“Lost in EU Regulation? Don’t Worry, AI Found the Obligation”

Extracting and Representing Legal Obligations in the GDPR the DSA and the AI Act

Thiago Dal Pont
Alma Mater Research Institute for Human-Centered Artificial Intelligence (AlmaAI)
University of Bologna
Bologna, Italy
thiago.raulino@unibo.it

Federico Galli
Department of Legal Studies
University of Bologna
Bologna, Italy
federico.galli7@unibo.it

Galileo Sartor
Department of Computer Science
Swansea University
Swansea, United Kingdom
galileo.sartor@swansea.ac.uk

Giuseppe Contissa
Deparment of Legal Studies
University of Bologna
Bologna, Italy
giuseppe.contissa@unibo.it

Abstract

Compliance with European Union (EU) digital regulations is challenging due to their increased complexity. This study develops and evaluates an AI-driven pipeline to automate the extraction and structuring of deontic obligations. Our framework, grounded in legal theory, applies classical Natural Language Processing (NLP) approaches in conjunction with Large Language Models (LLMs) to three major EU regulations: General Data Protection Regulation (GDPR), Digital Services Act (DSA), and AI Act (AIA) in a structured workflow. The workflow accurately identified and structured obligations, with strong performance on the AIA but challenges in the GDPR and DSA. Knowledge graphs improved the organisation and cross-referencing of obligations. Overall, the results indicate that LLMs significantly enhance the extraction and organisation of obligations. The research contributes to AI-driven legal compliance by facilitating efficient regulatory navigation and supporting the development of automated compliance tools for organisations, legal professionals, and policy-makers.

CCS Concepts

• **Applied computing** → **Law**; • **Computing methodologies** → **Natural language processing**; *Knowledge representation and reasoning*; • **Information systems** → *Information extraction.*

Keywords

Obligation Extraction, Large Language Models, Knowledge Graphs, Legal Compliance, EU Digital Law

ACM Reference Format:
Thiago Dal Pont, Federico Galli, Galileo Sartor, and Giuseppe Contissa. 2025. “Lost in EU Regulation? Don’t Worry, AI Found the Obligation”: Extracting and Representing Legal Obligations in the GDPR the DSA and the AI Act. In *20th International Conference on Artificial Intelligence and Law (ICAIL 2025), June 16–20, 2025, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3769126.3769260

ACM ISBN 979-8-4007-1939-4/25/06
https://doi.org/10.1145/3769126.3769260

1 Introduction

The rapid development of Artificial Intelligence (AI) and natural language processing (NLP) is reshaping the landscape of legal compliance, providing effective tools to manage, interpret, and operationalise complex regulatory frameworks. Among these, Large Language Models (LLMs) have been developed as advanced technologies, exhibiting a strong capability to process and generate legal text with high accuracy [14]. They have significant potential in regulatory compliance, particularly for tasks that involve analysing dense legal language and organising complex knowledge [8, 13].

In the European Union (EU), regulatory compliance has recently increased significantly, particularly in the area of digital technologies. Key regulations such as the General Data Protection Regulation, Data Act, Data Governance Act, Digital Services Act, Digital Markets Act, Cyber Resilience Act and Artificial Intelligence Act exemplify this trend. Altogether, these regulations constitute an emerging legal order [23], characterised by interconnected concepts, overlapping provisions, and similar regulatory approaches [20] aimed at governing various aspects of the digital society.

Digital operators face great challenges in managing and complying with this new intricate framework [17]. The texts of the regulations are particularly complex due to their dense, techno-legalistic language, substantial length, and the wide array of stakeholders they affect, including both EU and non-EU industry actors, national governments, administrative authorities, and EU institutions [6].

Grasping this framework is made even more difficult by linguistic ambiguities, inconsistencies in phrasing, and the interwoven nature of provisions across multiple regulations. For instance, compliance obligations are often scattered across numerous provisions, conditioned by implicit requirements, or dependent on cross-referencing related regulatory texts, national implementations, and supplementary guidance documents, creating a fragmented and highly complex regulatory landscape.

This complexity underscores the urgent need for innovative solutions, such as automated tools, to simplify compliance processes. These tools may help regulated entities engage in more efficient and effective compliance-oriented readings of regulatory texts, reducing the burden and risk of non-compliance.

In this context, this research aims to bridge the gap between the increasing demand for efficient compliance processes and recent developments in automated compliance technologies. Specifically, this study explores how advancements in NLP and LLMs can be used to develop an extraction and analysis pipeline for structuring and representing legal obligations from regulatory texts through knowledge graphs. Our experiments focus on three key EU digital regulations: the GDPR, the DSA, and the AIA. The relevance of this research is manifold. From a technical point of view, our approach demonstrates how LLMs surpass traditional methods by allowing abstraction from manual annotation, enabling the consideration of broader context, and offering flexibility in handling complex legal language. In addition, this research may open up opportunities for further technological development, such as the integration of expert systems for compliance support, logical representations of norms for automated reasoning, and the application of network analysis to uncover interconnections and dependencies among legal obligations across regulatory texts.

