The Court's policy of paragraph numbering and precise cross-references has further reinforced this mechanism, enhancing the recognisability and portability of JIFs [11]. As a result, interpretative continuity is achieved not through hierarchical bindingness but through textual repetition. The formula becomes the unit of transmission, ensuring that the law evolves through linguistic iteration rather than doctrinal reform.

4. Judicial Interpretative Formulas and the mechanics of precedent

The growing importance of JIFs in VAT case law depends not only on their internal coherence but also on the way they are transmitted, cited, and redeployed within the judicial system. The CJEU's jurisprudence operates without a formal doctrine of *stare decisis*; yet, paradoxically, its authority relies

[11] See to this end K. LENAERTS, *The Court of Justice of the European Union and the Rule of Law*, in *European Law Journal*, 2021, p. 3–20.

heavily on the self-referential citation of previous judgments. This section examines how the mechanics of precedent in EU law have evolved into a system of formulaic continuity—a mode of reasoning where interpretative stability is maintained through the reiteration of linguistic constructions rather than the binding force of prior decisions.

In its early case law, the CJEU seldom cited its own previous cases. The institutional legitimacy of the Court was still fragile, and the emphasis lay on presenting each decision as a self-contained syllogism. Over time, however, the expansion of EU law and the growing complexity of its subject matter made explicit references to earlier judgments necessary. The Court's gradual transition from a deductive to a cumulative reasoning style can be traced from the 1980s onwards, as paragraph numbering and standardised citation practices were introduced [12]. In VAT cases, this transformation coincided with the Court's increasing reliance on recurring interpretative formulas. Each citation served a double function: it linked the new decision to an established line of reasoning and confirmed the formula's continued validity. The Court thus constructed a web of interpretative consistency that mimicked the logic of precedent without acknowledging it as such.

The distinction between de jure and de facto precedent in EU law has been widely discussed in the literature. Some scholars maintain that CJEU judgments lack binding force even upon national courts, which remain formally free to interpret EU law independently [13]. Others argue that the persuasive authority of the Court's reasoning is so overwhelming that, in practice, it functions as binding law. The truth lies somewhere in between: the CJEU's authority derives not from the coercive power of precedent but from the internal compulsion of formulaic reasoning. When CJEU cites "settled case law", it is not invoking a specific ratio decidendi

---

[12] M. BOBEK, *Of Feasibility and Silent Elephants: The Legitimacy of the Court of Justice through the Eyes of National Courts*, in Cambridge Yearbook of European Legal Studies, 2013, pp. 263–296

[13] A.M. MANCALEONI, *National Judges and the case law of the Court of Justice of the European Union: Introduction*, in A. M. MANCALEONI, E. PUILLOT (eds.), *National Judges and the case law of the Court of Justice of the European Union*, Roma, 2021, pp. 1-7.

but reproducing a pattern of interpretation that carries institutional legitimacy precisely because it has been repeated. The JIFs, not the factual context of the original case, become the operative unit of precedent.

The Court's manner of citation illustrates this mechanism. When referring to prior cases, it rarely recounts their facts or procedural background. Instead, it extracts specific paragraphs and incorporates them verbatim into the reasoning of the new judgment [14]. This technique creates a chain of textual repetition in which the same linguistic construction appears across multiple decisions, slightly adapted but fundamentally unchanged. The result is a corpus of "floating paragraphs" that circulate throughout the jurisprudence, functioning as portable fragments of interpretative authority. The Court's increasing use of expressions such as "it is settled case law that" or "according to consistent case law" marks the point where citation turns into JIFs. The phrase itself becomes a rhetorical signal of continuity, performing the very stability it proclaims.

At a deeper level, this reliance on formulaic precedent reflects the structural tension between the CJEU's need for coherence and its avoidance of rigid hierarchy. EU law aspires to uniformity, but its legitimacy depends on flexibility. The Court's method allows it to reconcile these objectives: by repeating JIFs, it ensures predictable outcomes even if consistent with contextual adaptation. When necessary, the Court can subtly adjust the meaning of a JIF without openly overruling its predecessors. This technique, sometimes called "silent revision" [15], enables jurisprudential evolution without institutional rupture. In VAT law, silent revisions happened in areas such as bad payment, abuse of right, or proof of good faith.

The persuasive force of JIFs also operates horizontally, influencing national courts. Preliminary references are a dialogic process in which national judges seek clarification while contributing to the development of Union law. The language of "settled case law" reassures them that their

---

[14] K. LENAERTS, *Upholding the Rule of Law through Judicial Dialogue*, in *Yearbook of European Law*, 2019, pp. 3–17.

[15] See J. VERHOEVEN, *Belgium,* in P.M. EISEMANN (ed.), *The Integration of International and European Community Law into the National Legal Order: A Study of the Practice in Europe,* Leiden, 1996, p. 135.

decisions align with a transnational interpretative community. When national courts cite the CJEU, they often reproduce the same paragraphs, further disseminating the formulas within domestic legal systems. Over time, this creates a feedback loop: national courts internalise the formulas, employ them in their reasoning, and eventually generate new references that invite the Court to confirm or refine those same formulas. The system thus evolves through repetition, not command.

The mechanics of precedent in VAT law have therefore produced a new form of normativity that could be described as textual precedent. Instead of binding decisions, the EU legal order operates through binding language. As argued, every act of writing carries within it the potential for repetition and reinterpretation [16]. The CJEU's JIFs embody this logic: their authority derives from the expectation of recurrence. When the Court cites an earlier JIF, it does not merely recall a past judgment; it re-enacts its performative power. The JIF becomes a speech act that continuously reproduces the law.

This phenomenon raises important theoretical questions about the nature of precedent in supranational legal systems, particularly what ensures the integrity of interpretation if authority arises from repetition rather than hierarchy.

The cumulative effect of these mechanisms is a jurisprudence that is at once stable and dynamic. Stability arises from the Court's disciplined adherence to established JIFs; dynamism from its ability to recalibrate them incrementally. This duality explains why VAT jurisprudence has achieved such a high degree of internal consistency despite the absence of formal precedent. The law evolves through the self-replication of language. Each JIF functions as both memory and method: a repository of past reasoning and a template for future judgment.

The next section explores the implications of this linguistic replication for the evolution of judicial style. If the 1980s and 1990s were decades of conceptual innovation, the early twenty-first century marks the rise of a new phenomenon: the era of repetition. Through the mechanical re-use of

[16] J. DERRIDA, *Writing and Difference*, Chicago, 1978.

interpretative paragraphs, the Court has entered a phase of standardisation that both consolidates and exhausts the creative potential of its earlier formulas. Understanding this transformation is crucial for assessing the current and future trajectory of VAT jurisprudence in the European Union.

5. The era of repetition: Copy-pasting and the standardisation of legal meaning

By the second decade of the twenty-first century, the Court of Justice's interpretative style in VAT cases had entered what may be called the *era of repetition*. The proliferation of preliminary references, the increasing technicality of disputes, and the consolidation of a mature corpus of case law have transformed the Court's reasoning from creative construction to systematic replication. While the earlier period of VAT jurisprudence was characterised by conceptual innovation and the introduction of a significant amount of new JIFs, the recent phase is marked by their mechanical reiteration. The same paragraphs, phrases, and citations recur across judgments with minimal variation, producing what might be described as the "standardisation of legal meaning". This phenomenon, often lamented by scholars as "copy-pasting technique [17]", is not merely a stylistic degeneration but a structural response to the institutional pressures of modern adjudication.

The expansion of the Court's docket is the most immediate factor behind this development. As VAT disputes multiplied across the Union, the Court faced the challenge of ensuring uniformity while maintaining efficiency. Between 2015 and 2021 alone, VAT accounted for nearly half of all preliminary rulings concerning taxation [18]. To manage this workload, the CJEU began to rely increasingly on formulaic reasoning. The

---

[17] F. MILLET, *In the name of analogy: Judicial copy-pasting and competence creep in the connection data case law*, in *Common Market Law Review*, 2024, p. 1289. For an earlier use of that expression the repetition of the text of unpublished opinions in other unpublished opinions in the US, see B. SOUCEK, *Copy-Paste Precedent*, in *The Journal of Appellate Practice and Process*, 2012, pp.153-171.

[18] Court of Justice of the European Union, *Annual Report 2021: Judicial Activity*, Luxembourg, 2022.

use of pre-fabricated paragraphs from earlier decisions allowed chambers to produce consistent judgments in less time. Each formula served as a modular unit of reasoning, ready to be inserted, adapted, and redeployed. Over time, this practice solidified into a method: the jurisprudence became self-replicating, with past formulations providing the textual material for future judgments [19].

The "copy-paste" technique is particularly evident in the field of VAT. Here, the Court often repeats verbatim the canonical formulations: exemptions must be interpreted strictly as derogations, but such interpretation must comply with the principle of neutrality; the right of deduction constitutes an integral part of the VAT mechanism and cannot, in principle, be limited; supplies for consideration require a judicial link between the service rendered and the value received.[2] These sentences appear, sometimes identically, in dozens of judgments. They no longer serve to justify a specific conclusion but to signal adherence to established interpretative practice. Their repetition ensures stability but also risks semantic dilution: the formula becomes so familiar that its meaning is assumed rather than examined.

This transformation has provoked contrasting reactions among scholars. Some see it as evidence of judicial conservatism, a retreat from the creative vitality that once defined EU law [20]. Others interpret it as a sign of maturity: a phase in which the law has stabilised and the need for innovation has diminished. Both perspectives contain truth [21]. The standardisation of reasoning has undoubtedly curtailed the Court's conceptual experimentation, yet it also reflects a rational adaptation to the demands of uniform application. In a legal order premised on integration, the predictability of formulas may serve the rule of law more effectively than doctrinal novelty. Apparently, consistency in the interpretation

---

[19] The idea of the repetitive citation as a technique can be found in M. JACOB, *Precedents and Case-Based Reasoning in the European Court of Justice*, 2014, pp. 95 and 100-101.

[20] S. E. M. HERLIN KARNELL, T.KONSTADINIDES, *The Rise and Expressions of Consistency in EU Law: Legal and Strategic Implications for European Integration*, in *Cambridge Yearbook of European Legal Studies*, 2013, p. 139.

[21] P.M. TIERSMA, *The Textualisation of Precedents*, in *Notre Dame Law Review*, 2007, pp. 1187-1278.

of EU law is itself a form of justice, ensuring that "like cases are treated alike" across the Union [22].

Nevertheless, the mechanical replication of JIFs carries significant epistemic consequences. First, it obscures the historical contingency of the JIFs themselves. When paragraphs are reproduced without contextual explanation, they appear timeless and self-evident, detached from the factual and doctrinal circumstances that gave them birth. For instance, the formula asserting that "the prohibition of abusive practices is inherent in the system of VAT", originally articulated in *Halifax*, is now invoked routinely, even in cases that bear little resemblance to the artificial tax avoidance schemes that prompted the doctrine's creation. The effect is a form of *semantic ossification*: phrases that once conveyed dynamic reasoning harden into ritual incantations. The Court's language becomes liturgical–repeated for its own authority rather than for its argumentative force [23].

Second, repetition alters the Court's relationship to its own authority. The new phase of formulaic reasoning represents, paradoxically, a withdrawal from visibility. The Court hides behind its own precedent, allowing its voice to be carried by the anonymous authority of "settled case law". It is like the JIFs become the true authors of the judgments.

Moreover, the spread of copy-paste reasoning influences how national courts engage with EU law. When preliminary rulings reproduce familiar formulas, national judges learn to interpret those formulas as direct commands rather than as contextual arguments. The interpretative dialogue that once characterised the preliminary reference procedure risks turning into a monologue [24]. This phenomenon undermines the dialogical legitimacy of the EU judicial system, which depends on mutual

---

[22] K. LENAERTS, *New Horizons for the Rule of Law within the EU*, in *German Law Journal*, Special Issue 1: 20 Challenges in the EU in 2020, January 2020, pp. 29 – 34.

[23] As stated by P.M. TIERSMA, cit., "rather than treating statutes as common law, courts are beginning to treat the common law as legislation. Minds will differ on whether this transformation is good or bad. There are many consequences that flow from writing down the law in an authoritative way" (p. 1188).

[24] M. BROBERG, N. FENGER, *Preliminary References to the European Court of Justice*, Oxford, 2021, p. 42.

responsiveness between the CJEU and national courts. The formula, designed to facilitate communication, may in fact reduce it to repetition. The jurisprudence thus becomes more coherent but less conversational.

Yet it would be misleading to describe this evolution solely in negative terms. The standardisation of reasoning has produced undeniable benefits for legal certainty and administrative efficiency. The use of pre-formulated paragraphs ensures that similar issues are resolved in similar ways, reducing interpretative divergence among Member States. In an area as sensitive as VAT, where the same transaction may involve operators from multiple jurisdictions, such uniformity is crucial. Moreover, the formulaic method provides a transparent repository of interpretative guidance accessible to national courts, tax authorities, and practitioners.

There is, however, a subtle irony in this development. The CJEU's success in producing a coherent interpretative language may ultimately render its intervention less necessary. As the formulas stabilise and their application becomes routine, the interpretative burden shifts downward to national authorities [25]. The 2024 reform transferring preliminary VAT rulings to the General Court can thus be read not merely as an administrative adjustment but as the logical culmination of the standardisation process. The interpretative system has reached a point where the formulas are sufficiently entrenched to be applied by a lower court. The once-creative language of integration has become the grammar of administration.

The notion of "copy-paste" must therefore be understood dialectically: it represents both the triumph and the exhaustion of judicial interpretation. On one hand, it testifies to the

[25] As stated in one of the most recurrent JIF, not specifically related with VAT case-law, according to which "Similarly, it is solely for the national court, before which the dispute has been brought and which must assume responsibility for the judicial decision to be made, to determine, in the light of the particular circumstances of the case, both the need for and the relevance of the questions that it submits to the Court [...]. In particular, it is not for the Court to verify the accuracy of the legislative and factual context, which the national court is responsible for defining […]. Consequently, where the questions referred concern the interpretation of Union law, the Court is, in principle, bound to give a ruling" (amongst others: Judgement of 5 December 2013, TVI, Joined Cases C618/11, C637/11 and C659/11, ECLI:EU:C:2013:789.

CJEU's capacity to construct a self-sustaining interpretative framework, capable of reproducing itself across time and institutional levels. On the other hand, it signals the waning of the Court's formative power, the passage from invention to maintenance. The formulas that once created law now preserve it. The jurisprudence has entered what could be called its *classical period*: a stage where the canon is fixed, its language standardised, its creativity subdued by its own success.

The implications of this transition extend beyond VAT law. They illuminate a general tendency within mature legal systems: as doctrines stabilise, interpretation becomes increasingly ritualised. The repetition of authoritative phrases serves to affirm the system's continuity, even as the world it regulates continues to change. The danger lies not in repetition itself but in the loss of critical awareness of its origins. Every repetition contains both memory and difference; to repeat without remembering is to risk erasing meaning. The challenge for the future of EU VAT jurisprudence is therefore to maintain the balance between stability and reflection—to ensure that formulas remain instruments of interpretation rather than substitutes for thought.

In this sense, the "era of repetition" should not be seen as an endpoint but as a moment of self-awareness. The Court's reliance on formulaic reasoning invites scholars and practitioners to reconsider what it means to interpret law in a multilingual, multi-level legal order. If meaning is produced through repetition, then interpretation becomes an act of re-inscription rather than discovery. The next and final section explores how this process might evolve in the post-2024 landscape, as the competence for VAT preliminary rulings passes to the General Court.

6. Concluding remarks

The trajectory of VAT case law before the CJEU can be read as a long movement from invention to sedimentation, from interpretative experimentation to linguistic structure. JIFs that once served as instruments of creative reasoning have become the architecture of a mature legal system. They embody the paradox announced at the outset of this chapter: an unnamed concept that dominates discourse while remaining formally

unacknowledged. Present in every judgment, absent from every definition, the JIF has evolved from a pragmatic tool into the grammar of European fiscal legality.

This transformation speaks to the deeper nature of EU law as a legal order relying on interpretation rather than codification. Since *van Gend en Loos* and *Costa v. ENEL*, the Court's authority has rested on its ability to articulate meaning where the Treaties and directives remain silent. In VAT law, this function has been amplified by the Directive's technical complexity and its reliance on uniform application. The CJEU's JIFs have thus performed an integrative function: they have harmonised not only the application of tax rules but also the modes of legal reasoning across Member States. Each JIF acts as a microcosm of European legal culture, merging civil-law textualism with common-law pragmatism. Their proliferation has created a shared idiom of legal thought, an "Eurolect" [26] of adjudication that transcends linguistic and doctrinal boundaries.

Hence, the success of this idiom has produced its own limitations. The consolidation of formulaic reasoning has gradually reduced the space for interpretative plurality. In the early decades, the Court's judgments were laboratories of conceptual innovation. Cases such as *Tolsma*, *Apple and Pear Development Council*, and *Halifax* demonstrated the CJEU's willingness to craft new categories, including "direct judicial link", "economic activity", and "abuse of right", to ensure the functional coherence of VAT. In recent years, however, the recurrence of these same formulas has transformed innovation into maintenance. The Court now speaks through inherited language, and its interpretative creativity consists largely in managing the equilibrium between consistency and adaptation. This is not stagnation but institutional maturity: the jurisprudence has become self-referential, capable of reproducing its own logic without external input.

The 2024 transfer of competence for VAT preliminary rulings to the General Court marks a symbolic turning point in this evolution. It formalises a process that has long been

[26] B. LUCJA, *Eurolects and EU Legal Translation*, in M. JI, S. LAVIOSA (eds.), *The Oxford Handbook of Translation and Social Practices*, 2021, pp. 478-500.

underway: the administrative domestication of interpretation. By entrusting the General Court with cases once reserved for the CJEU, the Union acknowledges that VAT law has reached a level of interpretative stabilisation that no longer requires the highest judicial forum's constant supervision. The interpretation made through JIFs has become so entrenched that it can now be done by a lower court without risk of inconsistency. In this sense, the reform represents both the decentralisation and the institutionalisation of the JIFs system.

The consequences of this shift are both practical and theoretical. Practically, it is likely to enhance efficiency: the General Court, with its expanded membership, can process the high volume of VAT cases more rapidly. Theoretically, however, it raises questions about the future of interpretative authority. The CJEU's authority derived not only from its reasoning but from its symbolic position as the grantor of integration. The General Court, by contrast, is conceived as a court of first instance—closer to administration than to constitutional adjudication. The risk is that the use of JIFs, once vehicles of principled reasoning, may degenerate into bureaucratic routines. Without the aura of constitutional discourse, repetition might harden into rigidity.

The General Court, faced with the need to apply inherited JIFs to new economic realities—digital transactions, platform economies, and green taxation—may rediscover the creative potential of language. Just as the CJEU once transformed textual ambiguity into conceptual innovation, the General Court may transform procedural standardisation into renewed interpretative experimentation. The JIFs, far from being static relics, possess an inherent elasticity that allows them to accommodate new contexts. Their survival depends on their adaptability.

In VAT law, these formulas have come to function as *meta-norms*: they govern the way norms are understood, rather than prescribing substantive outcomes. Their authority lies in their repetition, but their legitimacy lies in their intelligibility. As long as the formulas continue to generate coherent reasoning and facilitate mutual understanding between courts, they remain legitimate. When they become opaque or when repetition replaces reasoning, they lose their normative force.

The evolution of JIFs also illustrates the intimate link between language and power in European law. The CJEU's

discourse has always been performative: to declare that a principle is "inherent in the system" or that a concept must be "interpreted in light of its context and objectives" is to create, through language, the very coherence it describes. The authority of EU law resides not in a hierarchical command but in a discursive consensus sustained by repetition.

The challenge for the post-2024 era is to preserve the communicative function of this syntax while preventing its petrification. The General Court, inheriting the CJEU's vocabulary, must ensure that formulas remain responsive to new realities. The neutrality principle, for example, may acquire new dimensions as the Union moves towards environmental taxation and digital levies. Similarly, the notion of "abuse of right" may evolve to encompass algorithmic fraud or platform-based avoidance schemes. Whether these developments will generate new interpretative formulas or revitalise existing ones remains to be seen. What is certain is that the European judicial discourse on VAT will continue to operate through formulaic reasoning – the anonymous, iterative language that both expresses and sustains the Union's legal identity.

In conclusion, the JIFs that populate the CJEU's judgments are not merely stylistic conventions; they are the architecture through which meaning circulates in a multilingual, multi-jurisdictional order. Their anonymity protects them from controversy, but their persistence ensures the stability of the legal system. As interpretation becomes structure, the line between reasoning and normativity blurs.

CHAPTER V

"PRINCIPI DI DIRITTO" AND JUDICIAL INTERPRETATIVE FORMULAS: BETWEEN NATIONAL DOCTRINE AND EUROPEAN INTEGRATION

Alessia Fidelangeli – Andrea Mondini*

1. Introduction

The Italian legal framework has developed a long-standing tradition of extracting, collecting, and publishing "principi di diritto" (hereinafter "interpretative principles of

* Alessia Fidelangeli is Research Fellow at the University of Bologna; Andrea Mondini is Associate Professor in Tax Law and Public Finance Law at the University of Bologna.

law") from the jurisprudence of the Italian Supreme Court, the "Corte di Cassazione" (hereinafter "Court of Cassation"). This extraction process is carried out manually by the *Ufficio del Massimario* of the Court (the Research Office of the Court of Cassation, hereinafter "ROCC"), in line with the Court's institutional *nomophylactic* function, which will be further examined in the following sections (par. 3.1). Moreover, while this work does not address the matter in depth, it is worth noting that the Italian tax courts have recently undertaken initiatives to compile *massimari* of national some tax decisions on the merits.

Although *interpretative principles of law* share certain similarities with Judicial Interpretative Formulas (JIFs), they also differ in significant respects.

This contribution pursues three main objectives: first, to outline the distinctive features of the Italian tax judiciary in comparison to other jurisdictions covered in the study, Sweden, and Bulgaria (parr. 2.1, 2.2, and 2.3); second, to analyse the conceptual differences between the JIF and the *interpretative principle of law* concepts (par. 3.2); and third, to investigate whether the extraction of both existing *interpretative principles of law* and potential future JIFs could contribute to the consolidation of case law. This inquiry, in turn, entails further reflection on the role of precedent and the broader debate on the creative function of case law in the Italian legal system (parr. 4.1 and 5.1), particularly in the domain of value-added tax (parr. 4.2 and 5.2). Finally, the paper presents empirical findings on the characteristics of JIFs in the Italian Supreme Court's case law (parr. 3.3 and 6).

2. The Italian tax judiciary and the allocation of supreme jurisdiction to the Court of Cassation

2.1. The structure of the tax judiciary

The current structure of the tax judiciary is the result of the 1992 reform that established an autonomous tax judiciary, which was implemented as of April 1st, 1996. The tax judiciary has competence on tax issues. Tax issues are considered as such, first, when they concern a tax and, second, when the applicant

challenges one of the acts listed in the relevant provisions regulating the tax process [1].

The most important tax justice regulation is contained in Legislative Decree No. 545 of December 31, 1992, which provides the regulation concerning tax courts (composition, structure, etc.) and Legislative Decree No. 546 of December 31, 1992, which provides the regulation of the tax process (content of the appeal, rules concerning competence, etc.) [2]. For matters not contained in Decree 546/1992, the rules of the Code of Civil Procedure apply.

Law No. 130 of August 31, 2022, introduced a reform of the tax judiciary which impacts both the composition of tax courts and tax process [3].

The tax justice system currently consists of courts of first instance ("Corti di Giustizia Tributaria di Primo Grado") and courts of second instance ("Corti di Giustizia Tributaria di Secondo Grado"). The courts of first instance are located in each provincial capital while the courts of second instance are located in each regional capital. A judge is at the head of each tax court. Each court consists of several chambers. Each chamber is composed of a president, who presides over the

---

[1] F. Batistoni Ferrara, *Gli atti impugnabili nel processo tributario avanti le Commissioni,* in *Diritto e Pratica Tributaria,* 1996, p. 1109 ff.; C. Glendi, *Contenzioso tributario*, in *Enciclopedia Giuridica*, IX, Roma, Treccani, 2001, p. 1 ff.; Id., *Processo XII) Processo tributario*, in *Enciclopedia Giuridica*, XXVIII, Roma, Treccani, 2004, p. 1 ff.; F. Maffezzoni, *Atti impugnabili e funzione del processo avanti alle commissioni tributarie. Parte prima*, in *Bollettino Tributario,* 1976, p. 1389; L. Perrone, *I limiti della giurisdizione tributaria*, in *Rassegna Tributaria*, 2006, p. 707 ff.; A. Poddighe, *Gli atti impugnabili dinanzi alle Commissioni tributarie: rassegna di giurisprudenza di legittimità e dottrina*, in *Rivista di Diritto Tributario*, 2012, p. 655 ff.; P. Russo, *Processo tributario (voce)*, in *Enciclopedia del Diritto*, XXXVI, Milano, 1987, p. 765 ff.; F. Tesauro, *Gli atti impugnabili e i limiti della giurisdizione tributaria*, in *Giustizia Tributaria*, 2007, p. 12 ff.

[2] F. Tesauro, *Processo tributario*, in *Digesto delle Discipline Privatistiche, Sezione Commerciale,* Torino, 2017, p. 339.

[3] A. Carinci, F. Pistolesi (eds.), *La riforma della giustizia e del processo tributario, Commento alla Legge 31 agosto 2022, n. 130*, Milano, 2022; S. Donatelli, C. Buccico, A. Cuva, F. Gallo, A.F. Uricchio, *Le novità introdotte dalla legge n. 130/2022 di riforma del processo tributario*, Bari, 2023, p. 103 ff.; C. Glendi, *La novissima stagione della giustizia tributaria riformata*, in *Dir. prat. trib.*, 2022, p. 1146 ff.; C. Consolo, G. Melis, A.M. Perrino, *Il giudizio tributario*, Milano, 2022.

first chamber, a vice-president, and not less than four tax judges. When a case is assigned to a chamber, decisions are adopted by a panel of three judges belonging to that chamber or by a monocratic judge for judgments up to 5,000 euros [4].

The composition of tax courts is evolving. Traditionally, tax judges were not full judges but honorary judges. Law 130/2022 provided that, from 2024, tax judges must be selected through a public exam, similar to that used to select full judges [5]. The self-governing body of the tax judiciary is called the "Consiglio di Presidenza della Giustizia Tributaria" [6].

The administrative support function for judicial activity is carried out by the staff of the Secretarial Office of the Tax Justice Courts, which belongs to the Department of Tax Justice of the Ministry of Economy and Finance (MEF). Hence, in Italy administrative support to tax courts is not provided by the justice ministry but by the finance ministry [7].

---

[4] C. GLENDI, *Riforma della giustizia tributaria. Basta con le "normicciuole" sulle liti minori*, in *Diritto e Pratica Tributaria*, 2023, p. 957 ff.

[5] Before the 2022 reform of the Italian tax judiciary, tax judges were not career magistrates: they were appointed on a part-time basis among professionals with specific qualifications, such as lawyers, certified public accountants, and other practitioners with recognised experience in tax matters, as well as certain categories of public officials, including retired judges, university professors, and civil servants from the Ministry of Economy and Finance. The full tax judges would have the same economic treatment as ordinary judges. Both law graduates and business graduates can sit the exam. See R.M. PALMIERI, *Il reclutamento dei nuovi magistrati tributari*, in *Questione giustizia*, 2024.

[6] Article 16 Legislative Decree 545/1992.

[7] There has been a lot of criticism towards this. See F. TUNDO, *L'indipendenza del giudice dipendente del MEF, ossimoro di una riforma che gioca d'azzardo*, in *Giustizia insieme*, 2022. By Order No. 403 of 31 October 2022, the Venice Tax Court of Justice referred to the Constitutional Court the amendments introduced by Law No. 130/2022 to Legislative Decree No. 545/1992, arguing that the reform not only maintained but strengthened the subordination of tax judges to the Ministry of Economy and Finance, undermining their independence and impartiality. The Constitutional Court, however, in Judgment No. 204 of 17 December 2024, declared the constitutional challenge inadmissible.

2.2. The supreme jurisdiction to the Court of Cassation

The tax process is divided into two levels of judgement: a first level before the Tax Courts of First Instance and a second level before the Tax Courts of Second Instance. Pursuant to Article 111 of the Italian Constitution, an appeal to the "Corte di Cassazione" (Italy's supreme court) is allowed against second instance judgments.

The tax process starts with the taxpayer challenging a tax act through serving an appeal ("ricorso") to the Tax Court of First Instance. The tax process thus has an appeal structure and is aimed at verifying the validity and legitimacy of the contested act. Appeals against the judgments of the Tax Courts of First Instance can be taken to the Tax Courts of Second Instance. In the court of second instance, the judge decides again on the merits of the matter. However, the judge can decide only on the requests alleged by the appellant.

Differently from many other legal systems, in Italy the Court of Cassation is the court of last instance also in tax matters [8]. The judgments of the Tax Courts of Second Instance may be appealed in the Court of Cassation on the grounds provided for in Article 360 para. 1 of the Code of Civil Procedure (violation of the rules on jurisdiction, violation or wrong application of the rules of law, nullity of the judgment and of the proceedings, failure to examine a decisive fact for the judgment that was the subject of discussion between the parties). The judgement before the Court of Cassation is brought by means of serving an appeal to the counterparty/resistant. According to Article 62 para. 2. Legislative Decree 546/1992, the rules of the Code of Civil Procedure apply to the case before the Court of Cassation. However, tax disputes are referred to a specific section of the court, Section n. 5 [9].

---

[8] For a comparative perspective, see European Commission for the Efficiency of Justice (CEPEJ), *Study on the functioning of judicial systems in the EU Member States*, 2022.

[9] It is composed of several chambers (three chambers deal with general civil law decisions, one with labour law decisions, and six with criminal law decisions). Decisions are taken by panels composed of five judges. A special panel of seven judges, the so called "Sezioni Unite", decides on: matters of jurisdiction; cases decided in conflicting ways by the ordinary chambers; and cases of special importance. The Court of Cassation is

The main functions attributed to the Court of Cassation by the Fundamental Law on the Judiciary No. 12 of January 30, 1941, ensure "the exact observance and uniform interpretation of the law, the unity of national law, and respect for the limits of the different jurisdictions". The crucial role of the Court of Cassation within the Italian legal system is related to its nomofilactic function, i.e. ensuring certainty in the interpretation of the law [10]. For this purpose, the provisions in force do not allow the Court of Cassation to know the facts of a case except when they result from the documents already acquired in the stages preceding the trial and only to the extent that it is necessary to know them in order to carry out its role.

Hence, the Court of Cassation is a court of legitimacy and not of merit and its decisions concern the legitimacy and correctness of the appellate judgement [11]. This means that the claimant could, for example, argue that the judgement misapplied a rule because it misinterpreted a provision or that the second instance decision is affected by a procedural defect that resulted in its nullity. However, they cannot ask for a review of the facts.

The final decision of the Supreme Court may be to uphold or dismiss the second instance decision [12]. The upholding may

---

composed of around 500 full judges who enter the court at the height of their judicial careers. They are neither chosen nor elected: they are appointed based on the rank they have reached in their career.

[10] For a critical reflection on this topic, see A. ALPINI, *La funzione "nomofilaetica" della Corte di Cassazione*, in *Il giusto processo civile*, 2016, p. 219 ff.; F.S. DAMIANI, *La funzione nomofilattica della Corte di Cassazione tra mito e realtà*, in D. DALFINO (ed.), *Problemi attuali di diritto processuale civile*, 2021, p. 217 ff.; For an overview of the institutional debate, see G. DE AMICIS, E. VINCENTI, M. ACIERNO, *Il ruolo della Cassazione: tradizione e mutamenti*, Relazione tematica a cura dell'Ufficio del Massimario e del Ruolo, 2011.

[11] Concerning the reform of civil proceedings before the Court of Cassation, particularly its procedural innovations, the impact on the Court's nomophylactic function, and the broader debate on its institutional role, see G. MONTELEONE, *Il nuovo volto della Cassazione civile*, in *Rivista di Diritto Processuale*, 2006, p. 943 ff.; A. PROTO PISANI, *Crisi della Cassazione: la (non più rinviabile) necessità di una scelta*, in *Foro it.*, 2007, p. 122 ff.; A. TEDOLDI, *La nuova disciplina del procedimento di Cassazione: esegesi e spunti*, in *Giurisprudenza Italiana*, 2006, p. 2002 ff.; M. TARUFFO, *Una riforma della Cassazione civile?,* in *Rivista Trimestrale di Diritto e Procedura Civile*, 2006, p. 759 ff.;

[12] For further analysis, see G.F. RICCI ALBERGOTTI, *Il giudizio civile di Cassazione*, Torino, 2025.

be with or without referral to the Tax Court of Second Instance. In the case of upholding without referral, the Supreme Court both eliminates and substitutes the previous decision with a new one. This happens when no further factual findings are necessary. When the Supreme Court decides with referral, the Court only rules on the legitimacy of the appellate judgment. It is up to the Tax Court of Second Instance to decide on the merits of the case [13].

