Dorso
12,63 mm

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

273714_cop_SANTIN (3 BIS).indd 2-3

03/12/25 13:18

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)

Table of Contents

Introduction

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

Piera Santin – Alessia Fidelangeli

Chapter I

ARTIFICIAL INTELLIGENCE FOR TAX COURTS

Marco Billi – Federico Galli – Giuseppe Contissa

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 VIII

THE ADDED VALUE OF JUDICIAL INTERPRETATIVE FORMULAS IN DISCOVERING JUDICIAL INTERACTIONS AMONG EUROPEAN AND ITALIAN COURTS

Federica Casarosa – Madalina Moraru

CHAPTER IX

FROM TEXT TO KNOWLEDGE: AUTOMATIC ANNOTATIONS THROUGH LARGE LANGUAGE MODELS

Rachele Mignone – Davide Audrito – Ivan Spada – Luigi Di Caro

CHAPTER X

EXTRACTING KNOWLEDGE: AUTOMATIC EXTRACTION THROUGH NATURAL LANGUAGE PROCESSING

Giulia Grundler – Andrea Galassi – Federico Ruggeri – Paolo Torroni

CHAPTER XI

STRUCTURING KNOWLEDGE: ONTOLOGY-BASED APPROACHES TO SIMILARITY AND CLASSIFICATION

Rachele Mignone – Davide Audrito – Ivan Spada – Luigi Di Caro

Chapter XII

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

Boycho Georgiev – Hristo Konstantinov – Vasil Oreshenski

CONCLUSION

THE EMERGING REALITY OF FORMULAS: CONCLUDING REMARKS

Alessia Fidelangeli – Piera Santin

CHAPTER I

ARTIFICIAL INTELLIGENCE FOR TAX COURTS

Marco Billi – Federico Galli – Giuseppe Contissa*

CONTENTS:

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, Arg-tuProlog 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.