From a practical standpoint, the automated extraction and structured representation of legal obligations provide regulated entities with tools to efficiently navigate complex regulatory frameworks. These tools can help organisations identify their obligations, streamline compliance efforts, and reduce the risk of non-compliance by offering precise and actionable insights tailored to their roles and responsibilities. In the long run, this research may offer valuable insights into the structure and interplay of EU (digital) regulations. This, in turn, can inform future legislative design and harmonisation efforts, helping to address inconsistencies and improve regulatory coherence.

The paper begins by situating the research within the broader context of AI and law, providing an overview of related works that focus on automatically identifying deontic statements, such as obligations, in regulatory texts (Section 2). Next, we introduce our framework, rooted in analytical legal theory, which defines how we analyse and represent legal obligations within legal texts (Section 3). Following this, we describe our proposed workflow for extracting, analysing, and representing legal obligations (Section 4). This workflow includes steps for obligation detection, deontic obligation filtering, deontic obligation analysis, and deontic obligation representation. We then present how this workflow was implemented in our case study, which examines the GDPR, the DSA, and the AIA (Section 5). This section covers the employed model, the validation approach, the experimental results, a discussion of the findings, and the resulting knowledge graphs. Finally, we summarise the key findings and propose future research directions and implementation possibilities (Section 6).

2 Related Works

Traditionally, legal professionals and compliance teams have relied on manual reading and interpretation of normative texts to extract legal obligations. These methods often involve the careful reading and analysis of dense and complex documents, which is time-consuming and prone to inconsistencies.

Semi-automated tools, such as keyword search systems and rule-based approaches, have been developed to assist in this process. For example, [11] introduced Gaius T., a tool based on the Cerno framework [12], which uses context-free grammars and semantic annotations to extract rights, obligations, actors, and constraints. While effective in certain scenarios, such tools often struggle to accurately link constraints to subjects in complex regulations.

Another example is the work by [27], who proposed a rule-based method for extracting conditional and deontic rules using the GATE framework and Stanford Parser. This approach was instrumental in identifying critical legal components, such as antecedents and agents, but remained constrained by the rigidity of rule-based systems. Similarly, [2] combined syntax- and logic-based rule extraction with lightweight ontologies, achieving high precision and recall in testing on the Australian Telecommunications Consumer Protections Code. However, these methods are limited by their reliance on predefined syntactic rules that may not appear the same in other legal texts.

The uptake of machine learning approaches has introduced more sophisticated techniques to obligation extraction, beginning with machine learning techniques. [25] explored active learning methods to classify deontic statements in German tenancy law, using multinomial naive Bayes, logistic regression, and multi-layer perceptron classifiers. This study highlighted the potential of machine learning to handle large corpora, although it required substantial annotated data for training.

Other researchers, such as [5], focused on combining syntactic and logic-based methods to identify six types of normative relationships, including prohibitions, authorisations, and sanctions. These methods have shown promise in addressing the limitations of rule-based systems, particularly for more nuanced regulatory texts. However, they still often depend on domain-specific training data and are constrained in their ability to generalise across legal systems and languages.

The advent of LLMs has transformed the landscape of legal text analysis. LLMs like GPT, BERT, and domain-specific models such as LegalBERT have demonstrated remarkable capabilities in processing, interpreting, and generating human-like legal language. [18] pioneered the application of neural networks for classifying deontic modalities in financial regulations, focusing on obligations, prohibitions, and permissions. More recently, [15] explored the use of GPT-3 for Legal Rule Classification (LRC), fine-tuning it on a GDPR-specific dataset. The study integrated symbolic knowledge from legal standards such as LegalDocML and LegalRuleML, achieving state-of-the-art performance with a weighted F1 score of 0.93, even with limited training data. This highlights the potential of LLMs to outperform traditional methods, particularly in scenarios requiring contextual understanding and generalisation.

This paper builds on these advancements by addressing key gaps in the existing literature and proposing a novel approach that leverages LLMs to extract and analyse legal obligations from regulatory text and represent them through knowledge graphs.

Unlike traditional rule-based systems, which rely on predefined rules and struggle to generalise across diverse legal texts and constructs, our approach employs LLMs, which are capable of capturing nuanced context and generalising effectively across varied regulatory frameworks.

Similarly, while earlier studies demonstrated high precision in obligation extraction, they often relied on extensive manual annotation to train models, which limited scalability. In contrast, we address this limitation by leveraging pre-trained LLMs, reducing the dependency on expensive and time-consuming annotation processes while maintaining high accuracy in identifying and structuring obligations.