A recent reform introduced the possibility for the tax courts to make the so-called "national reference for preliminary ruling" [14]. In this procedure the referring court may ask to the Court of Cassation which is the correct interpretation of a certain provision. The national reference for preliminary ruling is possible if: the resolution of the case depends merely on the interpretation of the law; the question has not already been decided by the Court of Cassation; there are interpretative difficulties at stake; and the question is likely to arise in many judgments [15]. In this case the Court of Cassation answers with an *interpretative principle of law*, clarifying how to interpret a legal provision. This is interesting for our purposes because it shows the importance of these interpretative statements for courts of first (or second) instance.

---

[13] For a critique perspective on the procedural and institutional developments of the Court of Cassation that undermine its nomophylactic role, due to the huge amount of cases to be decided, overly complex jurisprudence that weaken clarity and consistency in precedent, and excessive judicial individualism, see C. Consolo, *La Cassazione e il suo nuovo volto "gianuario" (doppio ma, infine "disambiguato"?)*, in A. Di Porto (ed.), *La nuova Cassazione civile*, Padova, 2017, p. 49 ff.; C. Di Iasi, *La fata ignorante (a proposito di Ufficio del Massimario e funzione di nomofilachia)*, in *Questione Giustizia*, 2017, p. 82 ff.

[14] Article 363-bis of the Code of Civil Procedure.

[15] *Ex multis*, see B. Capponi, *La Corte di cassazione e la «nomofilachia» (a proposito dell'art. 363 c.p.c.)*, in *Il processo*, 2020, p. 405 ff.; F.S. Damiani, *Il rinvio pregiudiziale in cassazione*, in *Il Giusto Processo Civile*, 2023, p. 55 ff.; F. Santagada, *Rinvio pregiudiziale in cassazione*, in R. Tiscini (ed.), *La riforma Cartabia del processo civile*, Pisa, 2023, p. 524 ff. On the possibility to make these referrals in the tax field, see P. Chiaranda, *Ammissibili i rinvii pregiudiziali alla Corte di cassazione proposti dai giudici tributari e su questioni di giurisdizione*, in *Rivista di Diritto Processuale*, 2024, p. 1379 ff.; F. Pistolesi, *Il primo caso di rinvio pregiudiziale alla Corte di Cassazione in materia tributaria*, in *Giustizia Insieme*, 2023.

2.3. The argumentative structure of the judgements

The decisions of the Italian Supreme Court have two main features. First, they follow the normal structure of higher court judgments, which is deductive. The objective is to present the final decision as a conclusion stemming from a chain of consistent logical steps, moving from given premises and arriving at a necessary conclusion [16]. The theoretical model that inspires the logical structure of the judgment is syllogistic deduction. Even when the actual structure of the judgment is more complex and not a mere logical syllogism, the prevailing tendency is to shape the justification of the judgment as a series of logical passages [17]. Second, judgments delivered by the Court of Cassation are answers to the "grounds of appeal" that the party has stated. Therefore, the judgment is also structured as a sequence of answers that the court gives to the issues raised by the parties [18].

The drafting style of higher court judgments has the following features [19]:

(i) legalistic: legal professional jargon is normally used; legal technicalities are always used; opinions are often written as if they were doctrine; the writing judge tries to demonstrate their legal and scientific culture; and choices of policy and discretionary evaluations are usually hidden behind technical legal arguments and formal reasoning concerning the meaning of the statutory rules involved.

(ii) magisterial: it is the court that formally delivers a judgment; there are no single judges expressing their own personal opinions about the case.

(iii) importance of interpretative role: the arguments supporting judgements deal with the interpretation of the principles and laws involved. ln many cases the court states in its decision that its interpretation is the 'proper and correct' interpretation of a statutory rule.

---

[16] N. MACCORMICK, R.S. SUMMERS, *Interpreting precedents. A comparative study*, cit., p. 146.

[17] *Ibid.*, p. 147.

[18] *Ibid.*

[19] *Ibid.*, p. 147-148.

3. The conceptual relationship between "principi di diritto" and the CJEU formulas

3.1. The growing importance of "principi di diritto"

Within the Italian Supreme Court, the ROCC plays a key role in ensuring the correct and uniform interpretation of the law [20]. The Office's task is the identification of *interpretative principles of law*. This identification takes place through the drafting of so-called "massime" [21].

The process leading to the publication of *massime* begins with the selection of the most significant decisions [22]. The selection process consists in reading all the decisions of the Civil and Criminal Sections of the Court of Cassation (except for the Orders of the Seventh Section) in order to extract the decisions containing *interpretative principles of law* with a nomofilactic nature, i.e. that are necessary to ensure the uniform application of law in the country.

The selection is followed by drafting, reviewing, and verifying the correctness of the *massime*. Each stage is performed by judges belonging to the ROCC. The final verification is carried out by the Directorate of the ROCC [23].

A *massima* can be defined as a concise statement of the *interpretative principle of law* affirmed by the Court on which a decision is based [24]. *Massime* are different from summaries and abstracts of judgments published, for example, by journals or databases. These summarise the facts of the case and the court's decision but do not focus on the extraction of *interpretative principles of law*. Furthermore, a *massima*, to be considered as such, must include the *interpretative principle of law*, cannot be the summary of the decision, and cannot be the mere reproduction of argumentative passages of the court's decision.

---

[20] M.R. San Giorgio, *L'Ufficio del Massimario e del Ruolo*, in M. Acierno, P. Curzio, A. Giusti (eds.), *La Cassazione civile. Lezioni dei magistrati della Corte suprema italiana*, III ed., Bari, 2020, p. 371 ff.

[21] The dissemination of *massime* is ensured by adding them to a national database (Italgiure) created and updated by the "Centro Elaborazione Dati" (CED) of the Court of Cassation.

[22] L. Nazzicone, *La massimazione delle sentenze*, Padova, 2021, p. 9 ff.

[23] G. Fanticini, G. Sessa (eds.), *Vademecum per lo spoglio e la massimazione*, 2023.

[24] *Ibid.*

To pursue the objective of granting the uniform application of law, the drafting of the *massime* is performed following some rules, which can be classified into three different types: f*ormal rules* to ensure the uniform drafting of massime; *selection rules* for the uniform identification of the *interpretative principles of law* that are used to build a *massima;* and *linking rules* to build a network of consistent precedents in the Italgiure database [25]. The latter create relations between *massime* (e.g. the tags "conforming" and "dissenting" show that a *massima*, respectively, coheres with or differs from another).

In addition, the drafting of *massime* must be performed through following some general principles: adherence to the decision; synthesis in the enunciation of the principle; and clarity and precision of the final text.

There are various types of *massime*: "ordinary" ones, which contain only a new nomofilactic *interpretative principle of law*; "conforming" ones, which reproduce an already maximised principle; and "massime with facts", which in addition to the *interpretative principle of law*, refer to the concrete case on which the Supreme Court has decided in application of the general principle [26].

The *massima* consists of two parts: (a) the preamble, which identifies the subject matter and the topic of the decision; (b) the enunciation of the principle. As for letter (a), the first part of each *massima* is a summary consisting of: a numerical code which refers to a classification scheme by subject matter and the keywords representing the content of the maximised principle [27]. The keywords follow a specific order as they must represent the logical argumentation of the decision.

The core part of a *massima* is the statement of the *interpretative principle(s) of law*. It represents the generalisation of the interpretation and application of the rule to a concrete case. The *interpretative principles of law* are not a source of law in the Italian legal system. Their formulation is the result of an interpretative activity. This action consists in the extraction from a specific case of a general rule that can apply also in future similar cases.

---

[25] *Ibid.*, p. 5 ff.
[26] *Ibid.*, p. 6 ff.
[27] *Ibid.*, p. 9 ff.

3.2. "Principi di diritto", CJEU formulas, and Judicial Interpretative Formulas

An *interpretative principle of law* is not an abstract principle, but a principle governing the law applied to the case [28]. Even if it is the criterion for deciding the concrete case that can be applied to similar or equal cases, the *interpretative principle of law* is different from the *ratio decidendi*. The *ratio decidendi* is the logical path (i.e. a series of logical steps) which justifies the adoption of a certain decision in one decision [29]. It cannot be generalised nor replicated to future cases. Hence, we can argue that the *ratio decidendi* is an abstract concept pointing at the essential passages of the argumentation of the judge, which are needed to pursue a certain conclusion in relation to a specific case. The *interpretative principle of law* is a statement which shortly gives evidence of the way in which the law has been interpreted and applied to a scenario and can be applied to similar cases in the future.

The *interpretative principle of law* aims at universalising the individual decision, i.e. creating a statement which can be applied to future similar cases [30]. Hence, the formulation of the *interpretative principle of law* has the objective to give hermeneutic indications to interpreters. The selection and publication of *interpretative principles of law* ensures the foreseeability of decisions, the coherence of the system, and a deflation/reduction of litigation. The final aim is to guarantee stability, predictability, and legal certainty.

Foreseeability in civil law systems is linked to the stability of precedents: the existence of "deep and persistent" case law divergences within national courts of last instance has a direct impact on fair trial [31] (see par. 4.1).

The Court of Cassation does not collect all the *interpretative principles of law.* As anticipated it creates

---

[28] G. Gorla, *Precedente giudiziale*, in *Enciclopedia Giuridica Treccani*, 1990, p. 11.

[29] G. De Amicis, *La formulazione del principio di diritto e i rapporti tra Sezioni semplici e Sezioni Unite penali della Corte di Cassazione*, in *Diritto Penale Contemporaneo*, 2019, p. 2.

[30] *Ibid*.

[31] L. Nazzicone, *Tenciche di massimazione delle sentenze*, cit., p. 10.

massime only when the decision contains important *interpretative principles of law*. These are considered "important" when they comply with substantive criteria and temporal criteria. As for substantive criteria the Supreme Court focuses on decisions including *interpretative principles of law* which solve a conflict of case law or a question of particular importance; new principles of law; *interpretative principles of law* which are divergent from precedents; *interpretative principles of law* which concern relevant cases, also because of the particular social impact of the issue [32].

However, the publication of a principle is excluded when it consists in the: mere repetition of a norm; definition of a notion; intermediate step functional to the statement of the *ratio decidendi*; incidental enunciation of principles; digressions from the *ratio decidendi* (*obiter dictum*); the formulation of general and abstract principles that are completely outside the subject-matter (for example, statements used to build parallelisms, distinctions, and analogies) [33].

Within the case law of the Court of Justice, "formulas" is an expression coined in European literature to designate legal statements that express abstract legal propositions intended to guide the application of European law [34]. These formulas are originally drawn up for specific disputes and then repeated in later decisions with no or only minimal changes (so called "copy-pasting" technique [35]). In doing so, they become relevant in fields of law others from those in which they originated [36]. This process of reiteration is facilitated by a drafting practice that has been described as "copy-pasting", i.e. the verbatim or slightly altered reproduction of passages from earlier judgments irrespective of their binding character [37].

---

[32] *Ibid*.

[33] G. FANTICINI, G. SESSA (eds.), *Vademecum per lo spoglio e la massimazione*, cit., p. 6.

[34] L. AZOULAI, *La fabrication de la jurisprudence communautaire* in P. MBONGO, A. VAUCHEZ (eds.), *Dans la fabrique du droit européen*, Bruylant, 2009, p. 163 ff.

[35] F-X. MILLET, *In the name of analogy: judicial copy-pasting and competence creep in the connection data case law*, in *Common Market Law Review*, 2024, p. 1289 ff.

[36] L. AZOULAI, *La fabrication de la jurisprudence Communautaire*, cit.

[37] B. SOUCEK, *Copy-paste precedent*, in *The Journal of Appellate Practice and Process*, 2012, p. 153 ff.

Hence, while the *interpretative principle of law* is a concrete statement showing how the law was interpreted and applied in a case, with the aim of guiding future similar decisions and ensuring coherence, predictability, and legal certainty, formulas are abstract legal propositions originating in specific disputes that are repeatedly reproduced, frequently without adaptation.

However, the role of *interpretative principles of law* has changed in the last two decades: once meant mainly to guide the judge hearing an appeal, they increasingly take the form of abstract normative propositions clarifying or specifying the meaning of statutory provisions, rather than reflecting the specific facts of the decided case [38]. From being a directive for correctly applying the law in remanded cases, it has become a tool for interpreting the law accurately, with potentially universal scope [39]. Where facts are mentioned, they are typically "distilled" into generic categories, detached from the concrete circumstances that generated them [40].

These principles are frequently invoked by successive case law as autonomous normative statements, often regardless of their original factual context [41]. Literature has argued that this phenomenon is due to the need for legal certainty, the number of decisions to be decided by the Italian Supreme Court and a change in the legal-cultural tradition [42]. The practice of reasoning based on legislation has led to shape case law precedents into abstract formulations resembling statutory norms, unknown to the common law world [43]. Recent legislative amendments reflect a new perception of the *interpretative principle of law.*

---

[38] L. Passanante, *Il postulato del primo Calamandrei e il destino della Cassazione civile,* paper presented on 11 November 2020, during the study meeting entitled *"Passato e futuro della Cassazione – A cent'anni dalla "Cassazione Civile" di Piero Calamandrei",* organized by the Italian School for the Judiciary – Decentralized Training Unit of the Court of Cassation.

[39] *Ibid*.

[40] L. Nazzicone, *Tecniche di massimazione delle sentenze*, cit., p. 28.

[41] L. Passanante, *Il postulato del primo Calamandrei*, cit.

[42] M. Barberis, *Contro il creazionismo giuridico. Il precedente giudiziale fra storia e teoria*, in *Quaderni fiorentini per la storia del pensiero giuridico moderno*, XLIV, 2015, p. 67 ff.

[43] F. Galgano, *L'interpretazione del precedente giudiziario*, in *Contratto e impresa*, 1985, p. 702.

Under the current version of Article 384 of the Code of Civil Procedure, the Court of Cassation is required to formulate a *interpretative principle of law* when deciding a case, while the reform of Article 363 introduced the so-called "principio di diritto nell'interesse della legge" (in Eng: *interpretative principle of law established in the interest of the law*). Moreover, since the 2006 reform, the Court may also pronounce such a principle *ex officio* when declaring an appeal inadmissible, provided that the question addressed is of particular importance [44]. Finally, in most recent decisions the Cassazione frequently reproduces, the jurisprudence of the CJEU, a practice which may also influence its drafting style.

Hence, there is a kind of approximation between formulas and *interpretative principles of law*. Both tend to become increasingly detached from the factual circumstances of the original case, and both are frequently invoked in subsequent case law as literal quotations, thereby acquiring a life of their own. Over time, they occupy a growing portion of judicial reasoning (par. 3.3).

3.3. Theoretical analysis, empirical analysis, and the unitary definition of formulas

Since formulas and *interpretative principles of law* pursue similar purposes, in our project we adopted the unitary concept of Judicial Interpretative Formula (JIF) for both European and national case law. A Judicial Interpretative Formula is a segment of text taken from the reasoning section of a judgment that expresses (i) the interpretation of a legal rule, part of a rule, or a general principle; (ii) the legal consequences arising from the interpretation or application of that rule or principle within the legal system; (iii) the subsumption of a fact under a legal rule; or (iv) the classification of a factual scenario as falling within a legal concept contained in a rule. Hence, every *interpretative principle of law* is considered a formula, but not every formula qualifies as *interpretative principle of law.*

[44] L. PASSANANTE, *Il precedente impossibile. Contributo allo studio del diritto giurisprudenziale nel processo civile*, Padova, 2018, p. 61.

An empirical analysis of the text of national decision, revealed several distinctions between formulas and *interpretative principles of law.*

First, in the CJEU, the copy-pasting of formulas through case law pursues the non-explicit objective of advancing harmonisation. Court of Cassation's *interpretative principles of law* preserve a more explicit role as interpretative statements of general scope. Second, the Italian Supreme Court systematically organises *interpretative principles of law* through the ROCC that has an institutional nomophylactic function. Moreover, CJEU formulas are never explicitly qualified as such. Formulas concerning the specific application of law to the case at hand can be found at the end of the motivation concerning each preliminary question. In most Court of Cassation's rulings, *interpretative principles of law* directly related to the case at stake are expressly labelled as such and generally appear at the end of the judgment.

The empirical analysis also revealed several distinctions between national and European JIF.

First, CJEU judgments tend to be longer, and they have more formulas. Second, the CJEU never explicitly overrules its previous formulas, whereas the Court of Cassation openly acknowledges revirements, distinguishing or overruling earlier orientations, often identifying the consolidated position, indicating the decision in which it was revised, and drawing explicit conclusions. Finally, from a formal perspective, numbered paragraphs are common in the CJEU decisions, but only occasionally present in Court of Cassation judgments, a feature that significantly affects automated extraction. CJEU formulas typically correspond to a single paragraph, while Court of Cassation decisions are less uniformly structured, with longer quotations and varying paragraph lengths; explicit *interpretative principles of law* usually occupy a single paragraph, but non-explicit ones may be shorter or considerably longer.

Given the structural and functional differences between national and supranational jurisprudence, we developed specific criteria for identifying and extracting formulas from national case law. Moreover, for national case law, we distinguish formulas by novelty, role, and origin. In terms of novelty, formulas may be "new", when they contain all the features of a formula without citing earlier decisions, or

"cited", when they reproduce pre-existing formulations; this classification is necessarily time-bound, as every new formula may later become a cited one. We distinguished formulas that are explicitly identified as such, corresponding to the end of the decision where the Court states the "principio di diritto" and expressly labels it as such. This distinction would not be meaningful in the context of the CJEU. Finally, with respect to origin, in national case law cited formulas may derive either from national case law of the Court of Cassation or from the jurisprudence of the CJEU.

4. The development of JIFs and the role of precedent in the VAT case law

4.1. Court of Cassation: No binding precedent but growing importance of precedents

JIFs are closely tied to the broader discourse on precedent. The growing prominence of these statements issued by both national and supranational courts reflects the increasing importance that prior decisions carry in resolving new cases.

The concept of precedent is typical of common law countries, where *stare decisis* applies, i.e. where courts and judges should be consistent with decisions, rulings, and opinions from prior cases [45].

Notwithstanding this definition, in the Italian system the word "precedent" is mainly used in a very broad sense, meaning any prior decision possibly relevant to a case. In this broader meaning, previous judgements of the Court of Cassation are often considered as precedents because they should be taken into account (= analysed, read) when deciding a case [46].

The Italian system usually refers to the proper meaning of

[45] In these countries a precedent is any prior decision dealing with the same legal issue as a later decision, with the former providing a model for the latter (N. MACCORMICK, R.S. SUMMERS, *Interpreting precedents. A comparative study*, Ashgate Publishing Limited, 1997, p. 1). On the fact that the use of the term has ended up involving Italian law, see F. GALGANO, *L'interpretazione del precedente giudiziario*, in *Contratto e Impresa*, 1985, p. 701 ff.

[46] N. MACCORMICK, R.S. SUMMERS, *Interpreting precedents*, cit., p. 151.

the term "precedent" (= a previous decision which is used as a model for a new one) in respect of the judgements of the Constitutional Court and Sezioni Unite of the Court of Cassation. These are usually considered to be judgments whose argumentation should be respected in a later case. In the last decades judicial precedents have become a major foundation for court decisions, often serving as a medium for interpreting laws, codes, and constitutions as judges rarely apply legal rules in their pure textual form, instead frequently interpreting them through prior case discussions[47]. Judgements of courts of merit, however, are usually not considered to be precedents, neither in the strict nor in the broad sense of the term.

In the Italian legal system there is no codification of the principle of binding precedent in the strict sense (*stare decisis*), nor has the Court of Cassation ever decided that it should be an inherent principle of our legal system. Nonetheless, there is a trend in Italian legal courts not to deviate from an established interpretation of the Court of Cassation if there are no strong reasons to do so[48].

There are three main reasons for this trend towards respecting precedents even if they are non-binding.

First, legislative sources of law are undergoing a crisis which is caused by several factors: the time taken by the legal procedure when compared with the rapidity of the phenomena it regulates; the poor quality (complexity, vagueness, etc.) of recent laws; the importance of courts in

---

[47] *Ibid.*, p. 152.

[48] Bibliography on this topic is extensive. See M. Bin, *Precedente giudiziario, "ratio decidendi" e "obiter dictum"*, in *Rivista Trimestrale di Diritto e Procedura Civile.*, 1988, p. 1004 ff.; G. Gorla, *Precedente giudiziale*, in *Enciclopedia Giuridica Treccani*, 1990, XVII, p. 2 ss.; A. Maniaci, *Il precedente giudiziale in Italia: verso lo 'Stare decisis'?*, in *Rivista Critica di Diritto Privato*, 2014, p. 567 ss.; U. Mattei, *Precedente giudiziario e stare decisis*, in *Digesto delle Discipline Privatistiche, sez. civ.*, XIV, Torino, 1996, p. 149 ss.; R. Rordorf, *Stare decisis: osservazioni sul valore del precedente giudiziario nell'ordinamento italiano*, in *Il Foro Italiano*, 2006, p. 279 ss.; G. Sartor, *Il precedente giudiziale*, in *Contratto e Impresa*, 1995, p. 1300 ss.; M. Taruffo, *Dimensioni del precedente giudiziario*, in *Rivista Trimestrale di Diritto e Procedura Civile.*, 1994, p. 411 ss.; G. Visintini (ed.), *La giurisprudenza per massime e il valore del precedente,* Padova, 1988.

the process of European integration; the welfare state and globalisation, which mean that citizens frequently ask courts for the protection of new or implicit rights [49]. These factors are common to all the legal systems of European countries. Hence, the trend towards case law in Italy is coherent with the trend in other civil law legal systems such as France and Germany. There are several national and supranational studies which show that case law is becoming increasingly important, and courts tend to rely on and follow previous case law [50].

Second, the procedural reforms that have taken place in Italy in recent years were inspired by the strengthening of the principle of binding precedent (e.g. reforms concerning art. 374 and 618 para. 1-bis of the Code of Civil Procedure) [51]. Since 2006, reforms have renewed interest in precedents in a "normative" sense which differ from common law, consisting of general and abstract legal rules stated by courts, similar to principles set by the Court of Cassation. They are rules consolidated through multiple rulings, forming a consistent judicial orientation or "living law", especially when endorsed by the Constitutional Court [52]. The enhanced emphasis on the role of *interpretative principles of law*, as previously discussed, can be seen as part of this trend.

Third, within the Court of Cassation, the case may be decided by a special panel of seven judges. This happens when the chairman of the court finds that the case deals with especially important issues of law, or in case of conflicting

---

[49] This reconstruction is based on the analyses and studies of M. BARBERIS, *Beccaria, Bentham and Legal Creationism*, in *International Review of the Philosophy of Law*, 2014, p. 569 ff.; M. CAPPELLETTI, *Legislating Judges?*, Milano, 1984, p. 19 ff.; S. CASSESE, *Global Law*, Torino, 2009; F. GALGANO, *Globalization in the Mirror of Law*, Bologna, 2005; S. RODOTÀ, *End-of-Century Repertoire*, Roma-Bari, 1999, p. 169.; G. ZACCARIA, *Case Law as a Source of Law*, in Id., *Understanding the Law*, Roma-Bari, 2012.

[50] See, for example, T. KOOPMANS, *Legislature and Judiciary: Present Trends. New perspectives for a common law of Europe/Nouvelles perspectives d'un droit commun de l'Europe*, Firenze-Leiden, 1978.

[51] G. AMOROSO, *Massime e principi del diritto vivente*, in A. DIDONE, F. DE SANTIS (eds.), *I processi civili in Cassazione*, Milano, 2018, p. 159

[52] G. DE NOVA, *La giurisprudenza fonte del diritto*, in *Juscivile*, 2016, p. 419; F. GALGANO, *L'interpretazione del precedente*, cit., p. 702.

judgements [53]. This creates problems in relation to precedents, since conflicting or inconsistent judgments on the same issue cannot represent an effective precedent. In the above-mentioned cases the special panel is expected to solve the conflict with a judgment that, being uniform and authoritative, amounts to a precedent for future cases [54]. Other courts or other chambers of the Court of Cassation are not formally obliged to comply with the judgment delivered by the Sezioni Unite. However, their judgments are usually considered as especially authoritative, and they are normally used as precedents in the strict meaning of the concept.

Stable case law is defined in Italy as "living law". The existence of "living law" does not exclude a subsequent evolution of case law. However, lower courts usually conform to the living law produced by higher courts because of its persuasive/authoritative value [55].

Especially in recent years, in the tax field the legislature frequently referred to "living law" or to "stable case law" when introducing its reforms (see Reform of the Statute of Taxpayers Rights 2023) [56]. This means that the legislature frequently codified principles which were first stated in case law, or topics in relation to which there were contrasting

---

[53] The "especially important legal issue" should be interpreted as meaning that there is an issue in relation to which it is crucial to set an authoritative precedent. The case of conflicting judgments means that different chambers of the Court of Cassation adopted different decisions on the same issue.

[54] On the importance of the decisions taken by Sezioni Unite, see F. Auletta, *Profili nuovi del principio di diritto (il «vincolo delle sezioni semplici al precedente delle sezioni unite»*, in E. Fazzalari (ed.), *Diritto processuale civile e Corte costituzionale*, Napoli, 2006, p. 1 ff.; R. Aprati, *Le sezioni unite fra l'esatta applicazione della legge e l'uniforme interpretazione della legge (commi 66-69 l. n. 103/2017)*, in A. Marandola, T. Bene (eds.), *La riforma della giustizia penale*, Milano, 2017, p. 278 ff.; G. Fidelbo, *Verso il sistema del precedente? Sezioni Unite e principio di diritto*, in *Diritto Penale Contemporaneo*, 2018.

[55] M. Cavino, *Diritto vivente*, in *Digesto delle Discipline Pubblicistiche*, 2010.

[56] For a comment to the reform, see A. Giovannini (ed.), *La riforma fiscale – I diritti e procedimenti*, vol. II, Ospedaletto (Pisa), 2024; M. Logozzo (ed.), *L'attuazione della riforma tributaria*, Ospedaletto (Pisa), 2024, E. Manzon, G. Melis (eds.), *Il diritto tributario nella stagione delle riforme*, in *Giustizia insieme*, 2024.

opinions in case law. For example, in the 2023 tax reform the Italian legislature introduced some principles which were codified first by the CJEU and recognised by the Court of Cassation's case law after some years of contrasting opinions (e.g. *ne bis in idem*[57]). Literature as well frequently looks at "living law" in dealing with specific issues. This is also quite frequent in the field of VAT.

4.2. The role of precedents in the jurisprudential harmonisation of VAT

In tax law, the importance of case law is generally increasing, but in the field of VAT this trend is particularly evident, especially due to the role that the Court of Justice has played in shaping and integrating this tax.

For harmonised taxes the CJEU is competent to provide the correct interpretation of EU law. When national judges have doubts concerning the application of EU law, according to Art. 276 TFEU, they must (if they are courts of last instance) or can (if they are not) ask the CJEU for a preliminary ruling. Hence, the role of case law is more important in harmonised taxation (see Ch. IV). Italy is one of the EU countries which make the highest number of preliminary references to the CJEU[58].

Many Italian scholars argue that the CJEU's interpretative activity in the field of taxation has reached its maximum intensity in respect to VAT, as this tax has comprehensive European legislation[59]. Its harmonisation started in the 60s

---

[57] E.M. BAGAROTTO, *Prime considerazioni sull'introduzione del divieto di bis in idem nel procedimento tributario*, in *Tax news*, 2024, pp. 14–22; C. CONSOLO, *Il divieto di* bis in idem *nel procedimento tributario*, in *La riforma fiscale – I diritti e i procedimenti*, cit., p. 65 ff.

[58] In 2022, the German, Italian, Polish, and Bulgarian courts made the highest number of requests for a preliminary ruling to the Court of Justice. The total number of references for preliminary ruling made by Italy in the field of VAT is 44 (the total number of preliminary rulings in the field of VAT is around 1.000). The countries which made a similar number of preliminary references are Belgium; Bulgaria; and Romania. The countries which made much more references are Germany and The Netherlands.

[59] R. MICELI, G. MELIS, *Le sentenze interpretative della Corte di Giustizia delle Comunità Europee nel diritto tributario: spunti dalla giurisprudenza*

and so CJEU case law is very extensive. Moreover, it has developed over time, and the CJEU feels "more confident" to adopt decisions in which the court has a significant margin of creation and discretion [60]. The Court has increasingly played a role in supplementing European law, giving case law a position of prominence over statutory law [61]. Some authors have described this as a creative function in tax matters [62] (see par. 5.1).

As explained in Ch. IV, the CJEU has an interpretative method which differs from the national method (e.g. it uses interpretation by principles much more than the Italian Supreme Court, and it uses comparative interpretation, and interpretative criteria, such as economic reality, which are not used in the national legal system). Some of these features are the result of the wording of the European legislation, which is characterised by conciseness, vagueness, (sometimes) non-coincidence of translations, and explicit references to the function of the legislation (in particular in the recitals). Moreover, the CJEU provides definitions of terms which are not defined by the legislation.

The use of these new interpretative methods, criteria, and concepts also influences national case law in the long run. For example, national courts in the field of harmonised taxes started to refer to concepts and criteria (e.g. economic reality [63]) which did not exist in the Italian legal system. Moreover, national case law concerning harmonised taxes

---

relativa alle direttive sulla "imposta sui conferimenti" e sull'IVA, in *Rivista di Diritto Tributario,* 2003, pp. 113-114.

[60] On the harmonization of VAT and the role of the CJEU in this process, see P. Adonnino, *L'armonizzazione fiscale dell'Unione europea,* in *Enciclopedia del Diritto*, Agg. III, Milano,1999; A. Comelli, *IVA comunitaria e IVA nazionale*, Padova, 2000; Id., *L'armonizzazione fiscale e lo strumento della direttiva comunitaria in relazione al sistema dell'Iva*, in *Diritto e Pratica Tributaria*, 1998, p. 1590 ff.; R. Cordeiro Guerra, *L'Iva quale imposta sui consumi: riflessi applicativi secondo la Corte di giustizia*, in *Rassegna Tributaria*, 1996, p. 322 ff.; F. Fichera, *L'armonizzazione delle accise*, in *Rivista di Diritto Finanziario e Scienza delle Finanze*, 1997, II, p. 216 ff.; P. Russo, R. Cordeiro Guerra, *L'armonizzazione fiscale nella Comunità Europea*, in *Rassegna Tributaria*, 1990, p. 629 ff.

[61] P. Adonnino, *Armonizzazione fiscale nell'Unione europea*, cit., p. 280.

[62] R. Miceli, G. Melis, *Le sentenze interpretative*, cit.

[63] A. Comelli, *La "realtà economica e commerciale" quale "criterio fondamentale" per l'applicazione della disciplina IVA*, in *Diritto e Pratica Tributaria Internazionale,* 2022, p. 99 ff. In supranational literature see A.

increasingly relies on the concepts and categories which have been developed at the EU level. The national courts frequently refer to the definition of concepts of the VAT Directive which are provided by the CJEU.

Finally, also the Italian legislature in the recent reform of taxation paid attention to the case law of the CJEU. Law 111/2023 for the reform of tax law emphasises the need for a revision of the definition of the taxable operations and exemption rules "in order to make them more consistent with European Union legislation". In particular, with regard to exemptions (one of the focuses of the POLINE project), the Law 111/2023 provides for a review of the provisions governing exempt transactions in order to ensure greater consistency with Court of Justice case law. The fact that even the legislature is modifying national VAT legislation to make it more consistent with EU case law shows the crucial role of case law in this field.