Furthermore, our framework extends beyond obligation identification by introducing structured analysis rooted in legal theory, which conceptualises obligations as deontic constructs with recurring elements (e.g. addresses, beneficiaries, conditions, etc.). In compliance, these elements are essential to understand the scope of applicability of an obligation and to determine what actions must be taken to fulfil it.

Finally, our approach includes a representation of these obligations through knowledge graphs. These graphs enable the visualisation of relationships between obligations, actors, and conditions, supporting a more intuitive navigation of regulatory texts. This structured approach also facilitates advanced functionalities, such as cross-referencing provisions across regulations, network analysis of norms to understand their interdependencies, and logic-based reasoning to identify overlaps, redundancies, or potential conflicts.

3 Theoretical Framework

Our approach to extract and represent deontic obligations in legal texts draws on legal theory and analytical philosophy (e.g., [7, 9, 10]). In particular, our framework considers the following elements: (0) deontic modality, (1) addressee, (2) predicate, (3) target, (4) specifications, (5) pre-conditions, (6) beneficiary.

We acknowledge that other important elements may define the normative content of obligations. These include, among others, temporal constraints, which establish the time frames within which obligations must be fulfilled, and logical connectors, which clarify the relationships between different components of an obligation (e.g., "AND", "OR", "IF-THEN"). Modelling and extracting this information is particularly challenging in the context of our framework [16, 19].

As detailed below, the proposed framework is designed to be modular and, in the future, can be easily extended with other elements of deontic obligations. Also, the framework is not intended as a new formal model for deontic statements, although its output could be used to represent obligations as logic rules, XML, or other existing formal models.

In our framework, elements can be mandatory or optional. The only mandatory element is the predicate that specifies the content of the obligation, namely a state of affairs to be realised or an action to be performed. The other elements are optional. Note that the addressee is not a mandatory element in our representation as it is not always explicit in legal texts (though, arguably, it is always implicitly identifiable through contextual reading or legal interpretation).

In general, the elements of an obligation may be explicitly defined in the text or implicit, requiring interpretation or contextual analysis to identify their presence. For instance, even if an obligation does not explicitly mention a beneficiary, this element can often be deduced from related provisions or the broader intent of the Regulation.

Deontic Modality. Obligations in English-written legal texts are typically conveyed through modal verbs such as "shall", "must", "ought to" or "has/have to". For instance, EU regulation usually employs "shall" to express mandatory actions or states, typically prescribing what must or must not be done in legal and regulatory contexts [3, 22].

At the same time, markers, such as "shall", may exhibit a high level of semantic ambiguity. For example, research in legal-linguistic has noted that regulatory statements introduced by "shall" may refer to different meanings [1, 4]. For example, "shall" may be used to provide a definition or explanation of a certain concept or entity, as in "any reference to an economic operator under Regulation (EU) 2019/1020 *shall* be understood as including all operators identified in Article 2(1) of this Regulation" (Article 74 of the AIA). Other times, "shall" may be used as a constitutive statement, as in "The Digital Services Coordinator *shall* be responsible for all matters relating to supervision and enforcement of this Regulation in that Member State" (Article 49 of the DSA). Disambiguating such instances is particularly relevant to extracting deontic obligations correctly. For this reason, as we will detail in Section 4.2, we have implemented an obligation filtering step to isolate genuine deontic obligations.

Our concept of an obligation also includes prohibitions since we assume that the prohibition of a certain action amounts to the obligation not to perform that action. For example, from the prohibition of providing or using AI systems to manipulate persons and create harm (Article 5(1)(a)), we can derive the obligation not to engage in such AI systems. This correspondence is also reflected at the linguistic level, as most prohibitions are denoted by the negative form of the verbs establishing obligations (e.g., "must not" or "shall not"). In the remainder of the paper, unless stated otherwise, we refer to "obligations" considering both duties and prohibitions.

Finally, our framework draws a distinction between obligations of action and obligations of being [24].

- *Obligations of Action*: These obligations require an action to be taken in order to achieve compliance. The level of specificity for the required action can vary. For example, compare these two provisions: "The AI provider shall implement appropriate safety measures" and "The data controller shall maintain a record of processing activities". Despite the difference in detail, the key requirement remains the same: an action—whether clearly defined or broadly described—is explicitly mandated.
- *Obligations of Being*: These obligations demand that an entity (a thing, a person, an organisation) meets certain requirements. For example, an AI system may be required to be secure by design, which implies meeting security standards but does not specify exact security measures.

This distinction applies equally to prohibitions, as prohibitions can either prevent specific actions or impose constraints on the state of being. For instance, consider a regulation stating that "providers shall not use AI systems to exploit vulnerabilities of specific groups" and that "AI systems shall not contain biases in their design resulting in unfair treatment of protected groups". In the first case, the regulation would introduce an actionable restriction, requiring the avoidance of a specific design practice. In the second case, it would determine a desired state of conformity, imposing a constraint on the attributes of the regulated target (the AI system).

Note that any obligation/prohibition of being can ultimately be translated into one or more obligations/prohibitions of action once the addressees, i.e. the entities responsible for implementing the requirements, are identified. However, legal texts do not always explicitly determine the specific addressees. Such a determination can be drawn contextually upon reading other provisions in the same text or implicitly by using interpretative means.

From this point forward, we will outline the elements composing a legal obligation. It is important to highlight that the function of each element may vary slightly depending on whether the Regulation imposes an obligation to exist in a certain state ("obligation of being") or requires a specific action ("obligation of action").

Predicate. The predicate defines the required state or action associated with the obligation.

In obligations of being, the predicate specifies the requirement that the target is expected to satisfy and is generally denoted by a linking verb (e.g., "shall be resilient").

In obligations of action, the predicate outlines the specific activity required. It is denoted by an action verb and may be in either active or passive form (e.g., "shall conduct", "must be communicated"). Regardless of the form, the predicate conveys the core content of the obligation, clarifying either the state the target must "be" in or the action that must be "done" affecting the target.

Addressee. An addressee is the individual, organisation, or entity responsible for fulfilling the obligation, bearing active responsibility for either realising, maintaining or affecting the target's condition (in obligations of being) or performing the specified action (in obligations of action). The addressee must individually or collectively possess the capacity to act, such as a natural or legal person. Consequently, entities like "AI systems", "a process", or "adopted measures" cannot serve as addressees of obligations.

In obligations of action, the addressee is generally explicitly stated in the regulatory provisions. At the same time, there may be cases where only the target and the predicate (generally expressed in the passive form) are explicit, and the addressee who has to carry out the action is implicit, requiring interpretive analysis to identify who must ensure the specified condition is met. For example, in obligations of action, the responsibility to conduct a particular task (like a risk assessment) is typically assigned directly to an entity such as a "provider" or "organisation".

In contrast, obligations of being require a regulated target to achieve certain standards without directly naming one or more addressees. Consider the provision that "the oversight measures shall be commensurate with the risks" (Article 14 of the AIA). In this case, the Regulation leaves it to the reader to infer who is responsible for ensuring that the oversight measures meet the required standard. This is often based on the broader regulatory context or the nature of the entity most logically associated with the target (i.e., the deployer of the AI system).

Target. The target is the entity associated with the predicate, defining what must be realised, maintained or affected. It is an optional element.

The role of the target differs depending on the obligation type: for obligations of being, the target is something that must meet or uphold a specific state, quality, or condition (e.g., "a system must remain secure"). Here, the target is always the grammatical subject of the sentence.

For obligations of action, the target is the entity that is subject to or concerned by the action and is the grammatical object of the sentence. When the predicate consists of a ditransitive verb (e.g., inform, grant, notify), the target is always the direct object of the sentence, while the indirect object is the beneficiary (see below). For instance, in the statement "the controller shall inform the data subject of any such extension within one month of receipt of the request" (Article 12 of the GDPR), the target is the "[information on] any such extension", while the data subject is the beneficiary.

Specifications. Specifications define the standard, time or methods required to fulfil the obligation. For obligations of being, they typically establish qualitative or quantitative criteria specifying the requirements the target is expected to meet (e.g., "in accordance with the standard procedure"). In obligations of action, specifications may outline the rigour or procedural benchmark necessary to perform the action. Not all obligations are accompanied by specifications; therefore, this element is optional.

Beneficiary. The beneficiary is the party who directly or indirectly benefits from the fulfilment of the obligation.[1] In obligations of being, the beneficiary may benefit from the maintained condition or state of the target, such as enhanced safety or transparency. In obligations of action, the beneficiary derives an advantage from the specific action performed, such as receiving a service, right, or information. While the beneficiary is not always explicitly stated, it is often possible to infer their identity from the context or prior knowledge about the relationships and responsibilities between the involved parties.

In theory, when a beneficiary can be identified, it should always be possible to establish derivative obligations (such as the duty to compensate for harm) that come into effect in the event of a violation of the primary obligation. For now, we do not consider these derivative obligations (also known as "contrary-to-duty obligations" [21]) in our experiments.

Pre-Condition. The pre-condition is another optional element that defines the circumstances or criteria that must be met (or not met) before an obligation becomes enforceable. In both obligations of being and obligations of action, pre-conditions act as triggering factors. Positive pre-conditions require certain conditions to be fulfilled for the obligation to take effect, while negative pre-conditions specify that the obligation applies only if particular conditions are absent.

In our framework, we also consider exceptions as instances of negative pre-conditions. These exceptions can be identified explicitly in the deontic statement or inferred from its context or referenced provisions. However, we acknowledge that not all exceptions can be modelled and identified as negative pre-conditions

[1] An indirect beneficiary refers to a situation where the party benefiting from the obligation differs from the one directly advantaged by its performance. Although this distinction may be relevant in our contexts, we have not implemented this distinction in the prompt so far.

of particular obligations. Exceptions may stem from broader contextual factors or require a cross-referenced interpretation of related provisions.