For this reason, the systematic extraction of JIFs from both national and European case law, followed by network analysis of these data, can play an essential role in strengthening the coherence of judicial interpretation in the area of harmonised taxation. Over time, the Court of Justice has progressively established the direct applicability of numerous provisions of Union law, affirmed the exhaustive nature of rules providing for tax exemptions or limiting the exercise of the right to deduction, and assigned fundamental importance to the principle of tax neutrality [64]. Making JIFs easily accessible alongside national judgments would allow domestic judges to identify and be consistent with the primacy of EU law. Since traditional *massime* do not cover all decisions (see par. 3.1 and 3.2), the availability of JIFs from national case law on specific topics could provide an invaluable complement. Furthermore, mapping JIFs would facilitate the verification of consistency between national references to EU case law and the CJEU's own use of prior decisions. As noted, the CJEU often cites earlier judgments in ways that differ from the reasoning in its newer rulings; awareness of such divergences

---

VAN DOESUM, F. NELLEN, *Economic reality in EU VAT*, in *EC Tax Review*, 2020, p. 215 ff.

[64] P. ADONNINO, *Armonizzazione fiscale nell'unione europea*, cit., p. 284.

could help national judges to align their reasoning more accurately with EU jurisprudence.

5. JIFs as an expression of the creative role of case law

5.1. The national debate on the creative role of judges

The importance of case law in the field of VAT evokes that of the "creative role" of the case law.

The Court of Cassation has adopted some decisions concerning the creative role of case law. For example, it said that this creative role does not cause an excess of jurisdictional power [65].

The idea that a judge could trespass on the terrain of legislative power is premised on the assumption that there is a clear distinction between the activity of normative production and interpretative activity. In this scenario, the former would amount to an appropriation of the powers of the legislature. However, according to the Court, interpretation does not have a merely heuristic function. On the contrary it is a creative work of applying of law to the case at hand. Hence, the judge's activity is always creative. Thus, the excess of power cannot be invoked.

This decision seems to support the idea that, since any interpretative activity entails a margin of creation, only in extreme cases would courts' decisions be so creative as to violate the division of powers. We believe that the statements of the court should be interpreted in relation to the specific cases.

Notwithstanding the above-mentioned decisions, there is an ongoing debate in Italy concerning the creative role of case law. This debate mainly concerns the Court of Cassation.

It is worth emphasizing that Italian doctrine does not make a consistent use of the term "creative jurisprudence" (or "creative role of the judge", or "judge-made law"). Indeed, sometimes these terms are used to refer to the legitimate choice made by a judge to pick the most plausible judicial interpretation within the framework of the possible meanings of a rule. Other times, it is

[65] Judgments (Sezioni Unite) of 15 July 2003, No. 11091, and 25 November 2009, No. 24763.

used to refer to the choice of normative meanings that are outside this framework and therefore are illegitimate [66].

The debate on the creative role of case law mainly spread in Italy through post-positivist doctrines. In particular, the judge's creative role has been widely debated in Italian legal theory since the 1980s [67]. Many experts on constitutional, administrative, and private law dealt with this issue [68]. However, the debate is especially important in criminal matters because, according to the Italian Constitution, criminal regulations must take the form of laws of parliament [69].

In the discussion regarding the creative role of jurisprudence we can identify three orientations [70].

The first orientation focuses on epistemological arguments to stress the importance of judge-made law. According to this

---

[66] L. FERRAJOLI, *Contro la giurisprudenza creativa*, in *Questione giustizia*, 2016, n. 4.

[67] L. FERRAJOLI, *Diritto e ragione. Teoria del garantismo penale*, Bari, 1989; R. GUASTINI, *Lezioni sul linguaggio giuridico*, Torino 1985; ID., *Lezioni sul linguaggio giuridico*, Torino, 1985; G. TARELLO, *L'interpretazione della legge*, Milano, 1980; F. Viola, *Autorità e ordine del diritto*, Torino, 1987; G. ZACCARIA, *Ermeneutica e giurisprudenza. I fondamenti filosofici della teoria di H. G. Gadamer*, Milano, 1984. For older works mentioning the topic see also G. LAZZARO, *Storia e teoria della costruzione giuridica*, Torino, 1965; M.G. LOSANO, *La dottrina pura del diritto dal logicismo all'irrazionalismo*, preface to H. KELSEN, *Teoria generale del diritto e dello Stato* (1945), trad. it. di S. Cotta e G. Treves, Milano, 1952; G. TARELLO, *Il problema dell'interpretazione. Una formulazione ambigua*, ora in ID., *Diritto, enunciati, usi. Studi di teoria e metateoria del diritto*, Bologna, 1974, p. 389 ff. For most recent ones, see L. FERRAJOLI, *Contro il creazionismo giudiziario*, Modena, 2018; R. GUASTINI, *Creazionismo giudiziario?*, in *Ragion pratica. Rivista semestrale*, 2020, p. 391 ff.; M. JORI, *Del diritto inesistente: saggio di metagiurisprudenza descrittiva*, Torino, 2010; R. PARDOLESI, G. PINO, *Post-diritto e giudice legislatore. Sulla creatività della giurisprudenza*, in *Il Foro Italiano*, 2017.

[68] M. CAPPELLETTI, *Giudici legislatori?*, Milano, 1984; S. CASSESE, *Il diritto globale*, Torino, 2009; F. GALGANO, *La giurisprudenza fra ars inveniendi e ars combinatoria*, in *Contratto e impresa*, 2012, p. 77 ff.; P. GROSSI, *L'invenzione del diritto*, Roma-Bari, 2022; S. Rodotà, *Repertorio di fine secolo*, Roma-Bari, 1992.

[69] G. FIANDACA, *Diritto penale giurisprudenziale e spunti di diritto comparato*, in ID. (ed.), *Sistema penale in transizione e ruolo del diritto giurisprudenziale*, Padova, 1997, pp. 5, 6, 11 e 14; ID., *Il diritto penale tra legge e giudice*, Padova, 2002, p. 33 ff.; M. VOGLIOTTI, *Lo scandalo dell'ermeneutica per la penalistica moderna*, in *Quaderni fiorentini*, 2015, p. 131 ff.; ID., *Tra fatto e diritto. Oltre la modernità giuridica*, Torino, 2008.

[70] L. FERRAJOLI, *Contro la giurisprudenza creativa*, cit.

orientation, whose most important representative at the supranational level is Gadamer, to adopt a "right" decision it is sometimes necessary to disregard the law. Moreover, the objective of interpretation is to make law real in relation to the case at hand [71]. This activity necessarily creates a certain distance between the literal interpretation of the rules and how they are applied in practice [72]. He argues that the judge chooses between multiple "interpretations that are all legitimately sustainable" within the perimeter drawn by the principle of legality. Moreover, in contemporary legal systems it is necessary to review and re-analyse the role of the principles of legality and of the prohibition of analogy. These principles were crucial in western legal traditions for several decades but a strict interpretation of them would no longer be appropriate to the current social and economic environment.

The second perspective is that of neo-constitutionalism. This perspective builds on the theories of Dworkin and Alexy. [73]. According to this theory, the existence of constitutions in most western countries led to the distinction between rules and principles. Rules have a more precise meaning than principles. The existence of principles in national legal systems had an impact on interpretation. The growing importance of principles means that judges frequently do not apply the law but rather balance different principles in relation to the case at hand. This is especially evident in the case of the Constitutional Court. For example, this court frequently has to balance between the "balanced budget" and "equality" principles, when deciding whether to extend specific tax benefits to certain categories of taxpayers. In addition, the rules must be interpreted and applied by national judges in the light of the existing principles of their legal system. For example, the existence of the principle of proportionality could result in a less burdensome interpretation of the rules concerning the application of sanctions.

---

[71] H.G. Gadamer, *Truth and method,* 1960.

[72] In Italy, this perspective is taken up by Zaccaria (G. Zaccaria, *La giurisprudenza come fonte di diritto*, cit.).

[73] R. Alexy, *Theorie der Grundrechte*, 1994; R. Dworkin, *Taking Rights Seriously*, 1977. In Italy, see G. Zagrebelsky, *Il diritto mite. Legge, diritti, giustizia*, Torino, 1992; Id., *Il giudice delle leggi artefice del diritto*, Napoli, 2007.

The third perspective is that of neo-pandettism, whose main Italian referent is Paolo Grossi [74]. He argues that the law is the "expression of the will" of the sovereign. From this perspective, the judge is legitimised to produce law because it represents the social conscience. Moreover, "absolute faith" in the principle of legality, as well as the subjection of the judge to the law, must be overcome because of the crisis of law which characterises most western legislatures in the 21st century.

Regardless of their perspectives, the above-mentioned doctrines seem to agree on the increased importance of jurisprudence in Italy. The causes of this increased importance are usually found in [75]:

- The responsibility of politics: disruption of legislative production through obscure, tortuous, and vague legal language; crisis of the general and abstract forms of legal norms; de-regulation, de-legification, and privatisation [76].
- The multilevel structure of current legal systems: existence of constitutions implying that the judge has the power to review the constitutionality of laws; European legal integration; dialogue between national and supranational courts, such as the Court of Justice of the European Union or the European Court of Human Rights.
- The rapid development of new economic models and emergence of new needs/rights: because of globalisation and cultures based on state-provided welfare, new needs frequently emerge before the legislature can intervene. When cases are brought to courts, judges must decide on the issues brought by plaintiffs even in the absence of a specific law. This brought the judiciary to expand its role and to develop a system of rights of jurisprudential origin [77].

---

[74] See, for example, P. GROSSI, *Un impegno per il giurista di oggi: ripensare le fonti del diritto*, in G. ALPA (ed.), *Paolo Grossi*, Roma-Bari, 2011, pp. 27 e 39.

[75] L. FERRAJOLI, *Contro la giurisprudenza creativa*, cit.

[76] On this topic, see also M. BARBERIS, Beccaria, *Bentham e il creazionismo giuridico*, in *Rivista Internazionale di Filosofia del Diritto*, 2014, p. 569 ff. e M. CAPPELLETTI, *Giudici legislatori?*, cit., p. 19 ff.

[77] On this topic, see also F. GALGANO, *La globalizzazione nello specchio del diritto*, Bologna, 2005.

5.2. JIFs in VAT: Between judge-made law and principle of legality

In the tax field many authors stressed the growing importance of case law. However, not many of them analysed the topic in the light of the above-mentioned studies carried out by experts of philosophy of law and constitutional law [78].

In the tax field, the topic of judge-made law is often treated in conjunction with that of the crisis of the law [79].

This topic is particularly important in tax matters as, according to Article 23 of the Italian Constitution, tax matters must be regulated through laws [80]. This principle, called "riserva di legge", is interpreted as meaning that Parliamentary Law (or Legislative Decrees, or Law Decrees) must provide the most relevant regulations of taxes, e.g. regulations of taxable persons, taxable events, taxable bases etc. [81]

---

[78] Some of the authors who have dealt with the creative role of judges in the tax field are G. Bizioli, *Building the EU Tax Sovereignty: Lessons from Federalism*, in *World Tax Journal*, 2022, p. 407 ff.; L. Carpentieri, *Riserva di legge e consenso al tributo*, in *Enciclopedia Treccani online*, 2015; C. Sacchetto, *Le fonti del diritto internazionale tributario e dell'ordinamento fiscale europeo*, in Id., *Principi di diritto tributario europeo e internazionale*, Torino, 2016, p. 21.

[79] On the deterioration of the quality of tax legislation, see, among many others, F. Bosello, *La fiscalità tra crisi del sistema e crisi del diritto*, in *Rivista di Diritto Tributario*, 1998, I, p. 1073 ff.; E. De Mita, *La legalità tributaria*, Milano, 1993; L. Ferlazzo Natoli, *Crisi del diritto tributario e funzione della giurisprudenza come scienza pratica*, in G. Ragucci, F.A. Albertini (eds.) *Costituzione, legge, tributi. Scritti in onore di Gianfranco Gaffuri*, Milano, 2018, p. 181 ff.; F. Paparella, *L'autonomia del diritto tributario ed i rapporti con gli altri settori dell'ordinamento tra ponderazione dei valori, crisi del diritto e tendenze alla semplificazione dei saperi giuridici*, in *Rivista di Diritto Tributario*, 2019, I, p. 587 ff.

[80] L. Carlassare, *Legalità (principio di)*, in *Enciclopedia Giuridica*, Roma, 1988; A. Fantozzi, *Riserva di legge e nuovo riparto della potestà impositiva in materia tributaria*, in *Rivista di Diritto Tributario*, 2005, p. 3 ff.; F. Gallo, *Ordinamento comunitario e principi costituzionali tributari*, in *Rassegna Tributaria*, 2006, p. 407 ff.; M.V. Serranò, *La riserva di legge tributaria ed il consenso al tributo*, Torino, 2008, p. 127 ff.

[81] E. Allorio, *La portata dell'articolo 23 della Costituzione e l'incostituzionalità della legge sui contributi turistici*, in *Diritto e Pratica Tributaria*, 1957, p. 86 ff.; S. Cipollina, *La riserva di legge in materia fiscale nella giurisprudenza costituzionale*, in L. Perrone, A. Berliri (eds.), *Diritto tributario e Corte costituzionale*, Napoli, 2006, p. 163 ff.; A. Fedele, *Commento all'art. 23 Cost.*, in G. Branca (ed.), *Commentario alla*

Nonetheless, specific aspects can be regulated by secondary sources. In fact, this principle is considered to be strictly related to the democratic legitimacy of taxation.

However, in recent years, researchers underlined that this principle is frequently derogated in practice. The reasons are: the emergency nature of decisions concerning taxation (they are frequently found within Budget Laws which must be adopted at the end of each year within three months); an excessive and improper use of Law Decrees and general administrative acts; and European legal integration and the growing importance of European rules within tax law [82].

The existence of European secondary law is not considered by the Italian Constitutional Court to be a violation of Art. 23. When Italy signed the Treaty on the Functioning of the European Union (TFUE) it attributed competence on indirect taxation to the EU. Nonetheless, Italian tax researchers frequently stress that European case law can be used as a source of law in harmonised matters, such as VAT (see Ch. IV).

In Italy [83], as in Germany [84], Denmark [85], and Belgium [86], EU law is limited by constitutional law [87]. Italy was one of the first countries in which the constitutional court created boundaries to European integration. This happened in the 70s. According to the Italian Constitutional Court, EU law cannot be implemented or applied when it violates the fundamental

---

Costituzione, Bologna-Roma, 1978; ID., *La riserva di legge*, in A. AMATUCCI (ed.), *Trattato di diritto tributario*, Padova, 1994, p. 159 ff.; M. LONGO, *Saggio critico sulle finalità e sull'oggetto dell'articolo 23 della Costituzione*, Torino 1968; G. MARONGIU, *L'applicazione dei principi costituzionali in materia tributaria*, in *Diritto e Pratica Tributaria*, 1962, p. 251 ff.

[82] A. MONDINI, *Corso di diritto della finanza pubblica*, Padova, 2021, p. 189 ff.

[83] Constitutional Court, Judgment of 8 November 1973, No. 183.

[84] Entscheidungen des Bundesverfassungsgericht 37, 217, 279, Solange I.

[85] Højesteret del 6 aprile 1998, I 361/1997.

[86] Decision of 16 October 1991, *Journal des Tribunaux*, 6670.

[87] On this issue, the positions taken by the Member States differ widely: in some countries, EU law enjoys full supremacy over constitutional law; in others, it is limited by constitutional law; and in others, constitutional law has supremacy over European law (C. GRABENWARTER, *National Constitutional Law Relating to the European Union*, in A. VON BOGDANDY, J. BAST (eds.), *Principles of European Constitutional Law*, Oxford, 2003, p. 84 ff.).

principles of the national legal system. This statement of the Italian Constitutional Court is called "teoria dei controlimiti" [88]. In the field of taxation, the "teoria dei controlimiti" could be invoked in relation to the principles of "riserva di legge" (Art. 23) and of ability to pay (Art. 53). These are the only principles concerning taxation which are explicitly mentioned within the Italian Constitution. For example, the Italian Constitutional Court could in principle claim that a European regulation in the field of taxation cannot be implemented because it does not comply with the ability to pay principle.

Nonetheless, the opposition to the EU that had led the national constitutional court to place external limits on European integration in the 70s has weakened with the progress of juridical integration. The doctrine of "controlimiti" has never been applied in the field of taxation and EU law has a leading role in this field. Moreover, as already said, national courts, including the Court of Cassation, frequently refer to the case law of the CJEU. The Court of Cassation has made many references to the CJEU case law since 2012. Hence, notwithstanding the increasingly important role of case law to carry forward the integration of VAT harmonisation it seems difficult that "teoria dei controlimiti" will be invoked.

6. Conclusions

The extraction of formulas is significant due to the increasing prominence of the Court of Cassation and the growing authority of its *interpretative principles of law*. This trend parallels the expanding relevance of *living law*, which compels the legislature, lower courts, and legal scholarship to engage with the pronouncements of the Supreme Court.

---

[88] A. Celotto, *L'efficacia delle fonti comunitarie nell'ordinamento italiano. Normativa, giurisprudenza, prassi*, Torino, 2003; F. Donati, *Diritto comunitario e sindacato di costituzionalità*, Milano, 1995; A. La Pergola, *Costituzione ed integrazione europea: il contributo della giurisprudenza costituzionale*, in *Studi in onore di Leopoldo Elia*, Milano, 1999, p. 815 ff.; A. Ruggeri, *«Tradizioni costituzionali comuni» e «controlimiti»*, now in *«Itinerari» di una ricerca sul sistema delle fonti*, Torino, 2003; F. Sorrentino, *Profili costituzionali dell'integrazione comunitaria*, Torino, 1996.

However, the CJEU approach relying on formulas has been criticized as it risks transforming formulas into prefabricated "building blocks" of judicial reasoning [89]. Moreover, this approach may serve rhetorical convenience rather than legal necessity, with the potential to blur the boundaries between judicial interpretation and legislative creation [90].

Comparable observations, *mutatis mutandis*, have been directed towards the *interpretative principles of law.* As noted, in reading certain decisions of the Court of Cassation that rely on lengthy lists of conforming precedents, the underlying rationale of the decision may be obscured. The reasoning is replaced by reference to prior holdings, resulting in a self-referential product devoid of substantive explanation, presented as a consumable item for the lower courts [91]. Hence, one of the challenges for legal analytics and tax law is to provide tools which could be useful to reverse or limit this trend. One of the future goals regarding JIFs could be to invest in their contextualisation with respect to the facts of the case or to improve network analysis to discourage inappropriate copy-pasting operations.

[89] M. JACOB, *Precedents and Case-Based Reasoning in the European Court of Justice*, Cambridge University Press, 2014, p. 100–101.

[90] U. ŠADL, I. PANAGIS, *The force of EU case law: An empirical study of precedential constraint*, *iCourts Working Paper Series*, 2016, p. 8.

[91] B. RIZZARDI, *Il giudice del merito e la Corte di cassazione: alla ricerca della nomofilachia perduta*, in *Questione Giustizia*, 2017, pp. 19-21.

CHAPTER VI

JUDICIAL INTERPRETATION AND LEGAL CERTAINTY: THE ROLE OF THE SUPREME ADMINISTRATIVE COURT IN BULGARIAN ADMINISTRATIVE JUSTICE

Dilyana Bozhanova – Lilia Kachoreva*

1. Introduction

The correct application of the law in a democratic society depends to a significant extent on the uniform interpretation of legal norms by the highest judicial authorities. This article examines the central role of the Supreme Administrative Court (SAC) of the Republic of Bulgaria in ensuring the coherence

* Dilyana Bozhanova and Lilia Kachoreva are in-house Lawyers at Apis Europe.

and predictability of administrative justice. First, it analyses the structure and functioning of the Bulgarian tax judiciary, highlighting the procedural safeguards and hierarchical mechanisms within which judicial interpretation takes place. Second, it links the concept of judicial interpretative formulas (JIFs) with examples from the SAC's practice in the field of value-added tax (VAT) law, illustrating how such formulas serve as instruments of legal reasoning. Particular attention is paid to interpretative judgments and the emerging constant case law that shape legal certainty and the uniform application of the law. Finally, the article explores the national debate on the creative role of judges and their indirect influence on the legislative process, emphasising the growing importance of judicial interpretation for the development of modern administrative law.

2. The Bulgarian tax judiciary

2.1. The sources of law in the field of tax justice

The primary body of tax legislation is contained in the Tax and Social Insurance Procedure Code (TSIPC), which regulates the assessment, collection, and appeal of taxes and social contributions. Substantive tax laws, such as the Corporate Income Tax Act, the Income Taxes on Natural Persons Act, and the Value Added Tax Act, further specify obligations and procedures. Implementing regulations issued by the Council of Ministers and the Minister of Finance provide technical detail for the application of tax laws.

2.2. The tax courts

Bulgaria does not maintain specialised tax courts in a narrow sense. Instead, tax disputes are resolved within the administrative court system, which consists of:

- Administrative Courts at the regional level – these are the first-instance courts competent to hear tax cases, including challenges to audit instrument, enforcement measures, and penalties.
- Supreme Administrative Court (SAC) – acting as the highest judicial authority in administrative justice, the

SAC serves as the cassation instance for tax cases. Its rulings ensure uniform interpretation and application of tax law across the country.

Although there are no dedicated tax judges, the administrative judiciary has developed considerable expertise in tax matters due to the high volume and complexity of cases.

2.3. The structure of the tax process

The Bulgarian tax process is characterised by a two-tier system of judicial control, preceded by a mandatory administrative review. This structure ensures both efficiency in dispute resolution and safeguards for taxpayers' rights.

2.3.1. Administrative appeal

Within the remit of their competence, the revenue authorities exercise tax and social-insurance control by means of conduct of audits and examinations. Audit proceedings are initiated and should be conducted in accordance with all relevant codes and regulations and within a certain period and should end with the issuance of an audit instrument, which can ascertain, modify and/or offset obligations for taxes and for compulsory social-insurance contributions; refund refundable results for a particular tax period, where so provided for by a law; refund amounts which have been unduly paid or collected.

The legality and appropriateness of administrative acts can be challenged within an administrative appeal procedure which regulatory framework is contained in Art. 152-155 TSIPC. The appeal should be brought before the immediately superior administrative body. Administrative acts can also be challenged within a judicial procedure, and the general rule is that the appellant can choose whether to appeal the act first before the administrative body and then by judicial procedure, or to submit the appeal directly to the court, i.e. the administrative appeal is optional. However, this rule does not apply to acts issued pursuant to the Tax and Social Insurance Procedure Code (TSIPC). The law requires that audit instruments and other types of acts of the revenue authority

must be challenged first before the administrative body as a prerequisite for their subsequent appeal before the court.

The audit instrument can be appealed before the administrative body, in whole or in particulars parts thereof, within fourteen days after service. The proceedings have a revising nature, since the matter is re-examined in substance by the institution that issued the act, which can itself revoke it. The decision-making authority should consider the appeal on the merits and issue a reasoned decision that may uphold, modify or revoke the audit instrument in whole or in part in respect to the appealed part thereof.

2.3.2. Judicial appeal of audit instruments before the Administrative Court

Challenging administrative acts before the relevant public body is not always as effective as a judicial appeal. This is so, because the ex officio principle has been strengthened in the court proceedings, which is why the court is not limited to only discussing the grounds stated by the appellant, but is obliged ex officio, based on the evidence presented by the parties, to check the legality of the contested audit instrument of all grounds.

Unlike the administrative review, in the judicial one the audit instrument can be appealed only regarding its legality, but not regarding its appropriateness. In addition, the audit instrument can be appealed to the administrative court only in the part in which it was confirmed, i.e. not revoked by the decision-making body under the administrative appeal procedure. The case should be heard by the administrative court in whose judicial district is the permanent address or registered place of business of the applicant at the time of carrying out the first action on the implementation of the tax-insurance control by the revenue bodies. A single-judge administrative court considers the appeal with the participation of the parties, with the possibility for a prosecutor to join the proceeding when deemed necessary, in order to protect a State or public interest. The court should adjudicate in the case on the merits, being competent to revoke the audit instrument in whole or in part, to modify the said act in the appealed part, or to reject the appeal.

According to Article 160(7) of the TSIPS, the judgement of the administrative court can be subject to cassation appellate review according to the procedure established by the Administrative Procedure Code (APC). The provision does, however, also introduce an exclusion – the judgement of the administrative court is final in cases of appeal established by audit instrument public claims a total of up to BGN 750, which does not include the accrued interest on arrears, when the audit instrument is issued to individuals, and a total of up to BGN 4000 shall not include accrued interest on arrears, when the audit instrument is issued to legal entities.

2.3.3. Cassation appeal before the Supreme Administrative Court (SAC)

The cassation proceedings before the Supreme Administrative Court are regulated in Ch. Twelve of the APC. According to the provisions of this chapter the first instance judgment can be subject to cassation contestation in whole or in separate parts thereof within fourteen days of its drafting by the relevant administrative court. The case is examined by a three-judge panel of the SAC. The cassation instance limits itself to considering the defects of the judgment as indicated in the appeal or protest, but the court should furthermore see to the validity, admissibility and conformity of the judgment to the substantive law ex officio, i.e. and on grounds not stated in the appeal or protest. The Supreme Court of Cassation assesses the application of the substantive law on the basis of the facts established by the court of first instance in the contested judgment. It is important to note that new written evidence is allowed before the Supreme Administrative Court but only for establishment of the grounds for cassation. No evidence is admissible for establishment of any circumstances irrelevant to the grounds for cassation.

According to Art. 221(2) of the APC the Supreme Administrative Court can leave in effect the judgment or reverse the judgment in the contested part thereof if the said judgment is incorrect. When the SAC leaves in effect the judgment, it provides reasons thereof and may also refer to the

reasons given by the court of first instance. According to Art. 223 of the APC the cassation judgment is final.

When reversing the judgment, the Supreme Administrative Court adjudicates in the case on the merits. The case could be referred for re-examination by another panel of the court of first instance where:

- the Supreme Administrative Court finds a material breach of the rules of court procedure;
- facts must be established, for which collection of written evidence is not sufficient.

In these cases, the orders and instructions of the Supreme Administrative Court on the interpretation and application of the law are binding upon a further examination of the case. Failure to comply with the instructions of the Supreme Court given under Article 224 of the Administrative Procedure Code is a material breach of the rules of court procedure [1]. The court of first instance examines the case according to the standard procedure, with the proceeding commencing from the first legally non-conforming procedural action which served as grounds for the referral of the case back to the said court. Where the judgment of the court of first instance is reversed again, the Supreme Administrative Court does not remand the case for a new review but adjudicates on the substance of the matter.

3. Judicial Interpretative Formulas and the case law of the Supreme Administrative Court

3.1 Definition and typology of Judicial Interpretative Formulas (JIFs)

Initially, within the framework of the POLINE project, the term *Judicial Principles of Law* (JPOL) was used, but it was later replaced by *Judicial Interpretative Formulas* (JIF) to avoid conflating the notion of *principles* with interpretative reasoning. This change was mostly terminological, but the basic concept remained intact.

[1] See Judgment No. 1204 of 4.02.2015 of the Supreme Administrative Court in administrative case No. 3855/2014, VIII o., Judge Aglika Adamova, Rapporteur.

For the purposes of this article, JIFs are understood in line with the concept developed under the POLINE project: as part of the reasoning of a court decision that involves the interpretation of a legal norm *or* principle. A necessary addition to this definition is that the interpreting body must be either the supreme jurisdiction of the respective Member State or the Court of Justice of the European Union (CJEU). This ensures that the interpretation acquires a universal dimension, capable of application beyond the particular dispute as a general formula for analogous cases.

By the scope of the applicable legal framework, JIFs can be classified as EU JIFs and National JIFs. Furthermore, based on whether they cite previous case law, JIFs may be divided into Cited JIFs (those referring to prior decisions) and New JIFs (those not relying on earlier case law).

3.2 Judicial interpretation and its limitations in the field of tax law

The concept of JIFs is not expressly recognised in Bulgarian legal theory, which generally employs the broader term *judicial interpretation*.

The notion of judicial interpretation encompasses the extraction of legal principles from case law and the methods of interpreting statutory provisions. Bulgarian legal doctrine traditionally distinguishes between official and unofficial interpretation. The official interpretation, in turn, may be authentic, legal, or judicial. The Supreme Court of Cassation itself employs the term *official judicial interpretation* when referring to interpretative judgments as a source of binding law.

The interpretative activity of courts is subject to legislative constraints. Article 46 of the Statutory Instruments Act lays down interpretative rules requiring that statutory provisions be applied according to their exact meaning. Where ambiguity exists, interpretation must align with the objectives of the law, the broader statutory framework, and the general principles of Bulgarian law.

Moreover, the Bulgarian Constitution explicitly provides that taxes and tax reliefs may be introduced only by law (Article 60(2)). As a result, courts are prohibited from interpreting tax laws expansively or creating new tax obligations through interpretation.

Under Article 46(3) of the Statutory Instruments Act, the application of analogy to justify administrative liability in tax matters is strictly forbidden. This principle has been confirmed in academic literature, emphasizing that tax procedural rules cannot be interpreted by analogy with civil law provisions but must be construed in accordance with the principles and objectives of the Tax and Social Insurance Procedure Code (TSIPC) itself.

Prohibition of placing the appellant in a worse position. The prohibition of "reformatio in peius" means that the Court may not place the auditee in a worse position for having exercised its right to appeal against the audit instrument issued by the administrative authority. The normative expression of this principle is the provision of Article 160(6) of the TSIPC: "The judgement may not modify the instrument to the detriment of the appellant". Thus, another limit is drawn to the discretion of the judge to rule on the specific tax case.

Taken together, these limits ensure that judicial interpretation in tax law remains strictly confined within the boundaries of legislative authority, preserving the separation of powers and legal certainty.

3.3 JIFs in the case law of the Supreme Administrative Court on VAT matters

In view of their binding force, we will consider 3 groups of decisions of the Supreme Administrative Court:

1. incidental practice (non-binding, *inter partes* effect);
2. interpretative judgments (binding *erga omnes*);
3. established (constant) case law (non-binding but persuasive authority).

3.3.1 Incidental judicial practice

Judgments delivered by the SAC as the last instance in a specific dispute bind only the parties to that case (*inter partes* effect) and lack general binding force. Unlike common law systems, such judgments do not create precedent obligatory for future cases. It is considered that the interpretation applied in these cases is casuistic and therefore cannot be automatically applied to similar cases.

3.3.2 Interpretative judgments

The Supreme Administrative Court exercises supreme judicial supervision as to the accurate and uniform application of the laws in administrative justice. Where the law is interpreted and applied in conflicting or incorrect case law, an interpretative judgement shall be adopted by the General Assembly of the Chambers of the Supreme Administrative Court (Article 124(1)(4) and (5) of the Judicial System Act). Where there is conflicting or incorrect case law between the Supreme Court of Cassation and the Supreme Administrative Court, the General Assembly of judges of the respective colleges of the two courts shall adopt a joint interpretative decree (Article 124(2) of the Judicial System Act). Pursuant to Article 130(2) of the Judicial System Act, interpretative judgments and interpretative decrees shall be binding on the judicial and executive authorities, on the local self-government bodies, as well as on all bodies issuing administrative acts.

The need to adopt an interpretative judgment arises when a normative act or individual provisions thereof are unclear or incomplete and are therefore interpreted differently by the administrative courts. The Chairperson of the Supreme Court of Cassation, the Chairperson of the Supreme Administrative Court, the Prosecutor General, the Minister of Justice, the Ombudsman or the Chairperson of the Supreme Bar Council, the Council of Ministers, a minister or a collective body of power established by law, when they have the authority to apply the relevant act may request the adoption of an interpretative judgement. What is specific to this type of acts of the SAC is that they are not ruled on a specific case but are formulated in the abstract in order to give a uniform interpretation of the legal provisions.

The Supreme Administrative Court has established a special Unit for Analysis and Interpretation. One of the main objectives of the Unit is to examine the signals received about conflicting or incorrect case law. Such signals may also be submitted by citizens and their organisations, lawyers, judges from different levels of the judiciary and other legal professionals.

Particularly in the field of interpretation of the legal provisions in the field of VAT taxation, a number of questions have arisen in the practice of the administrative courts in recent years, which have led to the adoption of several

interpretative judgments of the SAC: Interpretative Judgment No. 3 of 6.06.2008 in case No. 2/2008; Interpretative Judgment No. 1 of 11.02.2010 in case No. 3/2009; Interpretative Judgment No. 6 of 15.04.2021 in case No. 6/2019; Interpretative Judgment No. 4 of 10.05.2022 in case No. 2/2020.