4 Workflow

Figure 1 shows the proposed workflow to extract, analyse and represent legal obligations from the text of EU digital regulations.[2] The workflow is composed of four modules:

- Module 1: Obligation Detection;
- Module 2: Deontic Obligation Filtering;
- Module 3: Deontic Obligation Analysis;
- Module 4: Deontic Obligation Representation.

Modules 2-3 represent the core of the workflow and leverage LLMs to extract and analyse deontic obligations from legal texts. The following subsections contain an explanation of each module.

4.1 Candidate Obligation Detection

EU digital regulations are retrieved from the EUR-Lex database, i.e., the EU institutional database containing all legal sources of EU law.[3] The database stores the documents on EU legal regulations in XML format, with the text already divided into paragraphs. This structure facilitates the splitting of the documents into smaller, more manageable components for analysis or processing.

After the paragraphs are split, we retain only paragraphs containing "candidate obligations", namely those sentences potentially containing deontic obligations. For this, we rely on keywords commonly associated with obligations in the English version of EU legal texts [22], namely, "shall", "must", "should", "has" and "have to". To identify the presence of one or more of these keywords, we relied on regular expression, looking for the precise terms or variants when available.

Within the paragraphs containing these keywords, we also extract the specific sentence in which the keyword occurred, isolating the portion of text that the LLMs would focus on in subsequent analyses.

The formal XML structure of the document, with its division in paragraphs, is also used as context for the LLMs. In particular, when constructing the prompt, we include the entire article containing the analysed paragraph, along with any explicitly cited paragraphs from the same regulation, as additional contextual information. This context is often essential because it provides important details such as addressees, specifications, or preconditions that are often replaced in the analysed sentence by pronouns (e.g., "This shall...") or references to earlier provisions (e.g., "as stated in the previous paragraph").

Explicitly cited articles of the same regulation are also extracted and included as context. The relevance and impact of these contextual clues is evaluated after the LLM analysis and validation.

4.2 Deontic Obligation Filtering

Within the obligation filtering module, LLMs evaluate each "candidate obligation" to ascertain those that constitute genuine deontic obligations within the specified framework. Although modal verbs such as "shall" or "must" frequently indicate obligations, they may also be employed in definitions, authorisations, or constitutive statements. In order to ensure accuracy, the model is required to determine the semantic function of each sentence and to subsequently classify it into one of the following classes:

(1) *Definition*: The statement provides a definition of an entity or term.
(2) *Constitutive Statement*: The statement creates a new state of affairs or qualifies a fact with a legal effect, rather than prescribing behaviour or imposing obligations.
(3) *Deontic Obligation*: The statement imposes a duty on someone (Obligation of Action) or establishes a requirement on something (Obligation of Being).
(4) *Entitlements*: The statement empowers someone or grants a right.
(5) *Authorisations*: The statement indicates when an action is permitted or authorised under specific conditions.
(6) *Deontic Prohibition*: The statement imposes a prohibition on someone (Obligation of Action) or establishes a negative requirement on something (Obligation of Being).

An additional label, namely "Not Applicable", has been introduced to serve as a fallback category for statements that do not align with any of the defined classes.

The deontic obligation filtering prompt is composed of two elements. Firstly, there is the *system prompt*, the function of which is to define the task and require the model to justify its classifications. Secondly, there is the *user prompt*, the function of which is to provide the sentence to be analysed. Few-shot examples illustrate candidate obligations paired with their expected classifications. As demonstrated in Listing 1, the prompt employs Markdown formatting to ensure consistency with prevalent LLM outputs.

The user prompt (Listing 2) presents a single sentence along with its surrounding paragraph in order to provide context. This sentence-level focus has been demonstrated to enhance classification accuracy. Citations were excluded from the analysis, as preliminary tests indicated that they exerted minimal influence on model performance of the obligation filtering step.

The model's output is in JSON format to ensure consistency, readability, and smooth integration into downstream processing. Sentences that are classified as deontic obligations or prohibitions proceed to the subsequent module, Deontic Obligation Analysis.

Listing 1: System Prompt Overview for Obligation Filtering

```
# Classification of potential legal obligations
You are an expert in analysing regulatory texts.
## Context
Your task is to classify each occurrence of the term "
     shall" or "must" or "has/have to" into one of the
     following categories based on its use in the text:
[Presenting the details of Definitions, Constitutive
     statements, Deontic Obligations and Prohibitions,
     Entitlements, Authorizations.]
## Instructions:
[A set of instructions on how to analyse and extract the
     required elements and how to structure the JSON
     output.]
[Few shot examples]
[Final Remarks]
```

[2]Code, dataset, validation results and prompts are available at https://github.com/thiagordp/obligation_extraction_for_compliance.
[3]https://eur-lex.europa.eu.