The last of this series of judgments, Interpretative Judgment No. 4 of 10.05.2022, can be given as an example of the SAC's creative interpretation of provisions of the national VAT Act. The question which has been raised concerns the scope of joint and several liability under Article 177 of the VAT Act and, in particular, whether that liability also includes default interest. In its reasoning, the SAC applies a systematic interpretation of the provisions of the VAT Act in conjunction with the provisions of the Tax and Social Insurance Procedure Code. The SAC finds support for its opinion by citing its previous interpretative judgments. It also cites provisions of Directive 2006/112/EC on the common system of value added tax and refer to case law of the Court of Justice of the European Union on request for a preliminary ruling, in particular, Judgment of the Court (First Chamber) of 20 May 2021, „ALTI" LTD v Director of the "Appeals and Tax Social Insurance Practice" Directorate, responsible for the city of Plovdiv, within the Central Administration of the National Revenue Agency, Bulgaria. Thus, finding support for its arguments in national and EU legislation and case law, the SAC concludes in the operative part of the judgment as follows: "The scope of the joint and several liability under Article 177 of the VAT Act includes the liability for default interest on the tax due by the liable person".

This judgment will have significant practical implications – as a result of the interpretation given, the persons liable under Article 177 of the VAT Act will also have to pay default interest. It can be seen from the text of the judgment that the Supreme Administrative Court aims to clarify the meaning of the provision of the tax law in an interpretative way, without "rewriting" it, but seeking support in its previous interpretation and referring to the CJEU case law.

The above interpretative formula used by the court could be given as an example of JIF in the sense that we use within the framework of the POLINE project. This interpretation is mandatory for all courts and they must in the future decide

this issue in the same way, interpreting the provision of Art. 177 VAT Act in the sense that the Supreme Administrative Court has ruled in its interpretative decision.

3.3.3 Established (constant) case law

While not formally binding, a consistent line of SAC judgments on a particular issue gradually acquires persuasive authority. Legal scholars such as Prof. Ivan Apostolov and Prof. Zhivko Stalev have emphasised that the stability of judicial practice fosters legal certainty, predictability, and equality before the law.

Repetition of identical interpretations creates expectations among both judges and litigants that similar cases will be resolved consistently. Departures from established case law risk reversal on appeal and undermine public confidence in the judiciary.

Under a comparative perspective this phenomenon resembles the *dictum* in common law: not strictly binding (binding authority applies only to holdings), yet carrying significant persuasive weight when repeatedly affirmed.

An analysis of the appeals brought before the SAC relating to the application of the VAT Act shows that the parties and their lawyers often refer to constant case law of the SAC. The first instance administrative courts themselves also rely on and cite such SAC case law. For example in Judgment No. 6058 of 4.11.2020 of the Administrative Court, Sofia City, in Administrative Case No. 3000/2020 [...] "Consequently, the real estate which the audited person has disposed of is not an element of the independent economic activity of the person within the meaning of Article 3(2) of the VAT Act [...], i.e. it may be used both for the performance of business activity and for personal needs. In this sense is also the case-law of the Supreme Administrative Court, objectified in Judgment No. 164 of 05.01.2012 in Administrative Case No. 2642/2011, VII Department; Judgment No. 5935 of 26.04.2013 in Administrative Case No. 7552/21012, I Department, etc. [...]".

When considering cassation appeals, the SAC judicial panels also often cite and refer to previous cassation judgments rendered by other SAC panels as an argument in support of their findings. For example Judgment No. 9320 of

24.06.2013 of the SAC in Administrative Case No. 177/2013, VIII Department, Rapporteur Judge Emilia Mitkova: "The extent to which the registered entity will benefit from the exception granted by the law by choosing these supplies to be taxable depends on its free will. In this sense is the Judgment No. 15412/15.12.2009 of the Supreme Administrative Court, ruled in administrative case No. 7436/2009, whose findings are shared by this Chamber. [...]".

Lower administrative courts and litigants frequently invoke SAC constant case law as authoritative guidance. Similarly, SAC panels often cite earlier SAC judgments when deciding cassation appeals. Such citations exemplify Cited JIF**s** under the POLINE classification, as they rely on a stable line of interpretative reasoning rather than isolated decisions.

More about the creative role of the judge in the field of tax law and the current national debate regarding the admission of judge-made law will be said in the next section of the article.

4. The national debate on the creative role of judges

4.1. Making law through judicial practice

The question of whether judges merely apply the law or actively participate in its creation has been central to Bulgarian legal discourse. The "creative role" of judges refers to the interpretative and law-developing function of judicial decision-making, particularly the capacity to interpret and apply existing laws in new and innovative ways, especially when faced with complex issues not covered by precedent or legislation. While Bulgaria adheres to the continental legal tradition, where legislation is the primary source of law, judicial practice has progressively gained influence, particularly in areas where statutory provisions are vague, incomplete, or in conflict.

As a starting point to the thesis that judicial practice takes center stage in the evolution of the national legal system it is imperative to emphasise the role of the Constitutional Court of the Republic of Bulgaria. The Constitutional Court is not part of the system of courts included in the chapter of the Constitution on the judiciary, but it is regulated in a separate chapter. Its decisions, however, have binding effect according to Article 14(6) of the Constitutional Court Act: 'The

decisions of the Court shall be binding on all state bodies, legal entities and citizens'. Among the acts adopted by the Constitutional Court, it is important to highlight:

- Decisions providing a binding interpretation of the Constitution (Article 149(1) of the Constitution);
- Decisions establishing the unconstitutionality of laws and other acts adopted by the National Assembly (Article 149(2) of the Constitution).

The Constitutional Court acts on the initiative of at least one fifth of the members of the National Assembly, the President, the Council of Ministers, the Supreme Court of Cassation, the Supreme Administrative Court and the Prosecutor General. In addition, the latest amendment to Article 150(2) of the Constitution, which was adopted in December 2023, also provided for the possibility for any court, at the request of a litigant or on its own initiative, to refer to the Constitutional Court a request for a declaration of inconsistency between a law applicable to a particular case and the Constitution.

The authority to conduct constitutional review gives the Court the power to annul statutes and interpret constitutional principles. By doing so, it often clarifies the scope of fundamental rights and the limits of state power. These rulings, final and binding, function as normative standards that reshape the legal system in ways comparable to legislative reform.

Another dimension of judicial creativity lies within the powers of the Supreme Court of Cassation and the Supreme Administrative Court, which are entrusted with safeguarding the uniform application of laws. According to Article 124 of the Constitution 'The Supreme Court of Cassation shall exercise supreme judicial supervision as to the accurate and uniform application of the laws by all courts'. Similarly, Article 125(1) assigns the same function to the Supreme Administrative Court for administrative justice.

A primary instrument for achieving this objective is the issuance of interpretative judgements. These judgements represent a unique legal tool that differs significantly from ordinary judicial acts. Unlike case-specific decisions, interpretative rulings address abstract legal questions of principle, and their binding effect extends beyond the particular dispute that triggered their adoption. Interpretative judgements are not normative acts in the classical sense, as

they do not create new legal norms but clarify the meaning of existing legislation. It is guaranteed by the law that they are obligatory to all courts, bodies and institutions that apply the relevant legislation.

One of the frequently raised questions is whether the existence of this legal institution contradicts the law-making function of the National Assembly, which is proclaimed in the Constitution. In one of his articles "On interpretative judgements"[2] Dr. Vladislav Datsov examines the constitutionality of interpretative judgements issued by the supreme courts. These acts are not directly regulated by the Constitution but by the Judiciary System Act (JSA). According to the scholar their binding nature, established in Article 130(2) of the JSA, raises serious constitutional concerns. First, interpretative decisions effectively create new legal norms with the force of law, despite being issued by judicial bodies. This contradicts the principle of separation of powers (Article 8 of the Constitution), since the judiciary assumes a legislative function not explicitly granted by the Constitution. Unlike laws adopted by the National Assembly, interpretative judgements are not subject to constitutional review by the Constitutional Court, creating risks of unconstitutional yet binding rules. Second, these acts lack democratic legitimacy: unlike parliamentarians, judges are not directly elected by the people. Moreover, by making interpretative acts binding, the independence of judges (Article 117(2) of the Constitution) is restricted beyond what the Constitution allows, since judges become subject not only to laws but also to judicial interpretations that are not laws. As an alternative, the author proposes interpretative laws adopted by the National Assembly. Authentic interpretation by the legislature best ensures clarity, democratic legitimacy, respect for separation of powers, preservation of judicial independence, and the availability of constitutional review. Unlike judicial interpretative acts, interpretative laws derive directly from the law-making body and thus reflect the true legislative will without undermining constitutional principles.

Nevertheless, case law and legal doctrine accept that

[2] Дацов, В., За тълкувателните решения.

interpretative judgements occupy an intermediate position between judicial precedent and normative acts. They are not formally equivalent to law, but are treated as an informal (quasi-normative) source of law because of their binding nature and their role in legal certainty. In any case, it must be admitted that they have practical significant effect on the development of law not only for the judges as mandatory instructions when resolving specific cases but also for the legislator – as a signal for the existence of gaps, contradictions, or ambiguities in the regulatory framework.

Additionally, it is also worth mentioning the exclusive jurisdiction of the Supreme Administrative Court to hear appeals against secondary regulations issued by ministers or other authorised bodies. Pursuant to Articles 193-195 of the CAP, these judgments have effect for all and are promulgated in the Official Gazette. The Court may declare the contested act or part of it null and void, repeal it in whole or in part or reject the challenge. The revocation takes effect from the day of entry into force of the judgment. The legal consequences which have arisen from any statutory instrument of secondary legislation which has been declared void or which has been revoked as voidable shall be settled *ex officio* by the competent authority within a period that may not exceed three months after the entry into effect of the judgment of the Court.

In legal practice, certain difficulties have been observed with regard to the consequences of this type of annulment, including with regard to the recognition by the civil courts. Interesting considerations are set out in this regard in an article by Judge V. Petrov - "(Non)Recognition of the Judgments of the Supreme Administrative Court by the Supreme Court of Cassation and the Civil Courts" [3].

The academic debate on the role of judicial practice in Bulgaria has been profoundly influenced by the seminal article "Judicial practice as a source of law" by Prof. Zhivko Stalev [4]. In this paper prof. Stalev puts forward a new opinion that the long-established case law, which is formed in several court

[3] ПЕТРОВ, В., (Не)зачитане на решенията на ВАС от ВКС и гражданските съдилища.

[4] The article was first published in the Bulgarian legal journal "Contemporary Law", 1997, No. 6.

decisions containing the same interpretation of the same provision, should be considered as a source of law. The proposed thesis is that the statutory provision, together with its interpretation, constitutes a single indivisible normative unity – "law" in a broader and more general sense. Prof. Stalev sees the normative function of the Court in its power to fill gaps in the law. He supports his view with a comparative law argument from Article 1(2) of the Swiss Civil Code, which states that:

"In the absence of a provision, the court shall decide in accordance with customary law and, in the absence of customary law, in accordance with the rule that it would make as legislator".

The author's final conclusion is that the settled interpretation of a normative act motivates both the addressees of the act and the court to comply with it as everyone complies with the normative act itself and concludes that case law is not only a source of law in itself, but is also a factor in the effectiveness of other sources of law, among which normative acts occupy a central place. Prof. Stalev's opinion is not widely accepted, but it is gaining more and more supporters, including practicing judges [5].

4.2. Legislative initiative and judges in Bulgaria

The balance between the legislative, executive, and judicial branches of power lies at the heart of every democratic constitutional order. In Bulgaria, the principle of separation of powers is firmly established by the Constitution of 1991, which assigns specific competences to each branch. The legislative initiative, defined as the formal right to propose laws before the National Assembly, is a central prerogative of the legislative process. The right of legislative initiative is defined in Article 87 of the Constitution. It belongs to members of the National Assembly and the Council of

[5] K. KUNCHEV, *The article by prof. Zhivko Stalev "Judicial Practice as a Source of Law" and its Role in the Evolution of Legal Doctrine in Bulgaria* (in Bulgarian - Кунчев, К. Статията на проф. Ж. Сталев „Съдебната практика като източник на правото“ и нейната роля за развитието на правната доктрина у нас).

Ministers. This means that only parliamentarians and the executive branch may formally propose draft laws for consideration by the legislature. Judges, as members of the judicial branch, are not granted this right. The rationale is rooted in the separation of powers: the judiciary is tasked with applying and interpreting the law, not with creating it.

In 2020 the Bulgarian government introduced a widely criticised proposal for constitutional amendments, aiming to restructure elements of the judiciary and to redefine its relationship with other branches of power. One controversial aspect of the debate concerned the judiciary's role in the legislative process. While the proposal did not grant judges a formal legislative initiative, it sought to institutionalise their consultative participation in certain areas of law-making. Proponents of expanding judicial competences argue that judges are uniquely positioned to identify practical shortcomings in the legal framework. Their day-to-day experience with the application of laws allows them to propose reforms that respond to real needs. Granting judges a limited form of legislative initiative could improve legislative quality, reduce inconsistencies, and strengthen legal certainty. Conversely, critics argued that this reform was superficial, creating the appearance of judicial involvement without addressing structural challenges such as judicial independence, accountability, and the balance of power within the Supreme Judicial Council. Academic voices stressed that meaningful reform should focus on protecting the autonomy of judges rather than drawing them into political processes that risk compromising their neutrality.

Although judges lack formal legislative initiative, their professional role in applying the law provides them with substantial indirect influence over the legislative process. One of the most important channels for such involvement is the Supreme Judicial Council (SJC). Article 130 of the Bulgarian Constitution establishes the SJC as the central administrative and self-governing body of the judiciary. Its core functions include judicial appointments, promotions, disciplinary actions, and ensuring the independence of the judiciary. However, the SJC also plays an advisory role in the legislative process. The Council may issue opinions on draft laws and propose amendments concerning the organisation and functioning of the judiciary. Although these opinions do not constitute formal

legislative initiatives, they often serve as a significant source of expertise for those who do hold the constitutional right of legislative initiative. In practice, Members of Parliament and the Council of Ministers frequently adopt or adapt SJC recommendations into draft laws.

Another aspect of the indirect role of the judges in the legislative process are the inconsistencies, ambiguities, or lacunae in the legal system that judicial practice frequently uncovers and which legislators are then compelled to address. The interpretative judgements of the Supreme Court of Cassation and the Supreme Administrative Court, as well as the rulings of the Constitutional Court, often stimulate legal reforms. In this way, the judiciary helps refine the quality and coherence of legislation without overstepping the boundaries of its constitutional mandate.

Furthermore, professional organisations such as the Bulgarian Judges Association and the Association of Bulgarian Administrative Judges have an increasingly important role in shaping legislative initiatives, particularly in areas related to the functioning of the judiciary and the protection of fundamental rights. These associations serve as expert communities that consolidate the knowledge, experience, and perspectives of practicing judges. Their main contribution lies in offering professional expertise grounded in case law, procedural experience, and the principles of judicial independence. This makes their input valuable in the drafting and evaluation of legislative proposals. One of the most significant ways in which these organisations participate in the legislative process is through the development of opinions, statements, and policy concepts. When the National Assembly or the Council of Ministers introduces legislative proposals concerning judicial reform, criminal procedure, or civil justice, professional judicial associations often provide reasoned feedback. These written opinions identify potential risks, gaps, or inconsistencies in proposed texts. Such contributions are not merely academic. Members of Parliament or government officials can adopt or adapt these recommendations when preparing draft laws. For example, proposals by the Bulgarian Judges Association have, in several instances, informed discussions on reforms aimed at improving transparency, accountability, and the efficiency of the judiciary. This form of dialogue enhances the legitimacy of the legislative process by

ensuring that laws affecting the courts are shaped with the participation of those who apply them in practice.

5. Conclusion

The analysis confirms that the interpretative activity of the Supreme Administrative Court has become a key factor in shaping Bulgarian administrative justice. Through interpretative judgments, consistent case law, and abstract legal reasoning, the SAC ensures the uniform application of legislation, resolves conflicts in judicial practice, and indirectly stimulates legislative improvements. This growing body of case law not only guarantees legal certainty and equality before the law but also strengthens the rule of law by providing interpretative standards for all administrative and judicial authorities. In an era of complex legal relations and rapidly evolving legislation, the interpretative role of the SAC emerges as a cornerstone of both effective judicial protection and the coherent development of the legal system as a whole.

CHAPTER VII

EU DIALOUGE AND VAT PRECEDENT IN SWEDEN: OVERRULING AND DISTINGUISHING IN THE SUPREME ADMINISTRATIVE COURT

Magnus Kristoffersson – Kevin Aiderfors – Eleonor Kristoffersson*

1. Introduction

This chapter investigates the doctrinal mechanics through which the Swedish Supreme Administrative Court (sw: *Högsta förvaltningsdomstolen*, HFD) interprets Value Added Tax (VAT) law, with particular attention to the techniques of overruling and distinguishing. The analysis is grounded in an empirical corpus of approximately 670 VAT-related outcomes (sw: *referat* and *notiser*) and is conducted on two registers: theoretically, it

* Eleonor Kristoffersson is Full Professor of Tax Law at the Örebro and University of Gävle; Magnus Kristoffersson is Associate (Docent) Professor of Tax Law at the Örebro University and University of Gävle; Kevin Aiderfors is Student Assistant at the School of Behavioural, Social and Legal Sciences of the Örebro University. The authors have contributed equally to this work. Beside the EU funding, the working hours of Eleonor Kristoffersson were partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

interrogates how doctrinal change is signalled and constrained in a system without formal *stare decisis*; practically, it examines how those signals orient the day-to-day application of law by lower courts and administrative authorities.

Within the Swedish public law architecture, HFD's judgments, while formally binding only inter partes, exert considerable persuasive force and are ordinarily followed. The leave-to-appeal filter focuses the docket on questions of general importance for legal guidance and appeals from the Board for Advance Tax Rulings (*sw: Skatterättsnämnden*, BAR) reach HFD without leave, ensuring that issues with systemic significance are addressed at the highest level. EU law provides the principal methodological frame: because the Board is not a "court or tribunal" for the purposes of Article 267 TFEU [1], VAT interpretative questions reach the CJEU [2] through the ordinary court hierarchy, and where no preliminary reference is made HFD supplies the national last word. In this setting, even modest adjustments in the HFD's reasoning style or publication choices can produce meaningful doctrinal movement.

Methodologically, the HFD seldom declares overtly that earlier authority is overruled or merely distinguished. Instead, recalibration tends to be expressed through craft: selective recourse to plenary disposition; a calibrated publication practice that pairs *referat* with same-day *notiser* to propagate clarifications across adjacent fact patterns; and a concise integration of CJEU authority that identifies controlling propositions without extended quotation. These techniques allow HFD to maintain continuity while aligning domestic doctrine with the VAT Directive's structure and general principles.

The chapter proceeds by developing these themes across the sections used in the text. Overruling and distinguishing examines how HFD signals reinterpretation or departure from earlier guidance and how form and publication strategy convey such movement. Citations of previous European and national decisions analyses the Court's mode of citation and the functional role that brief, pinpointed references play in

[1] Consolidated version of the Treaty on the Functioning of the European Union, Official Journal of the European Union, C 326, 26 October 2012, p. 1–390.

[2] Court of Justice of European Union.

structuring the dispositive analysis. Non-self-standing Judicial Interpretative Formulas (JIF) explores how general principles are woven directly into VAT reasoning without a separate doctrinal compendium, while non-VAT JIF considers the role of principles such as proportionality and neutrality as constraints and enablers in fraud and abuse settings. One JIF composed by statement to be found in different paragraphs traces instances in which a single justificatory idea is articulated diffusely yet operates as a coherent doctrinal driver. Explicit qualification and JIF or similar terms assesses the extent to which the Court relies on explicit labels as opposed to implicit cues and structural devices. Summary and conclusions synthesise the findings and reflect on their implications for predictability, institutional design, and future research.

By combining a theoretical account of doctrinal signalling with a practical analysis of case processing and publication technique, the study aims to clarify how HFD functions as a doctrinal engine in Swedish VAT law; preserving stability where appropriate while retaining the capacity to recalibrate in step with EU law and administrative realities.

2. Overruling and distinguishing

The empirical corpus underpinning this study contains approximately 670 VAT-related HFD outcomes (*referat* and *notiser*). That magnitude is consistent with the HFD's mandate to create guidance (*sw: prejudikat*) and with the scope of officially published case law searchable via the judiciary's case law service and the HFD's decisions pages, even if the exact count of VAT cases depends on classification choices between referat and notiser.

In Swedish public law, HFD reviews both the merits of a tax dispute and applies constitutional review when necessary: under Ch. 11, Section 14 of the the Instrument of Government (1974:152) (*sw: Regeringsformen (1974:152)*), all courts must disapply a norm that conflicts with the Constitution or a superior law [3]. HFD's institutional role is expressly precedent-

[3] Chapter 11, Section 14 of the The Instrument of Government (1974:152).

forming, its primary task is to lead legal application by issuing decisions that guide lower courts and agencies [4]. Thus, although its judgments bind formally only inter partes, they carry heavy persuasive authority and are generally followed.

Two channels feed VAT questions into HFD. First, ordinary tax appeals reach HFD only if leave to appeal (*sw: prövningstillstånd*) is granted [5]. The leave regime centres on whether a case is important for guidance of the law (*sw: prejudikatdispens*), with additional "extraordinary reasons" grounds. [6] This structure focuses HFD's docket on issues of general importance rather than case-specific error correction. Second, advance rulings (*sw: förhandsbesked*) are issued by BAR, an independent body [7]. A ruling may be granted on an individual's application when the issue is significant either to the applicant or for uniform interpretation or application of the law, and the application must be lodged before the relevant event occurs [8]. Appeals from BAR go directly to HFD and, distinctively, do not require leave to appeal [9]. However, it appears that there is a tendency for HFD to have introduced an informal leave-to-appeal system by rejecting an increasing number of applications deemed unsuitable for advance rulings [10]. The statute also empowers the Tax Agency's public representative (*sw: Allmänna ombudet*) to seek advance rulings in tax matters [11]. These provisions promote uniformity in tax administration while preserving individual access to anticipatory clarification.

---

[4] 36 § Tax Administrative Court Procedure Act.

[5] 36 § The Administrative Court Procedure Act (1971:291) (*sw: Förvaltningsprocesslag (1971:291)*)

[6] 36 § The Administrative Court Procedure Act (1971:291).

[7] 1 - 4 §§ Act (1998:189) on Advance Rulings in Tax Matters, The Ordinance (2007:785) with instructions for Skatterättsnämnden (*sw: Förodningen (2007:785) med instruktion för Skatterättsnämnden*) and 20 § Ordinance (2017:154) with Instructions for the Swedish Tax Agency (*sw: Förordning (2017:154) med instruktion för Skatteverket*).

[8] 5 – 6 a §§ Act (1998:189) on Advance Rulings in Tax Matters.

[9] 22 § Act (1998:189) on Advance Rulings in Tax Matters.

[10] I. MELBI, *Förhandsbesked i skattefrågor – behöver prövningstillstånd införas?*, Skattenytt 2019, pp. 286-293; P. NILSSON, C. BERKESTEN HÄGGLUND, *Enklare att få prövningstillstånd än förhandsbesked – Högsta förvaltningsdomstolens strikta avvisningspolicy motverkar förhandsbeskedens syfte*, Skattenytt, 2022, pp. 62-73.

[11] 1, 6, 6 a, 10 §§ Act (1998:189) on Advance Rulings in Tax Matters.

On the plane of EU law interaction, BAR is not a "court or tribunal" entitled to seek preliminary rulings under Article 267 TFEU; in Victoria Film the CJEU characterized BAR as performing an administrative function and thus outside the meaning of "court or tribunal" [12]. As a practical matter, questions of EU-VAT interpretation therefore surface in Luxembourg only via the ordinary court hierarchy, with HFD exercising the final national word absent a preliminary reference.

Against that institutional backdrop, HFD's written reasons seldom declare in explicit common-law terms that an earlier line is being "overruled" rather than "distinguished". Instead, technique and form signal the move. One formal indicator is found in § 5 of the Administrative Courts Procedure Act, which authorizes the Court to order that a case be decided "in its entirety" (plenary) when a legal question is of special importance or when departure from earlier guidance is contemplated [13]. Plenary sitting is thus available, but not mandatory, as a mechanism for managing doctrinal recalibration [14]. Practice also illustrates a nuanced publication strategy: the Court continuously publishes both *referat* (*full reports*) and *notiser* (*short notes*), and it has sometimes paired a lead *referat* with same day *notiser* applying the same solution to adjacent fact patterns [15]. Both the plenary power and the *referat/notis* practice function as vehicles for either clarifying or quietly redirecting doctrine without formulaic labels.

The Court's reasoning style is well illustrated by HFD 2023 ref. 24, which addressed whether voluntary VAT liability for letting immovable property (sw: *frivillig beskattning*) can be denied solely because several tenants share the same space without exclusive rights. HFD anchored its analysis in the VAT Directive's option for Member States to permit taxation of lettings and in the CJEU's insistence on tax neutrality and

---

[12] CJEU, 12 November 1998, *Victoria Film A/S*, C-134/97, ECLI:EU:C:1998:535.

[13] E. KRISTOFFERSSON, *Att använda prejudikat och annan rättspraxis i rättstillämpningen*, Svensk skattetidning, 2011, p. 839 ff.

[14] *Ibid*.

[15] *Ibid*.

uniform application. In Turn- und Sportunion Waldburg, the CJEU recognized a broad national margin of discretion when implementing the option, tempered by neutrality constraints [16]. HFD read Swedish preparatory works on "permanent use" alongside that EU framework and concluded that multiple-tenant use does not, by itself, defeat voluntary liability, thus clarifying rather than contracting the reach of the Swedish regime.

That clarification builds on earlier decisions. In HFD 2015 ref. 62, the Court accepted voluntary tax liability where a property owner intended repeated short-term lettings for taxable use, decoupling "permanent use" from single long-term tenancy [17]. In HFD 2016 ref. 51, the Court affirmed that voluntary liability attaches to a specific property or physically delimited part (such as a defined set of rooms), not to an abstract "percentage" of the same area. Together, those rulings laid the groundwork for HFD 2023 ref. 24's holding that shared occupation does not negate permanence when the tenants' uses are taxable or public.

By contrast, HFD 2023 ref. 45 illustrates a tacit but concrete shift relative to earlier practice on input-VAT allocation in mixed activities [18]. Swedish law allowed allocation on a "reasonable basis", a standard treated in practice as compatible with the Directive's system. In HFD 2023 ref. 45, however, HFD held that a taxable person cannot be denied the turnover-based method set out in Article 174 of the VAT Directive, emphasizing the Directive's logic and the direct-effect consequences when national law does not fully implement the EU method [19]. In effect, the Court minimized the precedential force of HFD 2014 ref. 18 as to the direct-effect question, and

---

[16] CJEU, 12 January 2006, *Turn- und Sportunion Waldburg*, C-246/04, ECLI:EU:C:2006:22, para 27-30.

[17] T. MUINAFSHAR, *Momsregler hindrar asylboenden*, Svensk Skattetidning, 2022, pp. 228-236.

[18] M. SENYK, *Har rätten till avdrag för ingående mervärdesskatt utvidgats i svensk praxis under det senaste decenniet? – en översikt över rättspraxis från HFD och jämförelse med utvecklingen av EUD:s praxis*, Skattenytt 2024, pp. 165-185.

[19] T. KARLSSON, R. SCHIESS, *Högsta förvaltningsdomstolen – skälig fördelningsgrund för avdrag för ingående moms – HFD 2023 ref. 45*, Skattenytt 2024 p. 78.

without convening plenary it aligned domestic doctrine with EU law. This judgment operates as a "constructive overruling" in substance, although it does not say so in so many words.

Finally, the Court's use of *notiser* alongside *referat* confirms that distinguishing often proceeds incrementally. *Notiser* are used to propagate a clarified rule across factual variants, sometimes even after a plenary ruling, thereby entrenching the clarified line while preserving continuity [20]. That strategy makes Swedish overruling comparatively subtle, yet doctrinal effect is nonetheless real. In sum, HFD's VAT jurisprudence favors articulating shifts through reasoned alignment with EU law, calibrated publication strategy, and selective recourse to plenary procedure, rather than overt proclamations of overrule. The result is a system in which change and continuity co-exist: lower courts follow HFD's lead, while the Court retains flexibility to recalibrate practice.

3. Citations of previous European and national decisions

HFD routinely grounds its VAT reasoning in both European and national case law and typically does so without lengthy verbatim quotations. Instead, it cites the relevant judgment and pinpoints the controlling proposition in a compact way, integrates that proposition into its own analysis, and then states the consequence for the case at hand. The technique is visible in HFD 2021 ref. 18, where the Court first affirms the formal conditions for exemption of an intra-Union supply under national law corresponding to Article 138(1) of the VAT Directive and then immediately frames the dispute through settled CJEU doctrine on fraud and abuse [21]. In paragraph 10, the Court states that individuals may not rely on EU law in a manner constituting tax evasion or abuse and that national authorities and courts must refuse relief where, having regard to objective factors, the seller knew or ought to have known that the transaction formed part of a VAT evasion scheme, even if the evasion occurred in another Member State and the seller did not personally profit; the tax authority bears the

[20] E. KRISTOFFERSSON, *Att använda prejudikat*, cit., 2011, p. 839 ff.
[21] M. FRI, U. GREFBERG, *E Mervärdesskatt*, Skattenytt, 2022, p. 284.

burden of proof[22]. In paragraph 11, the Court adds that the prohibition of abuse is a general principle of EU law, so the duty to deny a right in cases of tax avoidance or abuse applies even absent national provisions capable of EU-conform interpretation[23]. HFD then applies those propositions to the national record: it rejects the intermediate appellate court's view that refusal of exemption presupposes a final, factual loss of VAT within the chain and holds that no such "final loss" requirement exists; instead, the decisive questions are whether evasion occurred somewhere in the chain and whether the seller knew or should have known[24]. The Court further emphasizes that proportionality and neutrality do not bar denial of rights in fraud contexts and that combating tax evasion would be significantly hindered if denial were limited to situations where double taxation can be ruled out ex ante.

Read against the background of the CJEU's line from Halifax (C-255/02), Kittel (C-439/04 and C-440/04), Mahagében and Dávid (C-80/11 and C-142/11), and Cussens (C-251/16), the HFD's method is orthodox: it imports a short, operative formulation from the CJEU, often introduced with a "see, for example" marker, then sets out the national implication. Kittel supplies the "knew or ought to have known" knowledge standard and treats the test as objective. Mahagében and Dávid clarify that mere irregularities or doubts are insufficient to deny rights and that the administration must substantiate, by objective evidence, participation in a fraud scheme; they also confirm that the tax authority carries the burden of proof. Cussens underscores that the abuse-of-rights principle is directly applicable as a general principle; Italmoda brings that approach firmly into the VAT context and extends it beyond the right to deduct to

---

[22] CJEU, 18 December 2014, *Italmoda*, C-131/13, C-163/13 and C-164/13, ECLI:EU:C:2014:2217, para. 43, 44, 45, 50, 62, and 69; CJEU, 13 February 2014, *Maks Pen*, C-18/13, ECLI:EU:C:2014:69, para. 27–29; CJEU, 17 December 2020, *Bakati Plus*, C-656/19, ECLI:EU:C:2020:1045, para. 80.

[23] CJEU, C-131/13, C-163/13 and C-164/13, *Italmoda*, cit., para. 51–59 and 62; CJEU, 26 February 2019, *N Luxembourg 1 and Others*, C-115/16, C-118/16, C-119/16 and C-299/16, ECLI:EU:C:2019:134, para. 117–119; cf. HFD 2013 ref. 12.

[24] HFD 2021 ref. 18, para. 13 and 15.

exemptions and refunds. HFD 2021 ref. 18 explicitly synthesizes these strands and, by anchoring its holding in "firm praxis (sw: *fast praxis*)" from the CJEU, treats the general principle as independently operative in Swedish law even without domestic transposition.

The Court's dialogue with national precedent follows the same pattern. In HFD 2013 ref. 12, decided before Italmoda, the Court reached essentially the same endpoint by EUconform interpretation of the VAT Act, holding that the buyer's right to deduct can be denied where objective circumstances show knowledge or constructive knowledge of fraud in the chain. HFD 2021 ref. 18 acknowledges that earlier Swedish approach but reframes the legal basis: after Italmoda, denial of rights in fraud or abuse settings follows directly from general principles of EU law, not solely from conform interpretation [25]. The evolution is doctrinally significant. It both standardizes the Swedish analysis with EU law and clarifies that the absence of a specific Swedish statutory rule against abuse does not prevent the denial of an exemption or deduction when the Kittel/Italmoda knowledge standard is met.