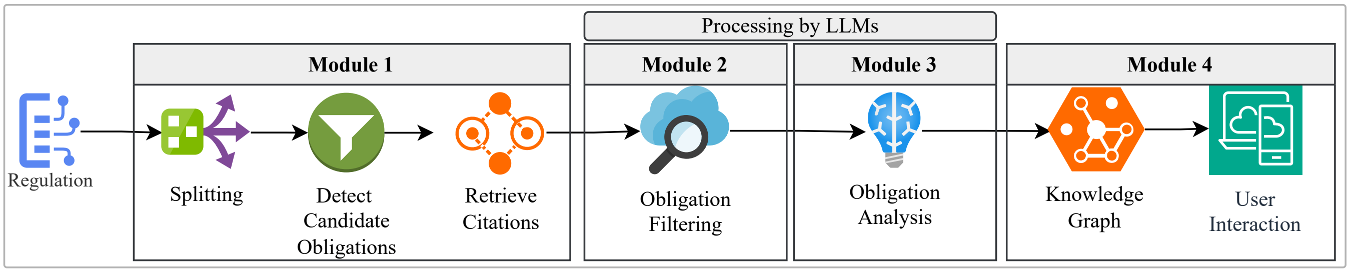

Figure 1: The obligation extraction workflow processes provisions through four modules: obligation detection, deontic obligation filtering, deontic obligation analysis and obligation representation. It integrates LLMs and knowledge graphs to identify, classify and structure legal obligations.

Listing 2: User Prompt Overview and Expected Output for Obligation Filtering

```
# User prompt
## Sentence to analyse
[Sentence here]
## Context
[Paragraph and surrounding paragraphs]
# Expected Output
```json
{   "classification":"[type of sentence]",
    "justification":"[Explanation here]"    }
```
```

4.3 Deontic Obligation Analysis

Based on the results of the obligation filtering step, LLMs are further required to process the text of the "deontic obligation" and break it down into the six constitutive elements delineated in our theoretical framework (i.e., addressee, predicate, targets, specifications, preconditions, beneficiary).

The prompt follows a structured approach similar to the obligation filtering module and includes a system and a user prompt.

A general overview of the system prompt for deontic obligation analysis is presented in Listing 3.

Listing 3: System Prompt Overview for Obligation Analysis

```
# **Legal Syntactical Analysis of Legal Obligations**
You are an expert in legal linguistics and you are
     required to make a syntactical analysis of legal
     obligations.
## **Obligation of Action vs. Obligation of Being**
[Definitions and short examples]
## **Syntactical Analysis Structure**
### **Obligation of Being Elements**
[Definitions and short examples]
### **Obligation of Action Elements**
[Definitions and short examples]
## **Extraction Method**
[Definitions of "Stated", "Context", "Citation", and "
     Background-Knowledge"]
## **Output Structure**
[Presentation and explanation of the expected JSON
     Structure]
[Few shot examples]
[Additional guidelines]
[Final remarks]
```

The prompt begins with a high-level task definition to ensure clarity. Then, it introduces the distinction between obligations of being and obligations of action, supported by their definitions and examples. Next, the obligation extraction framework is presented, detailing the key components as detailed in our theoretical framework: Addressee, Predicate, Target, Specifications, Pre-conditions and Beneficiaries. Each of these elements is interpreted differently depending on the type of deontic obligation (action/being). Note that multiple values can be extracted for each element of the framework (e.g., there may be more than one addressee for the same obligation).

The system prompt also includes instructions on how to determine the extraction method, i.e. where the source of information is taken from. These methods include:

(1) *Stated*, when the information is explicitly present in the deontic obligation analysed;
(2) *Context*, when it is inferred from the surrounding text;
(3) *Citation*, when it is inferred from referenced sections;
(4) *Background Knowledge*, when it is inferred based on implied knowledge of the model;
(5) *None*, when no reliable extraction is possible.

We included this additional classification task for the LLM for both technical and legal reasons. Technically, initial results showed that prompting the LLM to classify the extraction methods minimised the risk of hallucinations, especially when the information was not explicitly stated in the provision. From a legal standpoint, this approach ensures a comprehensive understanding of obligations by considering not only explicitly stated information but also the broader context and referenced sections. Methods like "Context" and "Citation" enable the reconstruction of obligations dispersed in the surroundings of the analysed provision.

Conversely, the "Background Knowledge" category ensures that when explicit information is unavailable in the provision or its context/extracted references, the model is allowed to make probabilistic inferences based on its pre-existing knowledge. Besides ensuring the completeness of the extraction, this category provided a means to test the model's understanding of general legal knowledge and its capacity for logical reasoning beyond the explicit text.

Finally, the system prompt includes a series of final remarks to reinforce key extraction principles and prevent the LLM from omitting details. This ensures a more reliable output.