Methodologically, the HFD's restrained citation style serves a functional purpose. By citing the judgment "as such" and pinpointing a few controlling paragraphs, the Court signals that the proposition is settled EU law and that its national decision is an application rather than an innovation. The technique also structures the burden-of-proof analysis: once the court identifies the applicable CJEU standard, it asks whether the tax authority has shown, through objective factors, that the taxable person knew or should have known; if so, denial of the right follows irrespective of where in the Union the evasion occurred or whether there was a proven final loss of tax at the last stage. The resulting synthesis, brief CJEU citations, concise statement of the governing rule, and tailored national application, has produced a stable line in Swedish VAT jurisprudence in which HFD's conclusions align with the CJEU's abuse-of-rights and fraud case law while

[25] M. Fri, U. Grefberg, *E Mervärdesskatt*, cit., p. 284.

maintaining the Swedish court's characteristic economy of quotation.

4. Non-self-standing Judicial Interpretative Formulas

In Swedish VAT jurisprudence, HFD often incorporates prior authority directly into its reasoning, using concise quotations or paraphrases of earlier decisions or principles. In the past, doctrine was woven into the legal narrative, rather than being treated as a separate doctrinal compendium.

A clear example is HFD 2023 ref. 48, where the court begins by restating what it deems a fundamental principle: the VAT burden should rest solely on the final consumer, and the taxable amount the authorities collect must not exceed the consideration actually paid by that consumer[26]. The Court anchors that statement in CJEU authority, Elida Gibbs (C-317/94) and Boehringer Ingelheim Pharma (C-462/16), citing the precise paragraphs in which the principle is expressed (paras 19, 28 in Elida Gibbs; para 35 in Boehringer Ingelheim).

It then connects this high-level premise to the structure of multi-stage VAT deduction: in a chain of supplies, each purchaser deducts the VAT added by its seller, leaving only the final consumer without deduction burdened with the tax. Using that frame, in paragraph 18 HFD addresses the scenario of post-sale discounts granted by an upstream supplier that "skip" intermediate steps. Quoting CJEU doctrine again, Elida Gibbs (paras 32–34) and Boehringer Ingelheim (C-717/19) (paras 40–45), the Court notes that intermediate stages need not adjust their taxable amounts in response to the discount, because the deduction mechanism insulates them.

In paragraph 19 the Court clarifies the domain of those precedents: they concern transactions taxed at all stages, including the final retail sale. The reasoning draws also from Commission v Germany (C-427/98, para 64), which supports the idea that a vendor's right to reduce the taxable base depends on the retail price, and any reduction thereof, already containing VAT. Since in that chain the final consumer's

[26] M. JACOBSSON, T. KARLSSON, J. ÖBERG, *EU-domstolens domar*, Skattenytt, 2025, pp. 368-369.

payment and the price reduction both embody VAT, upstream adjustments are justified.

On the facts, however, HFD then distinguishes. Because prescription medicines at retail are exempt from VAT under Swedish law (albeit with input-tax deduction upstream), no VAT is charged to the consumer or the regional health authority in reimbursement. Accordingly, the condition of a VAT final sale is missing; thus, the upstream manufacturer's later compensation to regions cannot be deemed a reduction of a taxable amount that originally included VAT. HFD, therefore, affirms the advance ruling denying the downward adjustment of the manufacturer's tax base, on the ground that the chain's last leg lacks taxable output to which a reduction mechanism might attach.

This example shows how HFD's quoting strategy is more than ornament, it is functional. The quoted passages supply the normative frame and constraints which shape the Court's reasoning. The method underscores the necessity of close reading: one must discern not only which authorities are cited, but how the Court *integrates* or *limits* them in relation to the facts. Because HFD seldom labels its moves as "overruling" or "distinguishing", analysts must detect whether a quotation is being deployed to extend a rule or to circumscribe it, even implicitly. Understanding Swedish VAT precedent thus requires parsing not just the referents the Court cites, but the rhetorical and logical work those quotes perform in situ.

5. Non-VAT Judicial Interpretative Formulas

HFD frequently draws on general principles that are not specific to VAT when resolving VAT disputes, and proportionality is one of the principles that recur in its reasoning. Read in context, proportionality serves as a limiting but not disabling constraint: measures must not go further than necessary to achieve their aims, yet Union law recognises that combating tax evasion and abuse is itself a legitimate and weighty objective. This balance is explicit in HFD 2021 ref. 18. There the Court of Appeal had reasoned that denying a right (here, the exemption for an intra-Union supply) should only be permitted if authorities can ensure that denial will not produce double taxation, invoking neutrality and

proportionality. HFD rejected that approach. It emphasised that settled CJEU case law treats the fight against VAT fraud and abuse as an objective promoted by the VAT Directive and that the neutrality principle does not shield a person who has participated in evasion from the denial of a right [27]. Building on that premise, the Court added that conditioning denial on ex ante certainty about the absence of double taxation would, in many situations, make it considerably harder to counter abusive practices; in real supply chains across borders, such a condition would invite strategic opacity and allow fraud to persist. Therefore, proportionality cannot be read to compress the abuse principle in the way the Court of Appeal proposed [28].

HFD's conclusion aligns with the CJEU's longstanding articulation of the proportionality test and its interplay with neutrality. In Garage Molenheide (C-286/94, para. 46–47), the Court held that Member States may adopt effective measures to protect the public purse, provided those measures do not go beyond what is necessary and do not undermine the structure and aims of the VAT system; the point is a calibration, not an immunity. The antifraud strand of case law sharpens that calibration through the knowledge standard: a Member State may refuse rights where, having regard to objective factors, the taxable person knew or ought to have known that its transactions were connected with VAT fraud. [29] HFD 2021 ref. 18 applies this "knew or should have known" test directly when it states that the decisive questions are whether evasion occurred somewhere in the chain and whether the seller knew or should have known of it; there is no further requirement that the authorities prove a final, factual tax loss at the last stage of the chain. That is a proportional response because the sanction, denial of a right, targets only the actors whose knowledge links them, on objective evidence, to fraud.

The Court's discussion also sits comfortably with the CJEU decisions sometimes cited on the other side of the ledger. Teleos (C-409/04, para. 25) and Collée (C-146/05, para. 23) reflect

---

[27] CJEU, C-131/13, C-163/13 and C-164/13, *Italmoda*, cit., para. 42, 48; CJEU, C-656/19, *Bakati Plus*, cit., para. 80.

[28] HFD 2021 ref. 18, para. 14.

[29] CJEU, 6 July 2006, *Kittel*, C-439/04 and C-440/04, ECLI:EU:C:2006:446, para. 55–61.

legal-certainty and neutrality concerns where compliant suppliers acted in good faith or where documentary proof was produced late, and they caution against measures that punish the diligent. But those cases do not install a "final loss" prerequisite. They police proof and procedure, not the substantive antifraud standard. Italmoda subsequently made explicit that the abuse principle in VAT is a general principle of EU law that can, where appropriate, require denial of rights even in the absence of a specific national implementing rule [30]. Bakati Plus confirms that neutrality and proportionality do not prevent refusal of exemptions where the objective evidence supports evasion [31]. HFD 2021 ref. 18 absorbs these strands: it recognises proportionality, but it also recognises that proportionality does not demand a proof-of-no-doubletaxation safe harbour for those who knew or should have known of fraud.

Finally, the Court's technique is characteristic. It cites judgments "as such", names the controlling proposition with paragraph references, and integrates that proposition into the dispositive reasoning with minimal quotation. The effect is to let general principles, here proportionality and neutrality, do their work inside a structured antifraud framework. Proportionality remains a real limit: authorities retain the burden to establish, on objective factors, a connection to evasion and the requisite knowledge; mere irregularities or suspicions are insufficient. But once that burden is met, proportionality is satisfied by denying rights to the knowing participant, even if it cannot be guaranteed in advance that no element of double taxation will arise somewhere in the Union. In this way, HFD's use of non-VAT JIF is consistent with the CJEU's proportionality jurisprudence, while preserving the practical effectiveness of the abuse-of-rights doctrine that underwrites the integrity of the common VAT system.

[30] CJEU, C-131/13, C-163/13 and C-164/13, *Italmoda*, cit., para. 51–59, 62.

[31] CJEU, C-656/19, *Bakati Plus*, cit., para. 80.

6. One Judicial Interpretative Formula composed by statement to be found in different paragraphs

In HFD's more recent VAT judgments, the reasons are presented in numbered paragraphs, which facilitates precise citation of propositions and their sequencing; older decisions are often set out in a more continuous narrative without paragraph numbering. The contrast is instructive in the line of cases on composite supplies. A good illustration from the older format is the judgment reported as HFD 2012 ref. 43. The matter reached the Court as an appeal from an advance ruling and concerned a contract under which Y AB undertook, for X AB Mervärdesskattegrupp, to handle the principal's customer correspondence. The setup comprised three bundles of elements: production packaging (paper, envelopes, pallets and protective packing), production services (printing, enveloping, sorting, bundling and packing on carriers) and distribution services (transport of individually addressed postal items posted on Åland for delivery to customers in Sweden). The applicant stressed that the legal materials did not make the tax treatment clear, that on transport services VAT could potentially be levied domestically (by reference to 5 chapt. 5 § OML) [32] while importrelated VAT could be levied by Customs (7 chapt. 8 § OML) [33], and that authoritative guidance was needed to avoid double taxation. Four questions were put, focusing on whether to apply a predominant approach or a splitting approach, whether the overall transaction was a supply of goods or services, and, depending on the answer, whether any exemption connected with import (3 chapt. 32 § OML) [34] could apply.

In its reasons, the Court did not erect a freestanding discussion of "principles" by name, nor did it present block quotations from earlier authority. Instead, it applied the CJEU's compositesupply method in a compact way: identify what the average customer chiefly seeks; treat other elements as ancillary if they merely enable better enjoyment of the

[32] 5 ch. 5 § OLM (*sw: Mervärdesskattelag (1994:200)*) is now 6 chapt. 33 and 34 §§ VAT Act (*sw: Mervärdesskattelag (2023:200)*).

[33] 8 ch. 24-26 §§ VAT Act.

[34] 10 ch. 63 § VAT Act.

principal element; and avoid artificial splitting where the parts are so closely linked that they constitute, economically, a single supply. That method traces directly to the CJEU's guidance in Card Protection Plan (C-349/96), Levob (C-41/04) and later cases such as Graphic Procédé (C-88/09): every supply is normally distinct and independent, but a single economic supply should not be artificially divided; there is a single supply in particular where one element is principal and others are ancillary, or where several elements are so closely linked that, objectively, they form one whole. Applying this matrix to the Åland correspondence chain, HFD characterized the printing, enveloping and sorting as subordinate to the distribution component, because what the customer principally sought was the actual delivery of the correspondence to addresses. The Court therefore treated the whole as one supply, a transport service, rather than two separate supplies or a supply of goods with subordinate services.

That classification carried immediate consequences. First, as a transport service the place-of-supply rule in 5 chapt. 5 § OML [35] located the supply in Sweden, so that the purchaser's VAT group was liable to account for the tax domestically. Second, the importrelated exemption in 3 chapt. 32 § OML [36] did not come into play on the facts because the letter consignments did not include goods imported by a Swedish acquirer to which the transport would be merely an accessory cost forming part of the customs value. In short, the attempt to bring the transaction within the importexemption path fell away once the Court determined that the composite supply was a domestic transport service and not an importrelated mixed package.

Two features of technique are notable. The first is stylistic: although the underlying dichotomy between a predominance approach (sw: *huvudsaklighetsprincipen*) and a splitting approach (sw: *delningsprincipen*) was squarely raised by the parties and by BAR's, the Court's own reasons implement the predominance analysis without naming it expressly [37]. The operative cues are the EU law tests (principal versus ancillary;

[35] See not. 33.

[36] See not. 35.

[37] Also HFD 2012 ref. 30 and O. KARLSSON, J. KELLGREN, E.

the "single, indivisible economic supply" standard) rather than domestic labels. The second is structural: the decision's unnumbered "*Skälen för avgörandet*" shows how, in older formats, the Court would integrate the compositesupply test into a concise doctrinal paragraph culminating in a clear classification and its statutory consequences, rather than parceling the reasoning into numbered propositions as seen in more recent judgments.

Placed against the broader EU jurisprudence, the outcome is orthodox. Card Protection Plan establishes the principal/ancillary framework; Levob confirms that where multiple elements are objectively one whole, a single supply must be recognized; Graphic Procédé clarifies the boundary between a supply of goods and a supply of services where the provider reproduces documents and transfers the copies [38]. HFD's conclusion that the correspondencehandling package is a single supply whose principal element is distribution fits those authorities and avoids the distortions that arise from artificially slicing an economically unitary chain into separate taxable events. It also illustrates the Court's characteristic economy of citation: earlier judgments are referenced for their controlling propositions, integrated into the Court's own voice, and used to drive classification, place of supply and the (non-) applicability of specific exemptions to the facts as found.

7. Explicit qualification and Judicial Interpretative Formulas or similar terms

In Swedish administrative adjudication, precedent from the HFD, as mentioned above, is not formally binding in a commonlaw sense, yet it is intended to guide the application of law and is ordinarily followed by the administrative courts and by agencies such as the Tax Agency. The leave-to-appeal filter is designed to bring to HFD issues of general importance, and the court's reasons therefore carry a normative message that extends beyond the parties. At the

---

KRISTOFFERSSON, *Funktionsförsäljning och beskattning*, Skattenytt årsskrift, 2019, pp. 117 ff.

[38] CJEU, 25 February 1999, *Card Protection Plan Ltd*, C-349/96, ECLI:EU:C:1999:93.

same time, lower courts and authorities retain the possibility to distinguish earlier case law where the factual matrix or the norm invoked is meaningfully different, or where subsequent legislation or Union law calls for recalibration. In practice, persistent deviation from HFD's guidance is rare; rather, the case law is internalized into the day-to-day administration of law, often through updates to agency guidance and the vocabulary of the lower courts.

The decision reported as HFD 2023 ref. 24 exemplifies this guidance function in the field of voluntary VAT liability on the letting of immovable property. Swedish law requires "permanent use" of the premises in a taxable activity. The issue before HFD was whether that permanence requirement bars voluntary tax liability when several tenants will occupy the same premises without any of them holding an exclusive right to a delimited area. By answering that the permanence requirement does not, in itself, preclude voluntary tax liability merely because multiple tenants share the same space, the court clarified the reach of the Swedish regime and harmonized it with the practical realities of modern tenancy structures. The court then granted leave to appeal in the dormant part of the case and remitted that part to the Administrative Court of Appeal in Jönköping for continued handling in light of its reasons.

Read against the earlier line of authority, the clarification is measured rather than disruptive. In HFD 2015 ref. 62 the court had already accepted that permanence can be satisfied by a landlord's plan to let a venue repeatedly in short terms for taxable activities; in HFD 2016 ref. 51 the court emphasized that voluntary tax liability can attach to a specific property or a physically delimited part of a property, but not to an abstract percentage of one and the same area. HFD 2023 ref. 24 judgment fits that architecture: permanence concerns the character of use over time, not exclusivity of square meters, while the delimitation rule from 2016 continues to govern the object to which voluntary liability may attach. The practical effect has been immediate at the administrative level, where guidance materials now recite both the delimitation constraint and the clarification that shared occupation does not, by itself, defeat permanence.

The older judgment HFD 2012 ref. 43, reported in the narrative style that preceded HFD's routine use of numbered paragraphs, shows the same technique of shaping doctrine

through concise reasons anchored in Union law method. The contract at issue combined production packaging (paper, envelopes, pallets and protective materials), production services (printing, enveloping, sorting, bundling and packing) and distribution services (transport of individually addressed items posted on Åland for delivery in Sweden). The parties and the BAR framed the dispute in terms of a predominance approach versus a splitting approach, and they raised ancillary questions about classification as goods or services and about the potential application of an importrelated exemption. HFD did not name the principles as such; instead, it applied the compositesupply method developed by the CJEU and asked what the customer principally sought. On the facts, the principal element was the distribution of correspondence; the printing and enveloping stages merely enabled that end. The supply was therefore a single transport service rather than several separate supplies or a supply of goods with subordinate services. That classification in turn determined the place of supply and foreclosed reliance on import rules; once the supply was domestic transport, the claimed import-exemption path was not engaged. The court thus confirmed the advance ruling. Although concise and unnumbered, the reasons communicate the operative tests and their consequences with sufficient precision for downstream application.

Taken together, these decisions illuminate how HFD's precedent operates in Swedish VAT law. The court states the controlling test at a level of generality that lower instances can use, situates the test within the relevant statutory framework and Union law method, and then resolves the concrete dispute. The doctrinal signal is usually clear enough that administrative guidance is adjusted to mirror the holding, which in turn steers future firstinstance and appellate adjudication. While the formal rule remains that HFD's judgments bind only the case at hand, the combination of the leave-to-appeal filter, the court's economy of reasoning, and the institutional expectation of uniformity means that the precedential force of HFD's VAT case law is both practical and substantial.

8. Summary and conclusions

This chapter examines how HFD shapes VAT doctrine in a system without formal *stare decisis* but where its decisions guide lower courts and agencies in practice. Based on a dataset of roughly 670 VAT outcomes, it analyzes how HFD develops case law through overruling and distinguishing, and how it signals doctrinal movement through reasoning style, publication practice (full reports and short notes), and occasional plenary sittings.

VAT issues reach HFD mainly via two channels: appeals that receive leave to appeal because they are important for legal guidance, and appeals from advance rulings issued by the independent Board for Advance Tax Rulings, which do not require leave. Courts also have a constitutional duty to disapply norms that conflict with superior law. EU law frames much of the analysis: the BAR is not considered a "court or tribunal" for preliminary references, so questions of EU VAT interpretation reach the CJEU through the ordinary courts, with HFD giving the final national word when no reference is made.

HFD rarely writes "we overrule". Instead, it uses form and technique to signal change: convening in plenary when appropriate and pairing reported decisions with short notes to propagate clarifications across fact patterns. A clarifying example is HFD 2023 ref. 24, holding that voluntary VAT liability for letting premises is not barred merely because several tenants share space without exclusive rights; this builds on earlier decisions that decoupled "permanent use" from a single long-term tenancy and confined voluntary liability to a specific property or delimited part. By contrast, HFD 2023 ref. 45 recalibrates Swedish practice on input-VAT allocation in mixed activities by recognizing the Directive's turnover-based allocation method where national law is underspecified, essentially a constructive overruling without saying so. HFD's citation technique is concise: it identifies controlling propositions from EU and national case law and applies them without lengthy quotation. HFD 2021 ref. 18 exemplifies this method in fraud and abuse cases, emphasizing the "knew or ought to have known" standard and rejecting any requirement to prove a final, factual loss of VAT. In pricing cases like HFD 2023 ref. 48, the Court restates that

VAT should burden only the final consumer and limits adjustments to chains where VAT is charged at every stage; because Swedish retail sales of prescription medicines are exempt, no downward adjustment was allowed on those facts. Older, unnumbered decisions such as HFD 2012 ref. 43 show the same EU-method approach to composite supplies, classifying the entire package as a single transport service based on the principal/ancillary test and thereby determining place-of-supply and excluding import-exemption routes.

In conclusion:

- Doctrinal change is signaled by craft, not labels. HFD relies on concise formulations, selective plenary sittings, and publication strategy (reported cases plus notes) to clarify or redirect doctrine without overt declarations.
- EU law provides the compass. Neutrality, abuse-of-rights, single-supply characterization, taxable-amount rules, and the knowledge standard in fraud cases structure HFD's reasoning and keep Swedish outcomes aligned with the EU acquis.
- Clarification versus recalibration. Some decisions consolidate and clarify existing lines (e.g., permanence in voluntary liability for lettings); others recalibrate practice to match the Directive's architecture (e.g., turnover-based allocation in mixed activities), functioning as constructive overrulings.
- Methodological consistency yields predictability. HFD states the governing rule briefly and applies it to the record, which lower courts and agencies can readily operationalize. Administrative guidance tends to track these holdings, reinforcing uniform application.
- Institutions focus the docket on precedential questions. The leave-to-appeal regime and the advance-ruling pathway channel issues of general importance to HFD, while constitutional review obligations and the courts-based route for EU interpretation reinforce HFD's harmonizing role.

Overall, Swedish VAT jurisprudence shows a stable balance: lower instances generally follow HFD's guidance, and the Court retains flexibility to refine domestic doctrine in step with EU law without unnecessary disruption.

CHAPTER VIII

THE ADDED VALUE OF JUDICIAL INTERPRETATIVE FORMULAS IN DISCOVERING JUDICIAL INTERACTIONS AMONG EUROPEAN AND ITALIAN COURTS

Federica Casarosa – Madalina Moraru*

CONTENTS: 1. Judicial interactions among European and national courts – an ongoing process of refinement of EU law. – 2. Judicial interactions – definition and typologies. – 3. The added value of Judicial Interpretative Formulas in analysing judicial interactions. – 4. New avenues for research. – 5. Tentative conclusions.

1. Judicial interactions among European and national courts – an ongoing process of refinement of EU law

The European legal system embeds a decentralised system of enforcement, which relies on the coordinated work of the judiciaries across the Member States. This system acknowledges every national judge as a 'European' judge who not only masters the European and national implementing legal provisions but is also able to interpret them to achieve harmonisation and consistency within this multi-level system of rules [1].

* Federica Casarosa is Research Affiliate at the Scuola Superiore Sant'Anna and Madalina Moraru is Associate Professor of EU Law at the Departement of Political and Social Sciences, of the University of Bologna.

[1] M. CARTABIA, *"Taking Dialogue Seriously" The Renewed Need for a Judicial Dialogue at the Time of Constitutional Activism in the European Union*, Jean Monnet Working Papers 12, 2004; F. PERRONE, *The Judicial Path to European Constitutionalism: The Role of the National Judge in the Multi-Level Dialogue*, in P. PINTO DE ALBUQUERQUE, K. WOJTYCZEK (eds.),

However, the normative hierarchy, overlap, and cross-referencing between national and European legislation create a complex web of fundamental rights that is difficult to untangle. Therefore, a set of tools is required to allow the players to interact and receive support. The coordinating role is played by the Court of Justice through the preliminary ruling system, as defined in Article 267 TFEU. This type of vertical and transnational judicial interaction involves the national court seeking advice on resolving specific conflicts of interpretation of EU legislation, while also influencing the subsequent application of the same provision by other courts [2]. However, this is not the only way courts may resolve conflicts, as other explicit or implicit forms of interaction may also emerge.

Judicial interactions can be found in several areas of law, and their importance has also been acknowledged by academic literature [3]. Several studies affirm the necessity of judicial interaction due to the decentralised and non-hierarchical nature

---

Judicial Power in a Globalized World: Liber Amicorum Vincent De Gaetano, Springer International Publishing, 2019; U. JAREMBA, *At the Crossroads of National and European Union Law. Experiences of National Judges in a Multi-Level Legal Order,* in *Erasmus Law Review*, 2013, p. 191 ff.

[2] A. WEBER, *A continuous dialogue between courts as an integration and harmony way through the preliminary reference of the CJEU*, in *YEUCL*, 2024, 3, pp. 927–973; J. KROMMENDIJK, *It Takes Two to Tango: The Preliminary Reference Dance Between the Court of Justice of the European Union and National Courts*, Special Issue of European Papers, 2020; M. BOBEK, *Of feasibility and silent elephants: The legitimacy of the Court of Justice through the eyes of national courts*, in M. ADAMS et al. (eds.), *Judging Europe's judges: the legitimacy of the European Court of Justice Examined*, Har, 2013, p. 197 ff.; T. DE LA MARE, C. DONNELLY, *Preliminary Rulings and EU Legal Integration: Evolution and Stasis* in P. CRAIG, G. DE BÚRCA (eds.), *The Evolution of EU Law*, OUP, 2011, p. 363 ff.

[3] F. CASAROSA, M. MORARU (eds.), *The practice of judicial interaction in the field of fundamental rights : the added value of the charter of fundamental rights of the EU*, Edward Elgar, 2022; M. MORARU, G. CORNELISSE, P. DE BRUYCKER, *Law and Judicial Dialogue on the Return of Irregular Migrants from the European Union*, Edward Elgar, 2022; U. RENÁTA, *The Perils of Defending the Rule of Law Through Dialogue*, in *European Constitutional Law Review*, 2019, p. 1 ff.; E. PAUNIO, *Conflict, Power, and Understanding - Judicial Dialogue between the ECJ and National Courts*, in *No Foundations: An Interdisciplinary Journal of Law and Justice*, 2010, p. 5 ff.; F.G. JACOBS, *Judicial Dialogue and the Cross-Fertilization of Legal Systems: The European Court of Justice*, in *Texas International Law Journal*, 2003, p. 547 ff.; K. LENAERTS, *Upholding the Rule of Law through Judicial Dialogue*, in *Yearbook of European Law*, 2019, p. 3 ff.; A. ROSAS,

of EU law, which may lead to inconsistencies and fragmentation in its application. However, the analysis has so far been based on qualitative rather than quantitative data, as sector-specific and Europe-wide analysis of the caselaw was still impossible as too resource-intensive [4].

This contribution aims to demonstrate the added value of the POLINE project in clarifying the impact of judicial interactions across Member States, thanks to the analysis of the judicial interpretative formula (JIF) [5]. To provide an overall methodological framework, it is crucial to identify the boundaries of the concept of judicial interactions (par. 2) and their connection with the JIF (par. 3). Then, some possible applications of JIF discovery will be presented (par. 4), followed by concluding remarks (par. 5).

2. Judicial interactions: Definition and typologies

The phenomenon, defined as judicial dialogue between courts at the European and national levels, is one of the most interesting themes in contemporary legal-political studies, and it has been widely discussed in doctrine; however, no clear and univocal definition of the concept has been achieved. At the beginning of the 1990s, the metaphor of "dialogue" was used to describe the application of the so-called comparative jurisprudence method: when a judge examined the case law of other courts, it was claimed that the choice of the best solution was the result of a dialogue with those other courts. In this respect, it was referred to as dialogue between different jurisdictions, judicial dialogue, horizontal and vertical communication, transjudicial communication, or "interjudicial

The European Court of Justice in Context: Forms and Patterns of Judicial Dialogue, in *European Journal of Legal Studies*, 2007, p. 121 ff.

[4] A. Egea-de Haro, *How Does Case Law Shape Civil Law Systems? An Analysis of Spanish Administrative Courts*, in *Liverpool Law Review*, 2024, p. 1 ff.

[5] Note that the terminology used at the beginning of the project was judicial principles of law (JPOL), however, the theoretical and comparative legal analysis carried out during the project brought to a change, due to the risks of ambiguity. The choice taken by the project team was to the above-mentioned judicial interpretative formula, which aligns with the project's conceptual and methodological aims. See more details in the Introduction by P. Santin and A. Fidelangeli, in this book.

dialogue", encompassing the progression from reception to dialogue [6].

As a result of the increasing relations among legal systems – and consequently the increased opportunities for judges of different countries to interact – dialogue between courts was strengthened: the reference to statements and principles on fundamental rights formulated by foreign jurisprudence flourished. Many studies have addressed the judicial interactions that occur between the CJEU and the ECtHR, between national courts from different jurisdictions and the CJEU, between national courts and the ECtHR, between European and international courts, and between national supreme or constitutional courts across Europe and the world [7]. In such cases, judicial dialogue involves an ongoing exchange of arguments to achieve common understandings. However, dialogue requires some form of reciprocity among the judicial actors involved, and it develops on a case-by-case basis over time, whereas judicial practice is more often a monologue rather than a dialogue. This is the reason for the use of a broader concept of "judicial interaction" that reflects instances of interaction that may differ in intensity, outcome, and typology, and aims to construct a better interpretation of a legal norm, without necessarily involving reciprocity or continuity over time [8]. At a basic level, judicial interaction can

---

[6] M. SLAUGHTER, *A Global Community of Courts*, in *Harvard International Law Journal*, 2003, p. 191 ff.; ID., *A typology of Transjudicial Communication*, in *University of Richmond Law Review*, 1994, p. 99 ff.; P. WOJCIKIEWICZ ALMEIDA, *The Asymmetric Judicial Dialogue Between the ICJ and the IACtHR: An Empirical Analysis*, in *Journal of International Dispute Settlement*, 2020; G. MARTINICO, O. POLLICINO, *The Interaction between Europe's Legal Systems: Judicial Dialogue and the Creation of Supranational Laws*, in *The Interaction between Europe's Legal Systems*, Edward Elgar Publishing, 2012.

[7] S. DOUGLAS SCOTT, *A Tale of Two Courts: Luxembourg, Strasbourg and the Growing European Human Rights Acquis*, in *Common Market Law Review*, 2006; K. DZEHTSIAROU, T. KONSTADINIDES, T. LOCK, N. O'MEARA, *Human Rights in Europe, The Influence, Overlaps and Contradictions of the EU and the ECHR*, Routledge, 2014; G. MARTINICO, *Judging in the Multilevel Legal Order: Exploring the Techniques of 'Hidden Dialogue*, in *King's Law Journal*, 2010, pp. 257-281; A. FØLLESDAL, B. PETERS, G. ULFSTEIN, *Constituting Europe: the Court of Human Rights in a national, European, and global context*, Cambridge University Press, 2013.

[8] M. MORARU, M. CLEMENT, *Judicial Interactions Upholding the Right to*

be characterised as a sequence of judgments that refer to each other, involving at least two courts and not necessarily in a hierarchical relationship, with the shared intention of improving the practice of interpreting or reviewing a supranational norm. Thus, judicial interactions embed a set of techniques used by courts and judges to promote coherence and coordination (or, at least, minimise the risk of conflicts) among different legal and judicial systems. The cooperative contraposition with foreign experiences, as a result, is not interpreted as a transplant of law through case law decisions, but rather as a form of dynamic knowledge flow aimed at ensuring respect for the law.

According to this approach, judicial interactions may occur 'vertically': between national judges and the CJEU, most notably in the form of the preliminary reference procedure, but also when national judges engage in other ways with the case law of the CJEU. Alternatively, interactions can be 'horizontal', between judges within the same Member State, and 'transnational', when national judges engage with the case law of courts in other Member States [9]. Judicial interactions can take various forms, of which the most important are (a) the duty of consistent interpretation of national law with EU legal obligations as interpreted by the CJEU; (b) the preliminary reference procedure; (c) disapplication of a national norm; (d) comparative reasoning with national legislation and jurisprudence from other Member States; (e) mutual recognition of foreign judgments.

Among them, the most used is the preliminary reference procedure. This direct judicial interaction technique is laid down in Art. 19 (3)(b) TEU and Art. 267 TFEU. According to Article 267 TFEU, Member States' courts may – and sometimes must – refer questions on the interpretation or validity of EU legal measures to the CJEU for a binding

---

Be Heard of Asylum Seekers, Returnees and Immigrants: The Symbiotic Protection of the EU Charter and General Principles of EU Law, in *The Practice of Judicial Interaction in the Field of Fundamental Rights*, Edward Elgar Publishing, 2022; F. Casarosa, M. Moraru, *Judicial Interactions in Action: A Tool for a More Powerful and Influential EU Charter of Fundamental Rights*, in F. Casarosa, M. Moraru (eds.), *The practice of judicial interaction in the field of fundamental rights: the added value of the charter of fundamental rights of the EU*, Edward Elgar, 2022), pp. 1–21.

[9] F. Casarosa, M. Moraru, *Judicial Interactions in Action*, cit., p. 8.

preliminary ruling. All national courts, regardless of their status in the national judicial hierarchy, can engage in direct interaction with the CJEU and submit preliminary questions regarding the correct interpretation or validity of EU law. The two main objectives and results of the judicial interaction technique of preliminary reference are: ensuring the coherence of the EU legal order and the respect of the fundamental principles of EU law (primacy, direct effect, and effectiveness) [10]. For this reason, the CJEU sometimes rephrases the questions formulated by national courts into principled questions of EU law, whose resolution is equally applicable in all Member States. Another outcome of this procedure is to provide the national judge with the tools to find a consistent interpretation of domestic norms in relation to EU law obligations, or to determine instead the disapplication of the latter [11].

In line with the case law, the CJEU clearly requires all Member States' courts to abide by its judgments. This is true not only with respect to the preliminary reference addressed to the judge of the main proceedings; all national judges must respect all judgments of the Court. Indeed, Court's judgments have an extended effect (*erga omnes*) as they clarify the interpretation of EU law rather than ensuring only the solution of the specific dispute. For instance, the preliminary rulings on the interpretation of the EU Charter are such principled judgments [12].

Along with the preliminary ruling, the judicial interaction technique most commonly used by the courts is consistent interpretation, which, in most cases, aims to remedy the

---

[10] In its preliminary ruling, the CJEU may either: state that the conflict of norms, judicial interpretation or between various fundamental rights and/or fundamental freedom(s) is non-existent; give guidance for its resolution by way of offering particular interpretation or the application of relevant tests; or state clearly the need to disapply the national law whenever it is applied in the context of the EU law. Through the preliminary reference mechanism, the CJEU may also determine that a specific legal provision in secondary EU law is in conflict with the EU Charter and invalidate it.

[11] T. TRIDIMAS, *The ECJ and the National Courts: Dialogue, Cooperation, and Instability*, in A. ARNULL, D. CHALMERS (eds.), *The Oxford Handbook of European Union law*, OUP, 2016, pp. 403.

[12] G. MARTINICO, *Retracing Old (Scholarly) Paths: The Erga Omnes Effects of the Interpretative Preliminary Rulings*, in *European Journal of Legal Studies*, 2023, p. 37 ff.

discrepancies between national and EU law. When applying national law that falls within the scope of EU law, national courts have a duty to interpret it as far as possible in line with the wording and purpose of the applicable EU law [13]. Consistent interpretation is also a crucial tool for upholding the autonomous meaning of legal terms in EU law and finding a 'fit' between EU and national law. Through the use of the consistent interpretation technique, national courts can achieve the result of remedying an apparent conflict between national legislation, judicial doctrine or administrative practice and a norm of EU law.

3. The added value of Judicial Interpretative Formulas in analysing judicial interactions

Courts do not decide cases by reinterpreting the procedural and substantial rules applicable to the factual circumstances every single time. Rather, courts rely on stratified concepts that have been interpreted and discussed in previous cases to support their reasoning [14]. This approach is in line with consistency and coherence of the legal system, but also allows for evolution whenever the interpretation is no longer adapted to the legal and social context. The reasoning shows this approach through the explicit and also implicit reference to previous decisions, by other courts, being them of the same level (for instance, when the Cassation court relies on Constitutional court decision) or of a different level (for instance, when tribunals rely on Court of Cassation decisions,

[13] According to CJEU, 13 November 1990, C-106/89, *Marleasing*, ECLI:EU:C:1990:395, national courts have a duty to interpret national law in conformity with EU law, even if the respective EU secondary provision has not yet been transposed by the domestic legislator. See S. Housman, *What Is the Role of the National Judge, through Conforming Interpretation, in Protecting the Law of the Union?*, in YEUCL, 2024, p. 691 ff.

[14] Although the doctrine of *stare decisis* is not applicable in civil law countries, the case law of previous courts has a persuasive role depending on circumstances such as the uniformity and consistency of courts' jurisprudential doctrines. See M. Cappelletti, *The doctrine of stare decisis and the civil law: A fundamental difference or no difference at all?*, 1981; F. Vincy, F. Parisi, *Judicial precedents in civil law systems: A dynamic analysis*, in *International Review of Law and Economics*, 2006, p. 519 ff.

or on European courts' decision)[15]. Thanks to these connections, the legal concepts may evolve and be applied to different circumstances.

When examining the CJEU, the previous description becomes even clearer, as the court reiterates 'formulas' based on its own jurisprudence to strengthen its legal authority[16]. Although the court has been criticised for a lack of consistency and coherence[17], self-citation is used as a tool for consolidating a jurisprudential criterion, which can result in clarifying the interpretation of the European legal framework and reducing litigation.

Even if the approach is well documented, a clear definition is still lacking, leaving scholars the option to recognise a formula as soon as they read it. The objective of the POLINE project is to fill this gap by identifying a set of criteria for recognising the so-called Judicial Interpretative Formulas (JIF). JIF provides a proxy for a set of concepts that are already embedded in the legal theory, such as *ratio decidendi*, holding or judicial dicta.

The proposed definition of JIF has been formalized in a set of criteria: the JIF is an interpretative statement of the judge as regards a rule or a general principle, or the consequences stemming from the interpretation of a rule or a principle in a legal system, the subsumption a fact within a rule, or also the qualification of a factual hypothesis as a concept contained within a rule[18]. This formalisation is crucial for achieving automatic extraction from national and European case law.

The analysis of the jurisprudence of the CJEU and the supreme courts of three Member States Italy, Bulgaria, and Sweden dedicated to VAT resulted in the collection of

---

[15] F. CAFAGGI, *The Impact of CJEU Judgments on National Legal Systems: Preliminary Thoughts on the Link with Judicial Dialogue*, in F. CASAROSA, M. MORARU (eds.), *The Practice of Judicial Interaction in the Field of Fundamental Rights*, Edward Elgar Publishing, 2022, p. 378 ff.

[16] J. KOMÁREK, *Reasoning with Previous Decisions: Beyond the Doctrine of Precedent*, in *The American Journal of Comparative Law*, 2013, p. 149 ff.

[17] L. AZOULAI, *La fabrication de la jurisprudence communautaire*, in P. MBONGO, A. VAUCHEZ (eds.), *Dans la fabrique du Droit européen: scénes, acteurs et publics de la Cour de justice des Communautéennes*, Bruylant, 2009, p. 153 ff.

[18] See 'First release of the corpus + draft annotation and validation guidelines', Deliverable 2.2, POLINE project, 2025, p. 18.

approximately one hundred judgments, dedicated partly to the exemption for activities in the public interest and partly to the taxable amount. Although the topic selected for the analysis is limited, the results of the extraction counted over 3,800 JIFs extracted from the text of such decisions [19]. Based on this set of data, the project team has developed a Pilot tool that enables the exploration of caselaw through the lens of concepts defined by JIFs.

This approach can become a groundbreaking tool, providing a quantitative analysis of the explicit and implicit ways in which judicial interactions may occur. The challenges emerging from the discovery of forms of citation and consistency in the legal reasoning of national and European courts are one of the factors that have hindered the conduct of thorough empirical research on the jurisprudence of courts. Thanks to the JIF-based analysis, this can be overcome, enabling legal experts to identify the correspondences and discrepancies that emerge from the interplay between the national and European legal frameworks.

4. New avenues for research

Based on the analysis of JIF, several research questions can be further explored to identify the emergence of judicial interactions among courts. One of the most interesting questions is the impact of European JIF on the national caselaw. In particular, how far the CJEU is considered by national courts in their own reasoning.

The impact can be clearly identified whenever an explicit citation is included in the text. This type of approach is primarily used when the national court consistently interprets national legislation in accordance with CJEU jurisprudence, explicitly mentioning the cases referenced in the argumentation. However, it is possible that national courts may use the JIF as a supporting argument without explicit reference, for instance as a result of the progressive

[19] For a detailed analysis of the process of extraction, thanks to advanced machine learning and NLP techniques, trained on expert-annotated data, see Ch. 10 and Ch. 12.

internalisation of the JIF in the caselaw. In this case, the identification of JIF can identify an implicit form of interaction that has not been previously documented [20], yet may highlight the process of harmonization occurring at the national level.

From a different perspective, it is essential to note that the jurisprudence of the CJEU in the tax law sector has facilitated harmonisation among Member States, even when legislative interventions were unable to achieve it. From the outset, the CJEU's interventions maintained an apparently unchanged structure built upon literal quotation; however, upon detailed analysis of the JIFs, a silent expansion of EU law becomes apparent [21]. On the one hand, the CJEU ensured predictability and continuity; on the other, it allowed for leeway in progressive interpretative changes. This emerged when the CJEU was able to shape the content of EU legislation in relation to its practical implementation, for instance, in cases involving exemptions or the concept of consideration [22]. Similarly, the court shaped the general principles in the field of tax law, in particular, the direct effect and neutrality principles [23].

Within this context, national courts may face difficulties in adapting to the silent evolution of the CJEU's interpretation of EU law. One of the potential further uses of the JIFs-based research could be to highlight the diachronic analysis of citations across the national courts' jurisprudence. The time

[20] See the results of the collection of the instances in the transnational training projects coordinated by the Centre for Cooperation at the European University Institute (namely TRIIAL, e-NACT, ACTIONES and JUDCOOP), available at https://cjc.eui.eu/data/. The database compiles all the case law selected during the training projects, thanks to the work of the CJC expert team in collaboration with national experts from the project's partners. Although the database counts 523 clusters of cases, the selection is based on the expertise and knowledge of the experts involved, without any aspiration to be exhaustive.

[21] Study of judicial principles of law, Deliverable 2.4, POLINE project, 2025, p. 13.

[22] For the concept of consideration see for instance, CJEU, 19 December 2012, C-549/11, *Orfey Balgaria*, ECLI:EU:C:2012:832.

[23] For instance, as regards direct effect see CJEU, 15 May 2014, C-337/13, *Almos Agrárkülkereskedelmi*, ECLI:EU:C:2014:328.

dimension can help in identifying if the national courts are more or less prone to adapt to modifications.

5. Tentative conclusions

This chapter has examined the phenomenon of judicial interaction in the European legal space through the analytical and methodological framework of JIFs. It began by situating judicial interaction as a defining feature of the multi-level European legal order—an order in which national and European courts cooperate, contest, and co-construct the meaning of legal norms. While earlier scholarship has focused primarily on the doctrinal and qualitative dimensions of judicial dialogue, this chapter has proposed the JIF as an operational concept capable of capturing the concrete linguistic and conceptual mechanisms through which such interactions take place.

The chapter's analysis revealed that judicial interactions occur in multiple forms—vertical, horizontal, and transnational—and that their intensity and structure vary depending on the interpretative techniques employed by courts. Among these, consistent interpretation and the preliminary reference procedure emerge as the most significant tools for ensuring coherence between EU and national law. Yet, beyond formal procedures, the chapter demonstrated that judicial reasoning itself becomes a vehicle for integration. Courts frequently embed interpretative formulas developed by the CJEU or by other national courts within their own decisions, thereby internalising shared reasoning patterns even without explicit citations. This phenomenon, which combines imitation, adaptation, and innovation, illustrates how harmonisation in EU law often progresses through implicit judicial learning rather than legislative intervention. The case studies in VAT law confirmed that such interpretative borrowing is both systematic and dynamic, producing a layered process of legal convergence that reshapes the understanding of judicial dialogue within the EU.

Looking forward, two main avenues for further research emerge.

First, a promising direction concerns the diachronic and

cross-jurisdictional analysis of JIFs. Future studies should examine how interpretative formulas evolve across time and legal systems, tracing both the diffusion of CJEU formulas into national jurisprudence and the feedback effects whereby national courts contribute to reshaping EU interpretative standards. This temporal and comparative perspective can illuminate the subtle dynamics of legal integration—how stable language may conceal evolving meanings—and provide a more nuanced understanding of how European law maintains coherence while accommodating national diversity. Such research would also enable the development of longitudinal indicators of judicial convergence, resistance, or innovation, offering empirical tools to assess the actual depth of European judicial integration.

Second, further inquiry should explore the normative and methodological implications of automating judicial reasoning analysis. The computational extraction of JIFs opens the possibility of combining doctrinal and empirical approaches in new ways, but it also raises theoretical and methodological questions about how meaning, reasoning, and authority can be represented in data-driven form. Future research could enhance our capacity to study judicial reasoning across languages and legal traditions, thereby deepening our understanding of how European courts collectively construct the law.

In conclusion, this chapter has shown that judicial interaction, when analysed through the prism of *Judicial Interpretative Formulas*, is not simply an expression of cooperation between courts but a constitutive mechanism of European legal integration. By exposing the implicit and explicit channels through which interpretative reasoning circulates, it can demonstrate how the daily interpretative practices of judges – more than institutional dialogue alone – sustain the dynamic coherence of the EU legal order.

CHAPTER IX

FROM TEXT TO KNOWLEDGE: AUTOMATIC ANNOTATIONS THROUGH LARGE LANGUAGE MODELS

Rachele Mignone – Davide Audrito – Ivan Spada – Luigi Di Caro*

1. Introduction

Under the Treaty on European Union and the Treaty on the Functioning of the European Union, the Court of Justice of the European Union (CJEU) is responsible for ensuring the uniform

* Rachele Mignone is PhD Fellow in Law, Science and Technology at the University of Bologna and University of Turin; Davide Audrito is PhD in Law, Science and Technology at the University of Bologna and University of Turin and Interoperability Consultant for the EU Commission; Ivan Spada is Research Fellow in Computer Science at the University of Turin; Luigi Di Caro is Associate Professor at the Departement of Computer Science of the University of Turin.

interpretation and application of EU law. A notable and evolving feature of this judicial practice is the Court's style of reasoning, which frequently involves the verbatim reproduction or subtle rephrasing of interpretative statements from previous judgments. Over time, this "copy-pasting" or "LEGO" technique (see Introduction) has become so pronounced that the significance of a precedent often resides not in the entire judgment, but in these specific, recurring paragraphs. We refer to these passages as Judicial Interpretative Formulas (JIFs), a term introduced to address this recognized drafting style.

While JIFs are a prominent feature of EU case law, similar recurring interpretative patterns are also prevalent in national legal systems, albeit with different characteristics. This chapter explores the use of Large Language Models (LLMs) to automatically extract JIFs from legal documents, encompassing not only CJEU judgments but also those from national supreme courts in Italy, Bulgaria, and Sweden. This study serves a dual purpose: it provides empirical evidence of the widespread nature of these interpretative patterns across different legal jurisdictions and facilitates the creation of a unique, multilingual case law dataset. This resource is expected to be used by legal scholars for conducting large-scale analyses on evolving trends in judicial reasoning across Europe, while simultaneously offering a practical reference for judges to ensure the consistent application of legal principles.

2. Dataset

This study is conducted on a dataset meticulously curated by experts in tax law. This dataset is a collection of court decisions specifically related to Value Added Tax (VAT). To ensure a broader perspective, the rulings were sourced from multiple jurisdictions and are presented in different languages. These decisions are categorized into four groups: those from the CJEU, and others from Italian, Swedish, and Bulgarian courts.

2.1. CJEU dataset

The European fraction of the dataset comprises CJEU judgments concerning VAT law, covering the period from

2006 to 2024. The CJEU was chosen as the primary source due to the accessibility and consistent structure of its judgments, which are well-suited for automated textual analysis. Judgments were retrieved from EUR-Lex [1], the official online repository of EU law.

While curating this section of the dataset, legal experts specializing in VAT law focused on two key subdomains: Taxable Amount and Exemptions for the Public Interest. A total of 101 relevant CJEU judgments were initially identified.

To optimize processing by LLMs, each judgment underwent a targeted preprocessing workflow. Recognizing that JIFs are exclusively found in the motivation section of the judgments, documents were truncated to this specific section. This approach significantly reduced the input size while preserving the most relevant content. The preprocessing resulted in an average document reduction of 15,317 characters or 2,384 words.

Further cleaning procedures were applied to eliminate extraneous characters, such as excessive newline characters, ensuring data cleanliness. The preprocessed text was then segmented into paragraphs and stored in CSV format, with original paragraph numbering retained to facilitate indexing and retrieval.

Although categorized by subdomain, the judgments were treated as a unified dataset for the primary analysis, given their common legal foundation in VAT law. The subdomain classifications were maintained to enable nuanced interpretation of the findings and to facilitate comparative analysis between the two areas of VAT law.

2.2. Bulgarian dataset

Bulgarian legal experts selected 120 decisions of the Supreme Administrative Court (SAC) of Bulgaria after defining jurisdiction-specific guidelines that could accommodate the unstructured nature of Bulgarian case law. These decisions related to VAT *Exemptions in the Public Interest* and *Taxable Amount*. The decisions were pre-

[1] https://eur-lex.europa.eu/

processed to convert them into plain text format and anonymize personal data.

The decisions were divided into three separate datasets:

- A Training Dataset (80 decisions).
- A Validation Dataset (20 decisions).
- A Test Dataset (20 decisions).

The 20 decisions in the Validation Set were manually annotated by legal experts using the specially adapted Bulgarian Annotation Guidelines. A double-tag annotation approach was employed, followed by a discussion to establish a unified annotation convention. The other 20 manually annotated decisions formed the Test Set.

2.3. Italian dataset

The Italian section of the dataset was curated by a team of Italian tax law experts. In selecting the judgments, they focused on decisions from the Court of Cassation (Supreme Court of Cassation). This subset comprises a total of 216 judgments spanning from 1989 to 2024.

Unlike the standardized structure of CJEU rulings, these Italian decisions have a less structured nature, including the lack of fixed, numbered paragraphs.

This structural difference necessitated the adoption of a distinct methodological approach for their analysis.

2.4. Swedish dataset

The Swedish dataset, compiled by Swedish experts, consists of a collection of 99 judgments deemed relevant within the domain of VAT in Sweden, spanning the period of 1995 to 2024.

These documents, similarly to the CJEU dataset, are categorized into two primary classifications: *Taxable Amount* and *Exemptions for Activities in the Public Interest*.

3. Extracting JIFs from CJEU judgments

3.1. Prompt and model selection

A systematic model selection process was implemented to

identify the optimal combination of LLM and model-agnostic prompt for extracting JIFs from CJEU judgments.

A set of LLMs, including Claude 3.7 Sonnet [2], DeepSeek-R1 [3], and Gemini 1.5 Pro [4], was evaluated against a manually annotated dataset of 11 randomly selected CJEU judgments. To facilitate efficient computational access and integrate the diverse LLMs within a unified environment, Quora's Poe [5] platform was used. This tool centralized access to multiple models, reducing the overhead of managing disparate API endpoints and significantly enhancing the experimental workflow's efficiency.

Each LLM was tested with four distinct prompting strategies to assess performance in terms of precision, recall, and F1-score for JIF identification, namely zero-shot, chain-of-thought, few-shot, and chain-of-thought few-shot prompting. The evaluation was conducted automatically using predefined metrics to ensure an objective comparison of model efficacy. The results of this process, which aimed to determine the most effective model-prompt pairing for legal text extraction, are presented in Table 1.

Table 1. Model selection evaluation table

		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

[2] Model snapshot: 2025-02-19. Temperature: 0.5, top-p: 0.7, top-k: 35.

[3] Model snapshot: 2025-05-28. Temperature: 0.35, top-p: 0.7, top-k: 35.

[4] Model snapshot: gemini-1.5-pro-002. Temperature: 0.20, top-p: 0.7, top-k: 35.

[5] https://poe.com/.

Following this quantitative analysis, a qualitative assessment of the extracted JIFs was performed by domain experts. This crucial step revealed the subjective nature of the task, as many instances initially classified as false positives by the automated metrics were, upon expert review, re-evaluated as valid JIFs, yielding performance results depicted in Table 2.

Table 2. Updated 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**

3.2. Extraction

The systematic model selection process identified Claude 3.7 Sonnet as the optimal LLM for extracting Judicial Interpretative Formulas. The model was chosen for its superior performance, as it demonstrated the highest precision, recall, and F1-score when combined with a few-shot prompting strategy.

This approach framed the extraction as a binary classification task: the model analysed each paragraph of a CJEU judgment to determine if it qualified as a JIF. To ensure a consistent and machine-readable output, the LLM was prompted to return its response in a structured JSON format. This structured output included the paragraph number, a Boolean field for JIF classification, and a detailed analysis.

A successful JIF extraction produces a structured JSON output. This output format ensures the data is machine-readable and includes the paragraph number, a Boolean field

indicating if the paragraph is a JIF, and a detailed analysis of the model's reasoning.

Consider the following paragraph from the case A & G Fahrschul-Akademie GmbH v Finanzamt Wolfenbüttel. (C-449/17):

"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."

The model's output for this paragraph would be:

"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 },...

Despite instructing the model to provide a structured JSON output, a crucial post-processing step was necessary to ensure data consistency and integrity. This step addressed deviations in the model's output that could hinder subsequent analysis. The corrections performed included removing extraneous model output that exceeded the JSON structure, sanitizing special characters that could disrupt parsing, and standardizing the capitalization of boolean fields, which were sometimes inconsistently returned as "true" or "false". Additionally, issues with inner quotations, a common feature in legal texts, were fixed to prevent JSON structure errors. This rigorous post-processing ensured the final dataset was uniformly

formatted, reliable, and ready for further quantitative and qualitative evaluation.

3.3. Results and evaluation

The evaluation process 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.

3.3.1. 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, amounting to an average of 14.08 JIFs per judgment.

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.

3.3.2. 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 disregard 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.

4. Extracting JIFs from national judgments

This section details the methodology used to extract JIFs from national judgments in Italy, Sweden, and Bulgaria. The process adapted the core principles from the CJEU extraction, while adapting them to expand the task.

4.1. Italian and Swedish extraction

Following the evaluation performed for selecting a model for the extraction of JIFs from CJEU (Table 2) judgments, the same model, Claude 3.7 Sonnet, was selected for this task. In this instance, however, the project was expanded to combine the extraction of JIFs with their classification within a custom

VAT taxonomy and the simultaneous extraction of the citations they contain (see the following chapter). To achieve this, the few-shot prompting strategy was augmented with a chain-of-thought approach. While this combined methodology proved to yield results of similar quality to simple few-shot prompting during the model selection phase, in this case, the new strategy was chosen and implemented to facilitate the model's adherence to the pre-established steps without omitting critical information.

Furthermore, to guarantee a robust and usable outcome, and mirroring the approach used for the CJEU extraction, the output was also requested in a structured JSON format, which facilitates the efficient processing of the extracted information in subsequent phases.

The prompt underwent an iterative and cross lingual evaluation, being tested in both English and the native languages of the national judgments. Performance analysis revealed that the model consistently achieved superior results when the prompt was presented in English, a finding that is consistent with the known English-language bias of many large-scale multilingual models [6].

For the few-shot prompting component, a series of distinct strategies was systematically evaluated to optimize the model's performance. The options assessed included: *i)* using the same English-language examples from the initial CJEU extraction, which maintained linguistic consistency with the prompt; *ii)* translating these examples into the respective languages of the judgments to test the model's ability to handle multilingual inputs; *iii)* employing a set of new, domain-specific examples selected by human experts from the examined judgments, with explanations provided in English; and *iv)* utilizing the same domain-expert-curated examples but with explanations written in the original language of the judgments. The thorough comparative analysis revealed that the fourth example style was the optimal approach. By providing the model with

[6] L. Schut, Y. Gal, S. Farquhar, *Do Multilingual LLMs Think In English?*, arXiv preprint – arXiv:2502.15603, 2025; Y. Guo, S. Conia, Z. Zhou, M. Li, S. Potdar, H. Xiao, *Do Large Language Models Have an English Accent? Evaluating and Improving the Naturalness of Multilingual LLMs*, arXiv preprint – arXiv:2410.15956, 2024.

domain-expert-curated examples and explanations in the native language of the judgments, this strategy significantly enhanced the quality of the extracted information and proved to be the most effective method for guiding the model.

The resulting prompts used for this task are provided in Section 6.2 (Italian) and Section 6.3 (Swedish). The classification of JIFs and the extraction of citations, which are part of this combined methodology, are detailed further in the following chapter.

4.2 Results and evaluation of JIF extraction from Italian and Swedish judgments

The extraction of JIFs from legal judgments produced the following quantitative results. A total of 1,532 alleged JIFs were extracted from 216 Italian judgments, yielding a mean of approximately 7.09 JIFs per judgment. Similarly, 479 alleged JIFs were extracted from 87 Swedish judgments, with a mean of approximately 5.51 JIFs per judgment.

To validate the extraction method, a manual *ex-post* evaluation was performed by legal experts. Due to the demanding nature of this task and the availability of national experts, we chose to limit the manual review to the Italian judgments, prioritizing a comprehensive evaluation of a single jurisdiction. This review focused on Italian Supreme Court judgments and aimed to identify false positives among the extracted JIFs.

The evaluation found that out of the total JIFs extracted from this subset, only 20 were misclassified as JIFs. This high level of precision is reflected in the final score, which is approximately 98.69%.

4.3. Bulgarian extraction and results

The extraction of JIFs from Bulgarian case law was performed separately.

4.3.1. Prompt creation and refinement

The first step involved developing a specialized prompt for

extracting JIFs from unannotated decisions. This was challenging because the concept of JIFs is not recognized in Bulgarian legal theory, and explicit interpretative formulas are not commonly used in SAC decisions on specific cases, unlike the Italian *interpretative principles of law*. Only a few interpretative rulings in VAT matters contain such explicit statements. Consequently, a detailed and explicit definition of JIF had to be included in the prompt.

Several rounds of experiments were conducted to revise the prompt, ensuring the correct identification of all sentences matching the JIF definition. Different prompt lengths and temperature settings were tested using OpenAI's large language model GPT-4o [7], with the final temperature set at 0.01. The results of the Bulgarian-language prompt experiments were documented, and the English translation of the final prompt was made available.

4.3.2. Validation and automatic annotation

The prompt was tested on the Validation Set and iteratively refined until its results closely matched the manual annotations. Evaluation was performed manually by the legal team, comparing the AI-extracted sentences with human annotations. Once the prompt demonstrated satisfactory and stable performance, it was applied to the Test Set. Further refinements minimized false positives. Finally, a procedure was implemented for the automatic merging of adjacent sentences initially identified as separate JIFs but belonging to the same context.

After achieving satisfactory performance on the Test Set, the prompt was applied to the remaining 80 unannotated decisions in the Training Set.

The final result yielded 445 JIFs extracted from SAC decisions: 108 JIFs were manually identified and annotated by APIS legal experts, and 337 JIFs were extracted automatically by the LLM using the developed prompts.

---

[7] OpenAI. GPT-4o [Large multimodal model], 2024. Retrieved from https://openai.com/the

5. Conclusion

This research shows that LLMs represent a viable method for automating the extraction of legal knowledge from multilingual corpora of case law, thereby overcoming significant structural and conceptual challenges in the legal domain. The project validated a systematic LLM-based methodology, demonstrating its superior performance on highly structured texts like those from the CJEU (achieving an average F1-score of 0.9315) and successfully adapting it to the less-standardized reasoning found in national judgments from countries like Bulgaria and Italy. By using advanced prompt engineering and iterative refinement, the process proved capable of handling jurisdiction-specific difficulties, such as the conceptual absence of JIFs in Bulgarian legal theory, culminating in the successful extraction of a substantial volume of JIFs. Ultimately, this work delivers a crucial, multilingual, annotated case law dataset that serves as an empirical resource for legal scholars to conduct comparative analysis of judicial reasoning and provides practitioners with a knowledge base for ensuring the consistent application of legal principles.

6. Prompts

6.1 CJEU extraction: Few-shot prompting

Text of the 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. Costs

- 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: [judgment text]*

### *6.2 Italian extraction: Few-shot chain-of-thoughts*

Text of the 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 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: Il riconoscimento dell'esenzione richiede, dunque, sotto il profilo soggettivo, che le prestazioni siano rese dai soggetti qualificati enumerati dalla normativa unionale e da quella interna, nonchè, sul piano oggettivo, che esse rientrino nel novero di quelle sanitarie per le quali sia ravvisabile la finalità perseguita dalle norme di esenzione, che è quella di agevolare l'accesso a determinate prestazioni nonchè la fornitura di determinati beni evitando i maggiori costi che deriverebbero dal loro*
```

assoggettamento all'iva (Corte giust 20 novembre 2019, causa C400/18, Infohos, punto 37).

Spiegazione: la Corte individua e specifica i requisiti oggettivi e soggettivi per l'applicazione dell'esenzione, alla luce delle ragioni per cui la stessa è stata prevista.

2. JIF: Va dunque affermato il seguente principio di diritto: in tema di IVA, ai sensi del D.P.R. n. 633 del 1972 , art. 13 , comma 1, e art. 73 della Direttiva IVA rifusa (2006/112/CE), nella base imponibile devono ricomprendersi tutti i corrispettivi, compresi eventuali integrazioni, dovuti al cedente o al prestatore di servizi sulla base delle condizioni contrattuali, non potendosi artificiosamente scomporre la prestazione economica, che è tendenzialmente unica ed indissociabile, fatta eccezione per gli elementi normativamente previsti (art. 79 della citata Direttiva).

Spiegazione: la Corte anticipa che sa enunciando un principio di diritto, che è per ciò stesso e sempre un JIF.

3. JIF: Va, quindi, ribadito il principio secondo cui "In tema d'IVA, alla domanda di rimborso non rientrante tra quelle previste dall'art. 30 del D.P.R. n. 633 del 1972, nel testo "pro tempore" vigente, e perciò non contemplata da disposizioni specifiche, si applica l'art. 21, comma 2, del D.Lgs. n. 546 del 1992, di carattere residuale e secondo il quale "la domanda di restituzione, in mancanza di disposizioni specifiche, non può essere presentata dopo due anni dal pagamento ovvero, se posteriore, dal giorno in cui si è verificato il presupposto della restituzione" (Cass., 31 luglio 2019, n. 20573; Cass., 23 ottobre 2015, n. 21674; Cass., 8 giugno 2011, n. 12433).

Spiegazione: la Corte cita esplicitamente un JIF originariamente formulato in una precedente sentenza

4. JIF: Il contenuto della categoria di "scuola riconosciuta" comprende - anche alla luce dell'evoluzione della prassi tributaria in materia - le imprese (enti o privati) in possesso di uno specifico riconoscimento di idoneità oggettiva e/o soggettiva allo svolgimento dell'attività di istruzione da parte della competente pubblica amministrazione, secondo la disciplina volta a

volta dettata dalla legge (condizione, quest'ultima, la cui sussistenza è incontestata nella specie).

Spiegazione: la Corte chiarisce quale sia l'interpretazione del termine "scuola riconosciuta" e quale sia l'ampiezza della sua applicazione.

Output Format (JSON):

```json

[

{

'jif_text': [JIF exactly how it appears in the original text, including final references if present],

"analysis":{

"reasoning": "text analysis and step-by-step full justification for JIF determination",

"meets_JIF_criteria": [True|False]

},

"references": [{"citation": legal references in the JIF if present , "formatted": according to one of the following, depending on citation type [Art. X, Regulation year/number/[EC, EEC or EU] or Art. X, Directive year/number/[EC, EEC or EU] or Art. X, c.c. or Art. X, c.p.c. or Art. X, Law number/year or

Art. X, d.p.r. number/year or Art. X, d.l. number/year or Art. X, d.lgs. number/year or Art. X, r.d. number/year}],

"JIF": [True|False],

"labels":list ol labels that best classify the JIF. Only choose the most relevant ones and the most specific ones that are applicable

},...