In the *user prompt*, the materials required for the LLM to analyse are presented in a clear manner. This includes the sentence to be analysed, the context, and any relevant citations. The following list provides a general overview of the prompt (Listing 4).

Listing 4: User Prompt Overview for Obligation Analysis

```
## Sentence to analyse
[Content of the sentence to be analysed]
## Context
[Surrounding text where the sentence appears]
## Citations to other paragraphs, sections, if any
[Content of citations if any]
```

In this step, too, LLMs are tasked with providing a structured, machine-readable output in a JSON format (Listing 5). This step acts as a bridge between raw legal text and structured representations.

Listing 5: Expect Output Structure for Obligation Analysis

```
[
    {
        "ObligationTypeClassification": "string", // "
     Obligation of Action" or "Obligation of Being" or "
     Unknown"
        "Addressees": [
            { "extraction_method": "string", "value": "
     string" }, [...]],
        "Predicate": { "extraction_method": "string", "
     value": "string", "verb": "string" },
        "Targets": [ { "extraction_method": "string", "
     value": "string" }, [...] ],
        "Specifications": [ { "extraction_method": "
     string", "value": "string" }, [...] ],
        "Pre-Conditions": [ { "extraction_method": "
     string", "value": "string" }, [...] ],
        "Beneficiaries": [ { "extraction_method": "string
     ", "value": "string" }, [...] ]
    }
    [...]
]
```

4.4 Deontic Obligation Representation

The Deontic Obligation Representation module is the final step of our framework, where the extracted and structured deontic obligations are represented semantically. To effectively visualise and enable querying of the extracted obligations, we leverage knowledge graphs, which highlight the relationships between entities, actions, and other elements involved in the obligations.

To construct a unified and semantically rich knowledge graph for legal obligations, an ontology is developed to normalise entities, actions, and relationships. The generalised concepts, previously selected by legal experts, are used to cluster similar terms, ensuring consistency and facilitating semantic searches.

Semantic clustering is then carried out using transformer-based models, such as Sentence Transformers. The clusterings align specific occurrences with the generalised ontology elements, grouping semantically related terms. This approach ensures that variations in syntax or phrasing do not hinder accurate obligation representation. Once constructed, the final knowledge graph can be exported in GraphML format, allowing it to be visualised or integrated into graph database systems.

5 Implementation

We have implemented and evaluated our framework across three major EU digital regulations: GDPR, DSA and AIA, which share a common focus on regulating high-impact automated technologies. However, they differ in scope of application, adoption timeline (2016, 2022, 2024, respectively), and terminology, making harmonised compliance difficult. The application of our framework demonstrates its potential to support the extraction of obligations across overlapping regulatory contexts.

5.1 LLM Selection

We chose LLaMa 3.3 70B for its open accessibility, strong balance of performance and efficiency in legal tasks, and solid benchmark scores (68.9% on MMLU PRO, 50.5% on GPQA Diamond). We set the temperature at 0.6, Top P at 1.0 and Top K at 50 to balance variability and coherence. The model was accessed via Together AI,[4] with all other parameters left at their default values. Due to cost and labelling constraints, we relied on prompt engineering rather than fine-tuning.

5.2 LLM Validation

The validation of the obligation extraction process was achieved through a structured evaluation of two key tasks: deontic obligation filtering and analysis. Four independent legal experts, unfamiliar with the system's development, assessed random samples of 30 sentences per regulation in a blind review setting. Each sentence was subjected to review by two evaluators. In the context of the obligation filtering task, evaluators employed a binary scoring system to ascertain the classification accuracy and the relevance of the provided justification.

In the context of obligation analysis, evaluators are tasked with the verification of the accurate extraction of obligation type, in addition to the assessment of key elements, namely the addressee(s), target, predicate, precondition(s), and beneficiary(s), with regard to correctness, with evaluators determining whether the extracted values were accurate or, in cases where background knowledge was applied, at least reasonable within the legal framework. The agreement rates between evaluators, as presented in Tables 2 and 5, functioned as a performance indicator. Disagreements were regarded as incorrect predictions to ensure rigorous evaluation.

5.3 LLM Experimental Results and Discussion

The performance of the LLM was evaluated using accuracy, as is required by the binary classification framework. While metrics such as precision, recall, and F1-score may offer additional insights, the current evaluation design does not take into account false positives or negatives. It is recommended that future research address this issue by adapting the methodology accordingly.

As demonstrated in Table 1 for the obligation filtering task, the results demonstrate consistent accuracy across all three regulations, with an overall accuracy of 0.84 for both classification and justification. The AIA attained the highest score (0.86), presumably due to the fact that the initial tailored version of the prompt was subsequently adapted for the GDPR and DSA, which received a slightly lower score (0.83). This finding suggests that the design of prompts according to regulatory specifications may enhance performance.