]

```

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

–

Labels: [indented list of labels]

–

judgment: []

6.3 Swedish extraction: Few-shot chain-of-thoughts

The prompt used for the extraction and classification of Swedish JIFs equals the one used on the Italian judgments. The difference between the two prompts lies in the provided examples. In this case the examples were the following:

> *1. JIF: I R 2005 ref. 81 konstaterade Hšgsta fšrvaltningsdomstolen att det inte fanns nŒgot lagstšd fšr att som mervŠrdesskatt behandla belopp som felaktigt angetts avse sŒdan skatt. Ett fšretag som debiterat och betalat mervŠrdesskatt fšr en omsŠttning som inte var skattepliktig hade dŠrfšr rŠtt till Œterbetalning av inbetalda belopp (jfr EU-domstolens avgšrande i de fšrenade mŒlen C-78/02ÐC-80/02, Karageorgou m.fl., p. 41, 42 och 53). Av rŠttsfallet fŒr anses fšlja att inte heller ett belopp som har berŠknats efter en fšr hšg skattesats ska behandlas som mervŠrdesskatt.*
>
> *Analysis: Texten innehåller en tydlig tolkning av rättsliga principer gällande mervärdesskattebehandling baserad på ett specifikt prejudikat (R 2005 ref. 81) och EU-rättspraxis. Den fastställer hur felaktigt debiterad mervärdesskatt ska behandlas och utvidgar denna tolkning till belopp som beräknats med felaktiga skattesatser.*
>
> *2. JIF: Beskattningsunderlaget fšr leverans av en vara ska utgšras av allt det som sŠljaren har erhŒllit eller ska erhŒlla fšr varan frŒn kšparen. Det belopp som felaktigt behandlats som mervŠrdesskatt och som kunderna betalat till bolaget fŒr anses ingŒ i ersŠttningen fšr de tillhandahŒllna produkterna. Beloppet ska sŒledes, i den mŒn det inte Œterbetalas till kunderna, rŠknas in i beskattningsunderlaget.*
>
> *Analysis: Texten ger en tolkning av vad som utgör beskattningsunderlaget för leverans av varor enligt mervärdesskattelagstiftningen. Den tolkar specifikt hur felaktigt debiterade belopp som betecknats som moms ska behandlas juridiskt - och klargör att sådana belopp, om de inte återbetalas till kunderna, måste inkluderas i beskattningsunderlaget. Detta representerar en rättslig tolkning av skattelagstiftningen som fastställer en princip för tillämpning i liknande fall.*

3. - JIF: 21. Om användaren väljer att betala ett högre belopp än minimibeloppet får den överskjutande delen anses vara en frivillig betalning utan direkt samband med den tillhandahållna dejtingtjänsten, vilket innebär att den delen inte ska beaktas vid beräkningen av beskattningsunderlaget (jfr Tolsma, C-16/93, EU:C:1994:80, punkt 19 och HFD 2019 ref. 68, punkterna 25 och 26).

Analysis: Texten ger en specifik tolkning av hur betalningar som överstiger minimibelopp ska behandlas i förhållande till beräkningar av beskattningsunderlaget. Den fastställer att sådana överskjutande belopp bör betraktas som frivilliga betalningar utan direkt koppling till den tillhandahållna tjänsten och därför bör uteslutas från beskattningsunderlaget. Tolkningen stöds av specifika hänvisningar till rättspraxis (Tolsma-målet från EU-domstolen och HFD 2019 ref. 68), vilket gör det till en tydlig rättslig tolkning av skatteprinciper som kan tillämpas på liknande situationer.

CHAPTER X

EXTRACTING KNOWLEDGE: AUTOMATIC EXTRACTION THROUGH NATURAL LANGUAGE PROCESSING

Giulia Grundler – Andrea Galassi – Federico Ruggeri – Paolo Torroni*

1. Introduction

Judicial Interpretative Formulas (JIFs) are general interpretative statements in case law that, while formulated in relation to a specific dispute, acquire broader normative relevance as they are cited and reused in subsequent rulings [1]. Despite their relevance, their identification is still largely manual and costly. To address this, we designed a machine-learning framework for their automatic extraction, based on datasets on four languages: English, Bulgarian, Italian, and Swedish.

From a technical standpoint, our work builds on prior

* Giulia Grundler is a Research Fellow at the University of Bologna; Andrea Galassi is a Junior Assistant Professor in Information Processing Systems at the University of Bologna; Federico Ruggeri is Research Fellow at the University of Bologna; Paolo Torroni is Associate Professor of Information Processing Systems at the University of Bologna.

[1] M. JACOB, *Precedents and case-based reasoning in the European Court of Justice: unfinished business*, Cambridge University Press, 2014.

research in the automated extraction of legal principles (or *rationes decidendi*) from judicial decisions. Earlier rule-based methods, based on regular expressions or handcrafted grammars, proved too rigid given the variability of judicial writing [2]. Subsequent machine-learning approaches, such as classifiers trained on annotated corpora, achieved promising performance but relied heavily on costly and time-consuming manual annotation [3]. More recently, LLM-based methods have been explored, showing potential but also raising concerns about transparency, reproducibility, and the need for expert validation [4].

Against this background, our work uses a hybrid approach: LLMs were employed to support dataset creation and reduce annotation costs, while for the core extraction task we fine-tuned BERT-based models. This chapter focuses on the extraction task, showing that these models achieve performance comparable to LLMs, with the added advantages of transparency, stability, and reproducibility, making them better suited for applications in the legal domain.

2. Corpus

As explained in the previous Ch., the experiments focused on two VAT subdomains: and specifically on taxable amounts and exemptions for the public interest.

For the English corpus, the source data consists of 101

---

[2] M. MOLINARI, M. QUARANTA, I.A. AMANTEA, *How do Case Law and Principles of Law interact, Computationally?* in *Proceedings of «Workshop on Innovation and Digitization of the Justice System»*, Within the XX Conference of the Italian Chapter of AIS - itAIS 2023, 2023.

[3] O. SHULAYEVA, OLGA, S. ADVAITH SIDDHARTHAN, E A. WYNER, *Recognizing cit-ed facts and principles in legal judgements* in *Artificial Intelligence and Law*, 2017, pp. 107–26; J. VALVODA, O. RAY. *From Case Law to Ratio Decidendi* in *New Frontiers in Artificial Intelligence – JSAI-isAI Workshop*s, JURISIN, SKL, AI-Biz, LENLS, AAA, SCIDOCA, kNeXI, Tsukuba, Tokyo, Japan, November 13-15, 2017.

[4] M. MOLINARI, M. QUARANTA, I.A. AMANTEA, *Using ChatGPT to Extract Principles of Law for the Sake of Prediction: an Ex-ploration conducted on Italian Judgments concerning LGBT(QIA+) Rights* in *Jurix '24 AI for Access to Justice Workshop*, 2024; I. CHALKIDIS et al. *LEGAL-BERT: The Muppets straight out of Law School*, in *Findings of the Association for Computational Linguistics: EMNLP 2020* [Online], 2020, pp. 2898–904.

preliminary rulings of the Court of Justice of the European Union (CJEU) in those subdomains.

For the national corpora, the Bulgarian source data consists of 120 rulings of the Supreme Administrative Court (see Ch. VI), the Italian data of 216 rulings of the Court of Cassation (see Ch. V), the Swedish data of 99 rulings of the Supreme Administrative Court (see Ch. VII).

Table 1 summarizes the details of the datasets. For each language, the test and validations set were manually annotated, following the methodology explained in paragraph 3.

Given the large costs of manual annotation by legal experts and the need for a large amount of data to train Transformer-based methods, we automatically annotated additional documents through Large Language Models (LLMs) to build the training sets. The details of this annotation process can be found in Ch. IX.

Language	Manually Annotated		Auto. Annotated	Total
	Test set	Validation set	Train set	
English (EU)	11	10	80	101
Bulgarian	20	20	80	120
Italian	19	19	178	216
Swedish	11	11	77	99

Table 1 Composition of the Dataset

3. Annotation methodology

We created our guidelines and datasets through multiple stages of annotation, each involving different documents and methodologies. As a reference, we report in **Table 2**, the number of documents and the corresponding annotation method involved in the annotation stages of the English corpus.

Set	Num of docs	Purpose	Annotators per doc	Annotators interaction	Split of destination
A	2	guidelines draft	2	collaborative	val
B	9	guidelines improvement	2	independent	val
C	2	guidelines validation	2	independent	test
D	8	increase test split	1	independent	test

Table 2 Document sets manually annotated at each stage and corresponding annotation setting, for the English corpus.

The annotation guidelines were conceived not only as a reference tool for human annotators, but also as a resource that could inform the automatic annotation methods. To enhance their utility, they incorporated illustrative examples drawn from authentic documents. Since this choice can facilitate automatic processing on those specific documents, to ensure a sound evaluation the corpus was divided into separate splits, ensuring that the test portion only included documents that were never used during the guideline development phase.

The initial draft of the guidelines was grounded in both theoretical considerations and preliminary empirical work. Two annotators jointly analysed and annotated an initial subset of documents (set A), progressively refining the guidelines until all sources of ambiguity were resolved. To assess their reliability, the annotators then independently annotated a new batch of decisions (set B), unseen during the drafting stage, in a double-blind setting. Any disagreements were subsequently discussed and reconciled, yielding to a consolidated version. Then, a third group of documents (set C) was annotated independently by both annotators and served to validate the final guidelines.

Inter-Annotator Agreement (IAA) was computed on set C, obtaining a Cohen's κ of 0.96, which indicates an almost perfect agreement.

Because sets A and B contributed to the refinement of the

guidelines and thus contained examples that could bias evaluation, they were excluded from the test split, and were combined into the validation split. Instead, since the guidelines were not altered after its annotation, set C was included in the test split. Given the limited size of set C, the test split was further expanded by adding a fourth set of decisions (set D), for which each document was annotated by a single expert.

4. Automated extraction

We addressed the automatic identification of JIFs by fine-tuning machine learning models on our dataset. The problem was modeled as a binary classification task: given a sentence or paragraph, determine whether it represents a JIF or not. The experimental setup relied on the train/validation/test partitions reported in Table 1, defined at the document level to prevent paragraphs from the same text appearing in different splits.

Importantly, both the validation and test splits exclusively contained manually annotated documents, whereas the training set consisted of automatically labelled data. This design choice guaranteed that evaluation was performed on high-quality human annotations, avoiding the risk of testing AI models against AI-generated labels.

A notable difference between English and the other languages is the unit of classification. In the English texts, the unit of classification is the paragraph, since the documents are clearly divided into numbered sections. Each paragraph is assigned a binary label (JIF or non-JIF). For the other languages, however, such a paragraph-level segmentation is not available. In these cases, we use the sentence as the unit of classification [5]: each sentence is labeled depending on whether it belongs to a possibly larger JIF or not. The task therefore remains binary in all languages, but for non-English texts the model only has access to more fine-grained, and thus

[5] Sentence segmentation was performed with a customized spacy pipeline that takes into account the common abbreviations of each language legal text, to avoid oversegmentation.

more partial, information. To address this limitation, we additionally introduce a context-aware classification setting, where the model is given not only the target sentence but also its immediate context, i.e., the preceding and following sentences in the text.

Here we report, for each language, the list of models we experimented with.

English:

- DistilRoBERTa [6]: a distilled version of the RoBERTa-base model. It follows the same training procedure as DistilBERT, and it is, on average, twice as fast as RoBERTa-base.
- DeBERTa [7]: an improvement of the BERT and RoBERTa models, using disentangled attention and enhanced mask decoder.
- LEGAL-BERT [8]: a family of BERT models for the legal domain, intended to assist legal NLP research, computational law, and legal technology applications.
- LinearSVC with TF-IDF features

Bulgarian:

- SlavicBERT [9]: a version of BERT initialized on Multilingual BERT and trained on Russian News and four Wikipedias: Bulgarian, Czech, Polish, and Russian.
- BERT multilingual [10]: a pretrained model on the top 104 languages with the largest Wikipedia, using a masked language modeling objective.

---

[6] V. SANH et al., *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*, arXiv preprint – arXiv:1910.01108, 2019.

[7] P. HE et al., *DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing,* arXiv:2111.09543*,* 2021.

[8] I. CHALKIDIS et al., LEGAL-BERT: The Muppets straight out of Law School. In Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 2898–2904.

[9] M. ARKHIPOV et al., *Tuning Multilingual Transformers for Lan-guage-Specific Named Entity Recognition,* in T. ERJAVEC et al. (eds.), *Proceedings of the 7th Worshop on Balto-Slavic Natural Language Processing Association for Computational Linguistics*, 2019, pp. 89–93.

[10] T. PIRES et al., *How Multilingual is Multilingual BERT?* in A. KORHONEN et al. (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2019, pp. 4996–5001.

- LinearSVC with TF-IDF features

Italian:

- Italian BERT [11]: an Italian version of BERT, trained on Italian Wikipedia and various texts from the OPUS corpora collection.
- ITALIAN-LEGAL-BERT [12]: a model based on Italian BERT with additional pre-training on Italian civil law corpora.
- LinearSVC with TF-IDF features

Swedish:

- Swedish BERT [13]: a BERT trained with the same hyperparameters as first published by Google, with text from various Swedish books, news, government publications, Wikipedia and internet forums.
- BERT multilingual.
- LinearSVC with TF-IDF features

For Bulgarian and Swedish a multilingual model was used due to the lack of monolingual models in these languages.

The BERT models were fine-tuned for 10 epochs with early stopping, a learning rate of 2e-5 and a batch size of 8. As a reference, we also report the performance of two baselines: a classifier that outputs a random class (Random baseline) and one that always predicts the majority class (Majority baseline).

5. Results

5.1. English dataset

Table 3 presents precision, recall, and F1 scores for each classifier across both classes, together with their macro-averages. DistilRoBERTa and LEGAL-BERT obtained the highest macro F1 score (0.76), while DeBERTa followed

---

[11] Bert-base-italian-cased (Revision 843e404), available on Hugging Face, 2025.

[12] D. LICARI, G. COMANDÈ, *ITALIAN-LEGAL-BERT: A Pre-Trained Transformer Language Model for Italian Law* in D. SYMEONIDOU et al. (eds.), *Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management*, CEUR, 2022, https://ceur-ws.org/Vol-3256/#km4law3. CEUR Workshop Proceedings.

[13] M. MALMSTE et al., *Playing with Words at the National Library of Sweden – Making a Swedish BERT*, arXiv:2007.01658, 2020.

closely with 0.74. On the positive class, LEGAL-BERT reached the best F1 value (0.76), with DeBERTa and DistilRoBERTa performing slightly below. DeBERTa achieved the highest recall (0.82) on the positive class, whereas DistilRoBERTa showed the strongest precision (0.75). With the exception of DistilRoBERTa, all models displayed higher precision on the negative class and higher recall on the positive one. Notably, LinearSVC, although simpler, produced results not far behind the transformer-based approaches, indicating that lexical features are highly informative for this task.

Overall, LEGAL-BERT is the most effective model: it not only ties for the best macro F1 score but also achieves the highest F1 on the positive class and nearly the best recall. This aspect is particularly valuable for our application, since in a tool designed for legal practitioners, it is preferable to return a few additional JIFs rather than risk omitting essential ones. Missing a relevant JIF would undermine trust in the system, while users can easily disregard a small number of irrelevant paragraphs.

Model	**Precision**			**Recall**			**F1 score**		
	yes	*no*	*avg*	*yes*	*no*	*avg*	*yes*	*no*	*avg*
Majority	0.00	0.54	0.27	0.00	1.00	0.50	0.00	0.70	0.35
Random	0.46	0.53	0.49	0.49	0.50	0.49	0.47	0.51	0.49
LinearSVC	0.68	0.76	0.72	0.75	0.70	0.72	0.71	0.73	0.72
LEGAL-BERT	0.73	0.82	0.77	0.80	0.74	0.77	**0.76**	0.77	**0.76**
Distil RoBERTa	0.75	0.79	0.77	0.75	0.78	0.77	0.75	**0.78**	**0.76**
DeBERTa	0.69	0.82	0.75	0.82	0.68	0.75	0.75	0.74	0.74

Table 3: Results for the JIF classification task on the English dataset

5.2. Bulgarian dataset

Table 4 reports precision, recall and F1 scores obtained by

each classifier for each class, as well as their macro-average. The best macro F1 score of 0.66 is obtained by SlavicBERT in the context-aware setting. However, this setting degrades the performance of BERT multilingual from 0.64 to 0.60, showing that having access to context is not always the best option. SlavicBERT with context obtains also the best F1 score over the positive class (0.43), and the best recall over the positive class (0.46). The highest precision of 0.42 is instead obtained by BERT multilingual without context. As in the English dataset, LinearSVC exhibits not much lower performance compared to BERT models, suggesting the relevance of lexical features in this context.

We observe that classification performance is consistently lower with respect to the English language. We hypothesize that this is due to the difference in segmentation: while English texts allow paragraph-level classification, in the other languages only sentence-level units are available, which provide more limited information and likely make the task harder.

Model	**Precision**			**Recall**			**F1 score**		
	yes	*no*	*avg*	*yes*	*no*	*avg*	*yes*	*no*	*avg*
Majority	0.00	0.85	0.43	0.00	1.00	0.50	0.00	**0.92**	0.46
Random	0.16	0.87	0.51	0.55	0.51	0.53	0.25	0.64	0.44
LinearSVC	0.42	0.87	0.65	0.19	0.96	0.57	0.26	0.91	0.59
SlavicBERT	0.39	0.89	0.64	0.38	0.90	0.64	0.38	0.90	0.64
BERTmultil.	0.42	0.89	0.66	0.35	0.92	0.63	0.38	0.90	0.64
with context									
SlavicBERT	0.40	0.91	0.65	0.46	0.88	0.67	**0.43**	0.89	**0.66**
BERTmultil.	0.39	0.88	0.63	0.24	0.93	0.59	0.30	0.91	0.60

Table 4 – Results for the JIF classification on the Bulgarian dataset

5.3. Italian dataset

Table 5 reports precision, recall and F1 scores obtained by each classifier for each class, as well as their macro-average. The highest macro F1 score of 0.62 is obtained by ITALIAN-LEGAL-BERT in both context-aware and non context-aware settings. However, we consider the context-aware setting to be slightly better because it has the best F1 score on the positive class (0.32). Italian BERT exhibits slightly lower performance with respect to ITALIAN-LEGAL-BERT, higher in the setting without context.

This is the only dataset where LinearSVC achieves low performance, comparable to the majority baseline, suggesting that the Italian judgments employ a less standardized language and that lexical features alone are insufficient for accurate classification.

Moreover, as already observed for the Bulgarian dataset, classification performance is consistently lower with respect to the English language, and we attribute this mostly to the difference in segmentation.

Model	**Precision**			**Recall**			**F1 score**		
	yes	*no*	*avg*	*yes*	*no*	*avg*	*yes*	*no*	*avg*
Majority	0.00	0.91	0.45	0.00	1.00	0.50	0.00	**0.95**	0.48
Random	0.10	0.92	0.51	0.54	0.49	0.52	0.17	0.64	0.40
LinearSVC	0.06	0.91	0.49	0.01	0.98	0.50	0.02	0.94	0.48
Italian BERT	0.26	0.93	0.60	0.25	0.93	0.59	0.26	0.93	0.59
IT-LEG-BERT	0.28	0.93	0.61	0.35	0.91	0.63	0.31	0.92	**0.62**
with context									
Italian BERT	0.24	0.92	0.58	0.16	0.95	0.56	0.19	0.93	0.56
IT-LEG-BERT	0.29	0.94	0.61	0.38	0.91	0.64	**0.32**	0.92	**0.62**

Table 5 – Results for the JIF classification task on the Italian dataset

5.4. Swedish dataset

Table 6 reports precision, recall and F1 scores obtained by each classifier for each class, as well as their macro-average. The highest macro F1 score of 0.66 is obtained by Swedish BERT in the context-aware setting. It also reaches the best f1 score on the positive class (0.37), while the highest score on the negative class belongs to the majority baseline. BERT multilingual with context is the second-best model, with a macro F1 score only 0.01 point lower than Swedish BERT (0.65). Both models without context and LinearSVC reach comparable macro F1 scores, ranging from 0.60 to 0.62.

As previously observed with Bulgarian, the multilingual model performs worse than the language-specific one. This is a common finding in the literature, where multilingual models often underperform compared to monolingual counterparts, especially when trained on languages with limited resources or less standardized linguistic features [14]. Additionally, research indicates that adding large amounts of multilingual data can harm performance, likely due to limited model capacity, a phenomenon known as the 'curse of multilinguality' [15].

Moreover, as already observed for the Bulgarian and Italian datasets, classification performance is consistently lower with respect to the English language, and we attribute this mostly to the difference in segmentation.

Model	**Precision**			**Recall**			**F1 score**		
	yes	*no*	*avg*	*yes*	*no*	*avg*	*yes*	*no*	*avg*
Majority	0.00	0.93	0.47	0.00	1.00	0.50	0.00	**0.97**	0.48
Random	0.08	0.95	0.52	0.62	0.51	0.57	0.15	0.66	0.41
LinearSVC	0.30	0.94	0.62	0.20	0.97	0.58	0.24	0.96	0.60

[14] K. YOOJOONG et al., *A pre-trained BERT for Korean medical natural language processing* in *Scientific Reports*, 2022, p. 13847 ff.; T. PIRES et al., *How Multilingual is Multilingual BERT,* cit.

[15] T. CHANG, A. TYLER et al., *When Is Multilinguality a Curse? Language Model-ing for 250 High- and Low-Resource Languages* in Y. AL-ONAIZAN et al. (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2024, p. 4074 ff.

Model	Precision			Recall			F1 score		
	yes	*no*	*avg*	*yes*	*no*	*avg*	*yes*	*no*	*avg*
Swedish BERT	0.23	0.96	0.60	0.49	0.88	0.69	0.31	0.92	0.62
BERT multil.	0.23	0.95	0.59	0.32	0.92	0.62	0.27	0.93	0.60
with context									
Swedish BERT	0.29	0.96	0.63	0.52	0.91	0.71	**0.37**	0.94	**0.66**
BERT multil.	0.30	0.96	0.63	0.46	0.92	0.69	0.36	0.94	0.65

Table 6 – Results for the JIF classification task on the Swedish dataset

6. Linking JIFs to ontology concepts

The task of linking JIFs to ontology concepts starts from the extracted JIFs and aims at classifying them into one or more of the ontology concepts explained in Ch. XI. It is therefore a multi-label and multi-class classification task.

Here we focus our analysis on the English dataset, since it is the only one with gold labels for this task, and therefore the only one where an evaluation of the models is possible. We focus on a subset of the ontology, particularly on the higher-level concepts: the division of the JIF concept into *non-vat* or *value added tax*, and the subelements of the latter. Of these subelements we consider the ones with at least 10 examples in the test set.

Table 7 shows the composition of our dataset. The test set is manually tagged (gold), while the validation and train set are automatically annotated with Claude 3.7 Sonnet.

ontology concept	test	val	train
non-vat	12	11	74
value added tax	90	156	996
exemptions	23	74	451
principle of fiscal neutrality	12	24	107
principle that national law must be interpreted in conformity with EU law	13	16	47
taxable amount	34	48	303
taxable persons	19	10	134
taxable transactions	13	9	192

Table 7 – Composition of the dataset of the ontology concepts used in our experiments

We fine-tune LEGAL-BERT, which was the best model for the extraction task, for 10 epochs with early stopping, a learning rate of 2e-5 and a batch size of 8.

We report in Table 8 the results of the classification task, with precision, recall and F1 score of each class and their macro average. As concerns the higher-level division in *vat* and *non-vat*, the *non-vat* minority class obtains perfect precision but low recall, which leads to an F1 score of 0.50. The *vat* sub-elements reach good scores, ranging from 0.67 of *taxable transaction*, to 0.92 of *taxable persons*, with the exception of *principle that national law must be interpreted in conformity with eu law*, which obtains a score of 0.00 in both precision and recall. This low result could be attributed to the lower representation of this concept in the training set, with only 47 examples.

Despite the low result for this class, the task reaches a macro F1 score of 0.69 with 0.75 precision and 0.67 recall.

Ontology concept	precision	recall	F1 score
non-vat	1.00	0.33	0.50
value added tax	0.92	1.00	0.96
exemptions	0.78	0.91	0.84
principle of fiscal neutrality	0.75	0.75	0.75
principle that national law must be interpreted in conformity with eu law	0.00	0.00	0.00
taxable amount	0.87	0.82	0.85
taxable persons	0.94	0.89	0.92
taxable transactions	0.73	0.62	0.67
macro	**0.75**	**0.67**	**0.69**

Table 8 – Results for the ontology concepts classification

7. Limitations

A recurrent issue in the legal domain concerns the reliance on proprietary LLMs, whose training data, internal architectures, and algorithms are not publicly accessible. This lack of transparency, coupled with the well-documented risk of hallucinations, makes their uncritical adoption problematic in sensitive contexts like law. In our work, however, LLMs are not the end point but serve only during the dataset construction phase, where they help accelerate annotation and lower its overall cost.

The actual extraction of JIFs is carried out by dedicated, task-specific models, reducing dependence on proprietary systems and providing a more transparent, reliable, and accountable solution for legal text analysis.

8. Conclusions

In this chapter, we presented the NLP models developed for

the extraction of JIFs from CJEU and national VAT decisions. For each language, we compare a set of fine-tuned BERT models with an SVM using lexical features. We observe that classification performance in non-English languages is consistently lower compared to English. We attribute this primarily to differences in segmentation, which in turn arise from variations in text structure. Specifically, while English judgments are clearly structured and allow for paragraph-level classification, the other languages only permit sentence-level segmentation, which provides more limited information and likely increases the difficulty of the task. As for the linking of JIFs to ontology concepts, we focus our analysis on English and on the main concepts of the ontology - those sufficiently represented in the annotated dataset - achieving good results.

CHAPTER XI

STRUCTURING KNOWLEDGE: ONTOLOGY-BASED APPROACHES TO SIMILARITY AND CLASSIFICATION

Rachele Mignone – Davide Audrito – Ivan Spada – Luigi Di Caro*

CONTENTS:

1. Introduction

The Court of Justice of the European Union (CJEU) ensures EU law is applied consistently across all Member States. A key part of its reasoning is the use of specific interpretative statements from past rulings, which we define as Judicial Interpretative Formulas (JIFs).

---

* Rachele Mignone is PhD Fellow in Law, Science and Technology at the University of Bologna and University of Turin; Davide Audrito is PhD in Law, Science and Technology at the University of Bologna and University of Turin and Interoperability Consultant for the EU Commission; Ivan Spada is Research Fellow in Computer Science at the University of Turin; Luigi Di Caro is Associate Professor at the Department of Computer Science of the University of Turin.

This work builds on a previous phase of this study, presented in Chapter IX, that employs Large Language Models (LLMs) to identify and extract JIFs from CJEU judgments, creating a dataset for legal research. This resource aims to help scholars analyse trends in judicial reasoning and to give legal professionals a practical reference tool.

Expanding on that work, we introduce a new, broader framework. Firstly, we create a multilingual ontology that describes the Value-Added Tax (VAT) domain. We then utilize LLMs for classifying previously extracted JIFs according to the ontology and for extracting legal references within JIFs. The retrieved information is finally used to provide deeper insight into the JIFs and how they relate to each other, thanks to pairwise similarity assessment and citation network analysis.

2. Multilingual ontology creation

This project phase focuses on the creation of a multilingual ontology. The purpose of this ontology was to serve as a structured framework for classifying JIFs.

The initial set of ontology labels was derived from the existing EUR-LEX [1] taxonomy used by the platform to categorize judgments of the CJEU in its Directory of Case law. This established a foundational structure based on a recognized and authoritative source. To enhance its utility and further adapt it to the project's scope, the taxonomy was manually enriched by experts specialized in tax law who contributed additional terms and relations, thereby expanding the taxonomy's breadth and depth beyond its original scope. Additionally, a dedicated 'Non-VAT' node was established to classify extracted JIFs that were determined to be outside the project's predefined scope.

This process resulted in a set of 127 labels (i.e., nodes) and 130 edges of two different kinds. The first type of edge is the *has_subtopic* one, representing the taxonomical relation between nodes, and the second one is *divided_in* one, which was added to the ontology to link the VAT and Non-VAT

[1] https://eur-lex.europa.eu/.

branches. This was done to symbolize that JIFs can be categorized into one of these two divisions.

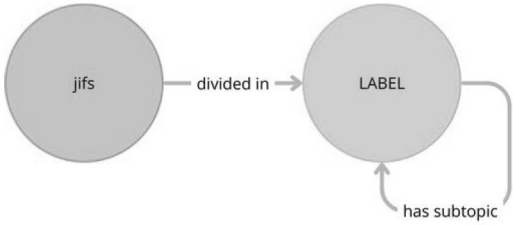

Figure 1: Ontology schema

Translating the labels into the four target languages was a critical step, and a tailored semi-automatic method was developed to ensure accuracy and reliability. This approach was necessitated by the inadequacy of available machine translation systems for handling the specific complexities of European Union juridical language.

The identified labels were grouped into three distinct categories, each necessitating a specialized translation methodology. This stratified approach was essential to maintain translation accuracy while minimizing the intensive manual labor required from legal experts.

First, for labels sourced from the EUR-LEX Directory of Case Law taxonomy, official translations were available directly on the platform. These were automatically extracted to ensure complete fidelity to the source. A second set of labels was manually added to the taxonomy, drawn directly from the keywords found in the headings of CJEU judgments. Their standardized location within the document structure made them readily identifiable and extractable via web scraping techniques with minor preprocessing. To find the officially accepted translation of these specific terms, it was necessary to: i) identify one of the judgments characterized by each keyword, ii) download the official translation of the judgment in all target languages, and iii) automatically identify the corresponding translation.

Finally, any remaining labels that did not fit into the previous categories were translated manually by a team of Italian, Swedish, and Bulgarian domain experts. Despite being

labour-intensive, this part of the process was crucial for maintaining precision and ensuring the terms were accurately represented in each language. Upon the completion of all translation tasks, the results were aggregated into a single, unified resource. This resource takes the form of a multilingual dictionary, in which each English term is mapped to its corresponding translations in Italian, Swedish, and Bulgarian.

The final step involved creating an ontology to represent the relationships between the labels. This ontology was developed using the Turtle (TTL) standard, a syntax for the Resource Description Framework (RDF) designed for serializing graphs. The use of TTL provides a structured and machine-readable representation of the taxonomy, allowing for its integration into various data processing and analysis pipelines. The ontology schema, as depicted in Figure 1, provides a visual representation of the relationships and hierarchy within the VAT domain taxonomy. Using the previously created multilingual dictionary, the ontology was translated from English into the other corresponding languages. This process yielded four linguistic versions of the same ontology.

3. JIF classification

Once the ontology was created, it served as the framework for classifying the JIFs that were extracted by LLMs, a process detailed in the previous chapter. This classification aimed to place each JIF within the VAT domain by associating it with a specific node in the ontology, thus enabling its potential indexing.

The classification itself was performed using LLMs, which employed two distinct methodologies depending on whether the JIF originated from the CJEU or a national jurisdiction (namely, Sweden and Italy).

3.1. CJEU JIF classification

Subsequent to the extraction methodology, the Claude 3.7 Sonnet model was employed to classify the JIFs according to the VAT ontology. A dedicated phase of prompt engineering was conducted to optimize the model's classification

performance. This involved exploring various representations of the ontology within the prompt, including an ASCII taxonomy, simple indentation, non-indented lists, and indentation with indices. Among the strategies evaluated, the most effective was found to be indentation with hyphens ('-') but without numerical indices, which would, in some cases, interfere with the model's ability to output the exact labels. To further enhance efficiency and minimize API requests, the subsequent task of citation extraction from the JIFs was concurrently performed within the same prompt.

Following experimentation with various prompting strategies evaluated by legal experts on a small subset of JIFs, the Chain of Thought (CoT) strategy was found to be the most performant, particularly for the combination of the two tasks. This approach directs the model toward a more structured, sequential execution, thereby mitigating the risk of skipping critical analytical steps. To facilitate downstream processing and further constrain the model's output, a structured output in JSON format was also requested. The used prompt can be found in Section 6.1.

3.2. National JIF classification

The national JIF classification was conducted simultaneously with the extraction process (as detailed in Chapter IX) to optimize costs and model resource requests. This integrated methodology is similar to the one previously implemented for the CJEU JIFs. The specific prompt employed to execute this task for the Italian judgments is fully described in Sections 6.2 and 6.3.

The prompt used for the national JIF extraction and classification is structured as a Few-Shot Chain-of-Thoughts (CoT) instruction designed to elicit explicit interpretative statements from legal texts. It begins by providing a clear definition of a JIF, establishing four distinct criteria for identification. Following the definition, the prompt includes four comprehensive examples of JIFs extracted from judgments. Each example provides the raw JIF text, followed by an accompanying explanation that justifies why the text meets the JIF criteria. These examples, provided in the language of the judgment, serve as the "few-shot" component,

training the model on the interpretative task and the required reasoning process. Finally, the prompt specifies a structured output format in JSON. This format requires the model to not only output the extracted JIF text but also to populate specific fields. The overall instruction guides the model to perform both the extraction and the subsequent classification within the same request, minimizing computational steps.

4. Assessing pairwise similarity

Following the extraction and classification of all JIFs, their pairwise similarity was assessed. This task presents a significant methodological challenge, as simple lexical or keyword-based comparisons are inadequate, as they fail to capture the deep semantic meaning and contextual nuances inherent in legal language. Two JIFs may use different terminology yet address the same legal concept, while others might employ similar phrasing to discuss entirely distinct legal issues.

To overcome these limitations, a hybrid methodology was developed for quantifying the pairwise similarity between JIFs. This approach moves beyond surface-level textual analysis by integrating two distinct dimensions: deep semantic similarity derived from advanced Natural Language Processing (NLP) models and structural similarity based on a formal legal ontology. This dual-pronged strategy ensures that our similarity assessment is both semantically rich and contextually grounded in the established structure of VAT law, providing a reliable measure of doctrinal proximity.

4.1. A hybrid methodological framework

This methodology is predicated on the understanding that true legal similarity is a composite concept, reflecting both the linguistic content of a statement and its formal classification within a broader legal domain. Consequently, our approach computes a unified similarity score by synthesizing these two analytical dimensions.

4.1.1. Semantic similarity: Capturing contextual meaning with BERT embeddings

The first dimension of our assessment is focused on semantic content. To capture the contextual nuances of each legal statement, we employ state-of-the-art Bidirectional Encoder Representations from Transformers (BERT [2]) models. Each JIF is converted into a high-dimensional vector, or "embedding", which represents its meaning in a dense numerical format. This process is crucial for handling the multilingual nature of our legal corpus, which includes judgments from the CJEU as well as national courts in Italy, Bulgaria, and Sweden.

To ensure the highest degree of accuracy, we utilize dedicated, language-specific BERT models that have been pre-trained on large corpora of text in each target language. The models employed are:

- Italian: dlicari/Italian-Legal-BERT [3].
- Bulgarian: rmihaylov/bert-base-bg [4].
- Swedish: AI-Nordics/bert-large-swedish-cased [5].
- English: joe32140/ModernBERT-large-msmarco [6].

For JIFs originating from CJEU judgments, we generate embeddings for the English text and for all available official translations in the project's target languages. This multilingual embedding strategy is critical for enabling cross-jurisdictional comparisons, allowing for a direct, language-specific comparison between a national JIF and its conceptual counterpart in CJEU case law. The cosine similarity between

---

[2] V. Sanh, L. Debut, J. Chaumond, T. Wolf, *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*, arXiv preprint arXiv:1910.01108, 2019.

[3] D. Licari, G. Comande et al., *ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law,* in *Ekaw (companion)*, 2022.

[4] R. Mihaylov, *rmihaylov/bert-base-bg* [Hugging Face model – 2024]. Retrieved from https://huggingface.co/rmihaylov/bert-base-bg.

[5] AI-Nordics. AI-Nordics/bert-large-swedish-cased [Hugging Face model – 2021]. Retrieved from https://huggingface.co/AI-Nordics/bert-large-swedish-cased.

[6] Joe32140. joe32140/ModernBERT-large-msmarco [Hugging Face model – 2024]. Retrieved from https://huggingface.co/joe32140/ModernBERT-large-msmarco.

the vector embeddings of two JIFs provides a precise measure of their linguistic and semantic proximity. However, as qualitative assessments by legal experts confirmed, a purely semantic approach can sometimes yield misleading results, grouping JIFs with similar phrasing but different legal applications. This observation highlights the necessity of incorporating a second, structural dimension into our similarity metric.

4.1.2. Structural similarity: Employing a multilingual VAT ontology

The second dimension of this methodology addresses the conceptual relationships between JIFs. To achieve this, we employed the ontology described in Section 2 and the JIFs' classification within it. This classification provides a structural collocation for each JIF within the legal domain. The structural similarity between two JIFs is then calculated based on their relative positions within the ontology. Specifically, we compute the shortest path distance between their assigned classification labels in the hierarchical graph. A shorter path signifies a closer conceptual relationship. This taxonomic component ensures that the final similarity score is sensitive to the formal legal classification of the statements.

4.2. A composite similarity metric for nuanced analysis

To create a single, unified measure of doctrinal proximity, we integrate the semantic and structural similarity scores into a comprehensive composite metric.

Given a pair of JIFs and 2 sets of classification labels:

$$L\left(j_{i\in 1,2}\right) = \left\{l_{ij} \middle| l_{ij}\ is\ a\ taxonomy\ label\ classifying\ j_i\right\},$$

the similarity between j_1 and j_2 can be computed as:

$$sim(j_1, j_2) = \cos Sim\left(emb(j_1), emb(j_2)\right) * \left(\frac{1}{1 + \alpha * meanDist(j_1, j_2)}\right)$$

where:

- emb(j_i) is the BERT-based embedding of JIF j_i;
- α is a weighting parameter, defaulting to 0.5;
- the meanDist metric is defined as the average of the shortest

paths connecting all possible pairs of classification labels (l_{1i}, l_{2j}). If a non-VAT JIF is compared to a VAT JIF, their mean distance will be infinite, thus making them incomparable.

This formula multiplicatively combines the two similarity components. The first term, cosineSimilarity, quantifies the textual similarity based on the BERT embeddings. The second term acts as a penalty factor, reducing the overall score when the JIFs are conceptually distant within the VAT taxonomy.

This composite metric, yielding a score between -1 and 1, ensures that two JIFs are only considered highly similar if they are both linguistically alike and conceptually proximate. This prevents the erroneous grouping of textually similar but legally distinct statements. A rigorous legal assessment of the results informed the implementation of a minimum similarity threshold of 0.4 on our research platform, guaranteeing that only JIFs with a significant and meaningful overlap are presented to the user.

4.3. Optimization measures

The computationally intensive nature of the pairwise similarity assessment, involving a large and growing multilingual corpus, necessitated the implementation of a scalable technical architecture. To achieve this, the optimization strategy was based on pre-computation and exploiting the inherent properties of the metric. Specifically, both the BERT-based embeddings and the pairwise similarity scores for all existing Judicial Interpretative Formulas (JIFs) are calculated only once and stored persistently. Furthermore, the system leverages the symmetrical nature of the similarity metric by storing the results in a symmetrical matrix structure, which minimizes both storage requirements and access time. This technical approach ensures a constant O(1) time complexity for the retrieval of the similarity score between any two already-processed JIFs, while maintaining an efficient O(N) complexity for the insertion and comparison of a new JIF against the existing N documents.

5. Conclusions

The chapter presented a hybrid framework for structuring legal knowledge, focusing on VAT judgments from the CJEU and national courts. Building upon a previous phase that used LLMs to extract JIFs, this chapter first establishes a multilingual ontology. This ontology, based on the EUR-LEX taxonomy and enriched by tax law experts, serves as a structured conceptual framework for classifying JIFs. LLMs were subsequently employed for the classification of the JIFs and the concurrent extraction of legal references, utilizing optimized techniques like CoT prompting to ensure structured, high-quality output.

The core methodological contribution is a composite similarity metric for assessing pairwise doctrinal proximity between JIFs. This metric moves beyond lexical analysis by combining two dimensions: semantic similarity, measured using language-specific BERT embeddings, and structural similarity, calculated as the shortest path distance between JIFs' classification labels within the ontology. This dual-pronged strategy prevents the erroneous grouping of legally distinct statements. The entire system is optimized for scalability, using pre-computation and a symmetrical storage structure, ensuring efficient retrieval for a growing multilingual corpus. This integrated architecture effectively organizes legal statements into a coherent, structured knowledge graph, providing valuable insight into judicial reasoning.

6. Prompts

6.1. CJEU JIF classification and citation extraction

Text of the prompt:

I want a list of labels that can be used to classify the input Judicial Interpretative Formula (JIF) and the legal references contained in the JIF.

The input JIF can be classified with at least one label from the attached taxonomy.

Analyse the input text.

Determine some candidate labels (only strictly relevant ones).

You can use the label's ancestors to better understand its meaning, but do not include them in the output.

Determine whether these labels share a parent label, if so discard one.

If one candidate label is parent to another, just keep the child label.

Keep all remaining labels.

Follow the following JSON output format:

{{

'reasoning': [reasoning behind label selection]

'labels': {{ 'A': [most relevant label],

'B': [second most relevant label if present, cannot have the same parent label as A],

more labels if present...

}},

"references": [{{"citation": legal references in the JIF if present , "formatted": [according to one of the following, depending on citation type. If the citation contains more than one referemce, split them into a list [Art. X, Regulation year/number/[EC, EEC or EU] or Art. X, Directive year/number/[EC, EEC or EU] or Art. X, c.c. or Art. X, c.p.c. or Art. X, Law number/year or

Art. X, d.p.r. number/year or Art. X, d.l. number/year or Art. X, d.lgs. number/year or Art. X, r.d. number/year]}}],

}}

Think step by step.

–

Input text:

{jif_text}

Labels:

{labels}

6.2. Italian extraction: Few-shot chain-of-thoughts

Text of the 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 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: Il riconoscimento dell'esenzione richiede, dunque, sotto il profilo soggettivo, che le prestazioni siano rese dai soggetti qualificati enumerati dalla normativa unionale e da quella interna, nonchè, sul piano oggettivo, che esse rientrino nel novero di quelle sanitarie per le quali sia ravvisabile la finalità perseguita dalle norme di esenzione, che è quella di agevolare l'accesso a determinate prestazioni nonchè la fornitura di determinati beni evitando i maggiori costi che deriverebbero dal loro assoggettamento all'iva (Corte giust 20 novembre 2019, causa C400/18, Infohos, punto 37).

Spiegazione: la Corte individua e specifica i requisiti oggettivi e soggettivi per l'applicazione dell'esenzione, alla luce delle ragioni per cui la stessa è stata prevista.

2. JIF: Va dunque affermato il seguente principio di diritto: in tema di IVA, ai sensi del D.P.R. n. 633 del 1972 , art. 13 , comma 1, e art. 73 della Direttiva IVA rifusa (2006/112/CE), nella base imponibile devono ricomprendersi tutti i corrispettivi, compresi eventuali integrazioni, dovuti al cedente o al prestatore di servizi sulla base delle condizioni contrattuali, non potendosi artificiosamente scomporre la prestazione economica, che è tendenzialmente unica ed indissociabile, fatta eccezione per gli elementi normativamente previsti (art. 79 della citata Direttiva).

Spiegazione: la Corte anticipa che sa enunciando un principio di diritto, che è per ciò stesso e sempre un JIF.

3. JIF: Va, quindi, ribadito il principio secondo cui "In tema d'IVA, alla domanda di rimborso non rientrante tra quelle previste dall'art. 30 del D.P.R. n. 633 del 1972, nel testo "pro tempore" vigente, e perciò non contemplata da disposizioni specifiche, si applica l'art. 21, comma 2, del D.Lgs. n. 546 del 1992, di carattere residuale e secondo il quale "la domanda di restituzione, in mancanza di disposizioni specifiche, non può essere presentata dopo due anni dal pagamento ovvero, se posteriore, dal giorno in cui si è verificato il presupposto della restituzione" (Cass., 31 luglio 2019, n. 20573; Cass., 23 ottobre 2015, n. 21674; Cass., 8 giugno 2011, n. 12433).

Spiegazione: la Corte cita esplicitamente un JIF originariamente formulato in una precedente sentenza

4. JIF: Il contenuto della categoria di "scuola riconosciuta" comprende - anche alla luce dell'evoluzione della prassi tributaria in materia - le imprese (enti o privati) in possesso di uno specifico riconoscimento di idoneità oggettiva e/o soggettiva allo svolgimento dell'attività di istruzione da parte della competente pubblica amministrazione, secondo la disciplina volta a volta dettata dalla legge (condizione, quest'ultima, la cui sussistenza è incontestata nella specie).

Spiegazione: la Corte chiarisce quale sia l'interpretazione del termine "scuola riconosciuta" e quale sia l'ampiezza della sua applicazione

Output Format (JSON):

```json

[

{

`jif_text': [JIF exactly how it appears in the original text, including final references if present],

"analysis":{

"reasoning": "text analysis and step-by-step full justification for JIF determination",

"meets_JIF_criteria": [True|False]

},

"references": [{"citation": legal references in the JIF if present , "formatted": according to one of the following, depending on citation type [Art. X, Regulation year/ number/[EC, EEC or EU] or Art. X, Directive year/ number/[EC, EEC or EU] or Art. X, c.c. or Art. X, c.p.c. or Art. X, Law number/year or

Art. X, d.p.r. number/year or Art. X, d.l. number/year or Art. X, d.lgs. number/year or Art. X, r.d. number/year}],

"JIF": [True|False],

"labels":list ol labels that best classify the JIF. Only choose the most relevant ones and the most specific ones that are applicable

},...

]

```

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

–

Labels: [indented list of labels]

–

judgment: []

6.3. Swedish extraction: Few-shot chain-of-thoughts

The prompt used for the extraction and classification of Swedish JIFs equals the one used on the Italian judgments. The difference between the two prompts lies in the provided examples. In this case the examples were the following:

1. JIF: I R 2005 ref. 81 konstaterade Hšgsta fšrvaltningsdomstolen att det inte fanns nŒgot lagstšd fšr att som mervŠrdesskatt behandla belopp som felaktigt angetts avse sŒdan skatt. Ett fšretag som debiterat och betalat mervŠrdesskatt fšr en omsŠttning som inte var skattepliktig hade dŠrfšr rŠtt till Œterbetalning av inbetalda belopp (jfr EU-domstolens avgšrande i de fšrenade mŒlen C-78/02ÐC-80/02, Karageorgou m.fl., p. 41, 42 och 53). Av rŠttsfallet fŒr anses fšlja att inte heller ett belopp som har berŠknats efter en fšr hšg skattesats ska behandlas som mervŠrdesskatt.
Analysis: Texten innehåller en tydlig tolkning av rättsliga

principer gällande mervärdesskattebehandling baserad på ett specifikt prejudikat (R 2005 ref. 81) och EU-rättspraxis. Den fastställer hur felaktigt debiterad mervärdesskatt ska behandlas och utvidgar denna tolkning till belopp som beräknats med felaktiga skattesatser.

2. *JIF: Beskattningsunderlaget fšr leverans av en vara ska utgšras av allt det som sŠljaren har erhŒllit eller ska erhŒlla fšr varan frŒn kšparen. Det belopp som felaktigt behandlats som mervŠrdesskatt och som kunderna betalat till bolaget fŒr anses ingŒ i ersŠttningen fšr de tillhandahŒllna produkterna. Beloppet ska sŒledes, i den mŒn det inte Œterbetalas till kunderna, rŠknas in i beskattningsunderlaget.*

Analysis: Texten ger en tolkning av vad som utgör beskattningsunderlaget för leverans av varor enligt mervärdesskattelagstiftningen. Den tolkar specifikt hur felaktigt debiterade belopp som betecknats som moms ska behandlas juridiskt - och klargör att sådana belopp, om de inte återbetalas till kunderna, måste inkluderas i beskattningsunderlaget. Detta representerar en rättslig tolkning av skattelagstiftningen som fastställer en princip för tillämpning i liknande fall.

3. JIF: *21. Om användaren väljer att betala ett högre belopp än minimibeloppet får den överskjutande delen anses vara en frivillig betalning utan direkt samband med den tillhandahållna dejtingtjänsten, vilket innebär att den delen inte ska beaktas vid beräkningen av beskattningsunderlaget (jfr Tolsma, C-16/93, EU:C:1994:80, punkt 19 och HFD 2019 ref. 68, punkterna 25 och 26).*

Analysis: Texten ger en specifik tolkning av hur betalningar som överstiger minimibelopp ska behandlas i förhållande till beräkningar av beskattningsunderlaget. Den fastställer att sådana överskjutande belopp bör betraktas som frivilliga betalningar utan direkt koppling till den tillhandahållna tjänsten och därför bör uteslutas från beskattningsunderlaget. Tolkningen stöds av specifika hänvisningar till rättspraxis (Tolsma-målet från EU-domstolen och HFD 2019 ref. 68), vilket gör det till en tydlig rättslig tolkning av skatteprinciper som kan tillämpas på liknande situationer.

Chapter XII

ACCESS TO KNOWLEDGE: STRUCTURE AND IMPORTANCE OF THE POLINE PILOT TOOL

Boycho Georgiev – Hristo Konstantinov – Vasil Oreshenski*

1. Introduction

The confluence of artificial intelligence (AI) and big data is catalysing a profound metamorphosis within the legal sphere, transitioning from traditional, text-based research paradigms to data-driven legal analytics. This shift offers unprecedented opportunities to enhance legal cognition, streamline judicial processes, and democratise access to justice. It is within this

* Boycho Georgiev is the GDPR Team Leader at Apis Europe; Hristo Konstantinov is Managing Director of Apis Europe; Vasil Oreshenski is Lead Software Engineer at Apis Europe.

dynamic context that the POLINE project (Principles of Law in National and European VAT), funded by the European Union's Justice Program and coordinated by the University of Bologna, finds its intellectual and practical impetus.

The project's primary objective is to develop and pilot an advanced, AI-powered online platform for the retrieval, analysis, and comparison of Judicial Interpretative Formulas (JIFs) concerning Value Added Tax (VAT) law. For the purposes of POLINE project, we define a JIF as a specific portion of a judicial decision that contains the generalisation of the interpretation and application of a legal rule to a concrete case. These formulas, often repeated across multiple judgments, function as interpretative standards, creating a de facto jurisprudential framework that guides future decisions in similar cases.

By leveraging state-of-the-art techniques in Natural Language Processing (NLP), machine learning (ML), and legal ontology engineering, the POLINE Pilot Tool aims to provide a concrete and trustworthy use case for AI in the judiciary. It seeks not only to advance the state of the art in legal analytics but also to promote more efficient tax procedures, increase the coherence in the application of EU and national law, and ultimately facilitate more equitable access to justice for all stakeholders.

This article will elucidate the structure of the innovative pilot tool developed within the framework of POLINE project and underscore its profound importance for the European legal community.

2. The architecture of the POLINE pilot tool: A modular approach to legal knowledge

The POLINE platform [1] is engineered around a central, groundbreaking concept: treating the Judicial Interpretative Formula as the primary semantic and structural unit of legal knowledge. This represents a significant departure from conventional legal databases, which are fundamentally document-centric. By atomising judicial reasoning into these

[1] POLINE Pilot Tool is freely available at https://www.poline-tool.eu/.

essential, reusable formulas, the tool offers a granular and conceptually rich mode of interaction with case law. The architecture is predicated upon a tripartite modular design, where each module serves a distinct yet interconnected function in the knowledge discovery process.

2.1. From "Judicial Principle of Law" to "Judicial Interpretative Formula": A conceptual refinement

Prior to engaging with the platform's modules, it is necessary to articulate a methodological clarification. In the initial phases of the project, we employed the term "Judicial Principle of Law" (JPOL) to delineate the interpretative statements targeted for extraction. However, our extensive theoretical and comparative legal analysis revealed that this terminology risked conceptual ambiguity. It could be easily conflated with the broader notion of general principles of law, such as legal certainty, subsidiarity, or proportionality, which have a different legal standing and function.

To ensure a more precise alignment with the project's conceptual and methodological aims, and to reflect the specific phenomenon of recurring interpretative statements observed across EU and national case law, the term was refined. The project thus adopted *"Judicial Interpretative Formula" (JIF)* as the conventional concept. It reflects the stylistic approach, which is increasingly mirrored in national high courts, that generates a corpus of standardised interpretative rules. This concept, aligned with existing legal scholarship on the drafting style of the CJEU, provides a more precise and methodologically sound foundation for our analytical and computational work. The adopted new concept underscores the focus on the formulaic, often verbatim, recurrence of judicial statements that act as interpretative standards, departing from common law traditions of "holding" or "dictum" to better suit the civil law context and the harmonised nature of EU VAT law. This refinement allows the POLINE Tool to concentrate on extracting and analysing these specific, instrumental segments of judicial reasoning.

2.2. The Legal Database Module: The semantic heart of the platform

The Legal Database Module serves as the primary gateway to the POLINE platform. It houses a multilingual, curated collection of over a 4,000 JIFs, extracted from the VAT jurisprudence of the CJEU and the supreme courts of three Member States: Italy, Bulgaria, and Sweden, specifically focusing on the VAT subdomains of "Exemptions" and "Taxable Amount". A significant portion of JIFs in this collection were automatically extracted using advanced machine learning and NLP techniques, trained on expert-annotated data.

Unlike traditional systems that offer full-text search across entire documents, this module provides a multi-faceted and structured search experience. Users can initiate queries via keywords but are empowered by a sophisticated filtering panel on the left of the interface. This allows for the refinement of search results based on a range of metadata attributes assigned to each JIF. Beyond the JIFs themselves, the Legal Database Module provides direct links to the full texts of the court decisions from which these formulas were extracted. This allows users to delve into the broader context of the JIF, examining the factual background and judicial argumentation that led to its formulation.

Furthermore, the module integrates essential EU and national VAT legislation, offering users a holistic view of the legal framework. Key instruments like the VAT Directive (2006/112/EC) and relevant national acts are readily accessible, with provisions for both current and repealed legislation where historical context is crucial for understanding current jurisprudence.

The user interface of the Legal Database Module is designed for intuitive navigation and sophisticated search. It features a command panel for accessing modules, language selection, and search functionalities, complemented by a filter panel. This panel enables users to refine search results based on jurisdiction (EU, Bulgaria, Italy, Sweden) and annotation source (AI or legal expert).

The true semantic backbone of this module, however, is the POLINE VAT Ontology. Recognising the limitations of existing taxonomies, which are often based on legislative terminology,

the project developed a bespoke, multilingual ontology. This knowledge structure integrates the established EUR-Lex classification system with a rich layer of keywords extracted directly from CJEU case law. This hybrid approach ensures that the classification scheme reflects not only the letter of the law but also its evolving judicial interpretation. The ontology is presented to the user through a hierarchical browsing component, enabling navigation through complex legal concepts and facilitating searches that transcend simple lexical matching, moving towards genuine conceptual retrieval.

2.3. The Link Visualisation Module: Unveiling connections in the judicial web

If the Legal Database Module is the repository of knowledge, the Link Visualisation Module is the engine for its analysis and discovery. This module transforms a static collection of JIFs into a dynamic network of interconnected legal reasoning, serving two primary functions.

First, it enables the exploration of conceptual similarity. Upon selecting a specific JIF, the module employs advanced textual and semantic similarity algorithms to identify and display a list of related JIFs. This functionality operates both within a single jurisdiction – revealing the evolution or consolidation of a particular line of reasoning – and across different legal systems. This cross-jurisdictional comparison is invaluable for researchers and judges engaged in comparative law, fostering a deeper understanding of the "dialogue between courts" and the process of legal harmonisation.

Second, the module provides a visual representation of jurisprudential networks. It maps the explicit citations and implicit semantic relationships between decisions relevant to the JIF in question. This feature allows users to trace the lineage of a legal interpretation, identify seminal cases, and understand how different judicial precedents influence one another. It moves beyond a simple list of results to a navigable map of legal influence, providing powerful insights into the structure and dynamics of case law.

2.4. The Customised Detection Module: An interactive tool for legal assessment

The Customised Detection Module is an interactive and practical component designed to support legal professionals and citizens directly. Its purpose is to bridge the gap between the vast corpus of established JIFs and a specific legal document at hand.

Users can upload a judicial decision into the system. The module then processes the document, identifying passages that correspond to known JIFs within the POLINE database. Thus, it presents a customised legal assessment, highlighting the relevant JIFs present in the text and providing an immediate overview of the interpretative standards applied or referenced. Furthermore, the module can evaluate the correspondence of these identified principles with the most current or relevant JIFs at both national and EU levels. This functionality is particularly beneficial for taxpayers, enhancing their awareness of the correct application of VAT law and empowering them to make informed decisions regarding appeals or claims for violations of EU law. By facilitating an active role in the enforcement of VAT law and the pursuit of fair taxation, this module contributes directly to strengthening access to justice.

3. The significance of the POLINE platform: Fostering access, coherence, and trustworthy AI

The value of the POLINE Pilot Tool extends far beyond its technical architecture. Its importance lies in its potential to address systemic challenges within the European legal ecosystem, from judicial workload to the transparency of law.

3.1. The JIF as a paradigm shift in legal analytics

The most profound innovation of the POLINE Pilot Tool is its conceptual reorientation of legal information retrieval. By establishing the Judicial Interpretative Formula as the core unit of analysis, the platform pioneers a concept-centric rather than a document-centric approach. This is the first time a legal analytics application has been architected around this granular element of judicial reasoning. This paradigm shift enables a

more precise, coherent, and semantically meaningful exploration of case law, allowing users to engage directly with the building blocks of legal interpretation, independent of the voluminous texts in which they are embedded. This innovation provides a powerful new lens through which to study the harmonisation of EU law and examine the evolution which the EU's judicial system in VAT domain has undergone.

3.2. Empowering legal professionals and enhancing judicial efficiency

The primary target group for the POLINE tool is the national judiciary. Judges in tax matters face immense caseloads and the complex task of aligning national law with the ever-expanding body of CJEU jurisprudence. The platform is designed to significantly reduce their workload in retrieving and evaluating relevant case law. By providing immediate, one-stop access to an organised set of JIFs, it can expedite legal research and enhance decision-making consistency.

In particular, it will facilitate a more informed application of the *acte clair* and *acte éclairé* doctrines, potentially streamlining the process of preliminary references to the CJEU and thereby reducing its workload as well. For lawyers, tax administrations, and academics, the tool offers an unprecedented resource for litigation, consulting, and research, enriching their knowledge of both national and European VAT frameworks.

3.3. Coherence, harmonisation, and dialogue of courts

Systematic exposure to EU and national JIFs and their similarity networks supports consistent interpretation across jurisdictions and over time. The ontology contributes a shared semantic layer that aligns tax vocabularies and concepts to harmonised VAT law. By visualising cross system analogies and divergences, the Pilot Tool fosters dialogue between courts on harmonised grounds, advancing both doctrinal clarity and practical coordination.

3.4. Democratising access to justice and promoting fair taxation

A foundational principle of the POLINE project is its commitment to open access. The platform is freely available to all user groups, a deliberate choice that stands in contrast to the often-prohibitive cost of commercial legal tech solutions. This democratisation of access is a critical step towards empowering citizens, taxpayers' associations, and smaller legal practices with a sophisticated analytical tool. By enabling taxpayers to better understand the judicial reasoning applied in their cases, the platform enhances the transparency of the tax system and supports the effective protection of their rights. Moreover, by raising awareness of JIFs related to specific areas, such as favourable tax regimes for care-related services – costs often disproportionately borne by women – the project can contribute to greater awareness of gender equality issues within the tax system.

3.5. Advancing trustworthy AI in the judiciary

The deployment of AI in the justice sector rightly calls for rigorous ethical scrutiny. The POLINE project is committed to the paradigm of trustworthy AI, adhering to the principles outlined in the European ethical Charter on the use of AI in judicial systems and their environment [2]. The system is not designed to predict outcomes or replace human judgment, but to augment legal cognition. The human-in-the-loop methodology, where expert annotations are used to train and validate the ML models, ensures a high degree of accuracy and legal soundness. The platform's transparency – always linking JIFs back to their original context – allows for constant human verification. Finally, the decision to release the software architecture under the European Union Public

[2] The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe adopted in December 2018 the first European text setting out ethical principles relating to the use of artificial intelligence (AI) in judicial systems. See: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment. The full text of the Charter is available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

License (EUPL) ensures its sustainability, promotes further research, and allows other institutions to build upon our work, fostering a collaborative and open ecosystem for e-justice technologies.

4. Conclusion

The POLINE Pilot Tool represents more than a sophisticated search engine. The online platform is an integrated environment for the discovery, analysis, and understanding of judicial knowledge. Its innovation lies not only in the application of advanced AI techniques but, more fundamentally, in its conceptual framework centred on the Judicial Interpretative Formula. By structuring legal information around these pivotal units of reasoning, the platform offers a novel and powerful way to navigate the complexities of European VAT case law.

Through its modular design, the tool provides a comprehensive suite of functionalities, from structured retrieval and network visualisation to customised document analysis. Its commitment to free and open access underscores a mission to democratise legal knowledge and enhance the transparency and coherence of the European legal order. Thus, we believe the POLINE platform provides a robust and ethical blueprint for the future of AI in the service of justice, contributing to a more efficient, consistent, and accessible legal system for all.

Conclusion

THE EMERGING REALITY OF FORMULAS: CONCLUDING REMARKS

Alessia Fidelangeli – Piera Santin*

Our work aligns with the most recent developments in legal informatics, which do not seek to replace judges, but rather to provide tools supporting their interpretative tasks. Conscious of the concerns raised by the growth of AI, in POLINE we have developed methods that assist interpretation by legal professionals, rather than substituting them. However, the introduction of AI into the legal domain entails a shift in the interpretative activity, affecting even the preliminary phase that precedes what we traditionally define as "interpretation" (Ch. II).

In our case, this process involved the analysis of case law and legal doctrine across four jurisdictions, the identification of common definitions and features of JIFs, their transposition into annotation guidelines, and finally, prompt engineering. As for the first stage (identifying definitions and features) every act of selection and definition is a form of interpretation and entails a subjective choice by the legal experts involved. This inevitably influence subsequent steps, including the extraction phase. Yet this is not a problem unique to AI: interpretative subjectivity is intrinsic to legal reasoning. In our case the initial selection was performed by the research team. The multidisciplinary composition of this team, combining expertise in tax law, philosophy of law, legal informatics, and

* Alessia Fidelangeli is Research Fellow at the University of Bologna; Piera Santin is Research Associate at the Robert Schuman Centre for Advanced Studies of the European University Institute.

computer science, helped mitigate the risks inherent in this choice.

One of our aims was to test whether legal informatics could help reveal the structural strength of the concept of JIF, whose meaning is somehow blurred in existing literature, and the project has indeed confirmed it. Two aspects proved crucial. First, the iterative process of drafting annotation guidelines and conducting error analyses forced us to clarify our conceptual choices. Second, the need to automate the process required a high degree of consistency, which brought to satisfactory results: the automatic extraction performed well when confronted with manual annotation (Ch. IX). Our findings confirm that JIFs are not only frequent but also increasingly relevant in the field of VAT, with recent judgments having a higher number of these formulas than older ones (Ch. III).

The analysis of the CJEU case law and the extracted JIFs further confirms that the CJEU "cites to confirm and writes to be cited", thereby reinforcing the coherence of its jurisprudence. Unsurprisingly, the European legal system is the one in which the extraction of JIFs works best because of the larger number of JIFs available, reflecting the Court's interpretative role, which entails that it uses JIFs most frequently than other courts (Ch. IV).

Regarding national case law, the analysis demonstrate that tax systems still differ considerably across jurisdictions. Nevertheless, two tendencies are common to all of them: (i) the growing use of formulaic expressions in judicial reasoning, and (ii) the influence of EU case law, including CJEU JIFs, on national decisions (Ch. V, VI, and VII).

Frequently, national JIFs reproduce EU ones. The POLINE tool captures this dynamic, evidencing how EU jurisprudence influences not only national reasoning but also drafting styles. This is particularly visible in Bulgarian case law, despite the Administrative Supreme Court decides also on the merit (Ch. VI), and in the Italian Court of Cassation, whose "principi di diritto" can be regarded as a subcategory of JIFs (Ch. V). This is coherent with the fact that across all the examined systems, there is an expansion of the creative role of judges and of the importance of precedent (even if not formally binding), as illustrated by the Italian notion of "diritto vivente" (Ch. V).

In the Introduction, we noted that contemporary legal

practice increasingly relies on tools that extract portions, summaries, or key statements from decisions. While this facilitates access to large corpora of case law, it also raises the risk of decontextualised interpretation. In our project, we addressed this concern by ensuring that each JIF in the POLINE platform maintains a dynamic link to the full text of the decision. In future developments, reinforcing the network structure and enriching metadata, including references to factual contexts, could further reduce these risks.

In more general terms, the existence of a platform capable of visualising JIFs, their interconnections, and their associations with legal concepts is crucial. These platforms achieve their full potential when they are open and easily accessible. The POLINE platform was built paying attention to implement an effective retrieval system, an essential component in navigating the corpus of case law which is often underestimated (Ch. XII).

From a methodological perspective, one of the most relevant findings concerns prompt engineering. After extensive manual work defining features of JIFs and guidelines for their annotation and extraction, we observed that the most effective prompts were not always the most detailed. This may surprise lawyers but is crucial when managing large datasets (Ch. IX). Moreover, given identical definitions and prompts, results can vary significantly across models. This confirms that consistency in model selection and targeted prompt engineering are essential.

Even when using LLMs, substantial human work remains indispensable, even if this is shifted from the annotation phase to validation and supervision. Nonetheless, the manual work in the validation phase enhances scalability, making it possible to extend our approach to larger datasets or different legal domains with small adaptation.

In addition to LLMs, POLINE also relied on NLP as, only through it, results become replicable over time without further LLMs extractions to be peformed with additional costs. In our case NLP allowed the creation of the customised detection module. Our experience with NLP has demonstrated that reliable results depend on robust methodology: manual annotation, double-blind validation, and detailed guidelines (Ch. X).

As emphasised since the beginning, knowledge is of little

use without structured organisation. The manual structuring of large-scale legal materials would have been unfeasible without computational tools. Nonetheless, ontology creation still requires significant expert input, revealing the centrality of human expertise. In this regard, the method that brought the most effective connection between JIFs and ontology concepts was the "chain-of-thought" approach. Interestingly, this method is the most closely resembling human reasoning (Ch. XI).

Whether this is coincidental remains an open question that we intend to explore in future research. Further future developments of this work could benefit from the full availability of national case law, which would enhance the effectiveness of the tool. We also plan to deepen our use of network analysis to show the evolution of JIFs over time and to improve users' contextual awareness when navigating interpretative materials.