The evaluator agreement rate in Table 2 reinforces these findings, with an overall agreement of 0.85 across all regulations, highlighting the reliability of the human validation process.

In the Obligation Filtering step, one can observe that the majority of deontic obligations and prohibitions detected are represented

[4] https://api.together.ai/

Table 1: Obligation filtering evaluation results for the three datasets in terms of Accuracy.

Element	AIA	DSA	GDPR	Overall
Classification	**0.86**	0.83	0.83	0.84
Justification	**0.86**	0.83	0.83	0.84

Table 2: Obligation filtering evaluation results for the three datasets in terms of Evaluator's Agreement Rate.

Element	AIA	DSA	GDPR	Overall
Classification	**0.86**	0.83	**0.86**	0.85
Justification	**0.86**	0.83	**0.86**	0.85

by the term "shall", suggesting that "shall" is predominantly employed to denote obligatory mandates across the three datasets, as presented in Table 3.

Table 3: Obligation Filtering: Frequency of candidate obligations per category and dataset.

Category	AIA	DSA	GDPR
Authorisations	4	2	1
Constitutive statement	74	64	46
Definition	4	0	1
Deontic obligation	603	392	323
Deontic prohibition	31	24	29
Entitlements	11	16	23
Not applicable	2	1	0
TOTAL	**729**	**423**	**499**

In the deontic obligation analysis task, the system exhibited strong performance across most elements, as shown in Table 4. The highest values for each dataset are highlighted in bold. For obligation type classification, the system achieved near-perfect accuracy (> 0.99) on the AIA, 0.97 on the DSA, and 0.84 on the GDPR, yielding an overall accuracy of 0.94. Other key components, such as addressees and predicates, also achieved high overall accuracies of 0.90 and 0.91 for the corresponding value, and 0.88 and 0.97 for the extraction method, respectively. However, the system struggled more with specifications and pre-conditions, particularly for the GDPR, where the accuracy for value recognition was 0.61 and 0.45, respectively. Despite this, the AIA consistently yielded the best results across all elements, demonstrating the robustness of the pipeline for newer regulations with more structured language.

For the obligation analysis, too, the evaluator agreement rates, as shown in Table 5, support the observed trends, with agreement rates closely mirroring the accuracy results. The highest agreement rates are highlighted in bold.

Confidence intervals were calculated for the best-case (AIA obligation type) and worst-case (GDPR pre-conditions) scenarios to assess variability in the model's performance. Using a 95% confidence level with Wilson's method [26], the AIA obligation type's accuracy was estimated between 0.90 and 1.00, indicating potential

Table 4: Obligation Analysis evaluation results for the three datasets in terms of Accuracy.

Element	Feature	AIA	DSA	GDPR	Overall
Obligation Type	Value	**>0.99**	0.97	0.84	0.94
Addressees	Value	**>0.99**	0.95	0.74	0.90
	Extraction	0.94	**0.97**	0.71	0.88
Targets	Value	0.86	**0.87**	0.77	0.84
	Extraction	0.91	**0.97**	0.84	0.91
Predicate	Value	**0.94**	0.92	0.87	0.91
	Extraction	0.97	**>0.99**	0.94	0.97
Specifications	Value	**0.89**	0.74	0.61	0.75
	Extraction	**0.89**	0.79	0.74	0.81
Pre-Conditions	Value	**0.89**	0.74	0.45	0.70
	Extraction	**0.89**	0.74	0.48	0.71
Beneficiary	Value	**0.89**	0.74	0.77	0.80
	Extraction	**0.89**	0.66	0.77	0.77

Table 5: Obligation Analysis evaluation results for the three datasets in terms of Evaluator's Agreement Rate.

Element	Feature	AIA	DSA	GDPR	Overall
Obligation Type	Value	**>0.99**	0.97	0.84	0.94
Addressees	Value	**>0.99**	0.95	0.77	0.91
	Extraction	0.97	**0.97**	0.81	0.92
Targets	Value	**0.94**	0.87	0.77	0.87
	Extraction	0.91	**0.97**	0.84	0.91
Predicate	Value	**0.94**	0.92	0.87	0.91
	Extraction	0.97	**>0.99**	0.94	0.97
Specifications	Value	**0.91**	0.84	0.61	0.80
	Extraction	**0.91**	0.89	0.74	0.86
Pre-Conditions	Value	**0.89**	0.82	0.48	0.74
	Extraction	**0.89**	0.79	0.52	0.74
Beneficiary	Value	**0.89**	0.76	0.77	0.81
	Extraction	**0.89**	0.74	0.77	0.80

overfitting or the influence of the small sample size. For GDPR pre-conditions, the confidence interval was much wider, ranging from 0.29 to 0.62, reflecting significant variability and areas needing improvement.

These findings highlight the importance of larger datasets to reduce uncertainty and improve the precision of accuracy estimates. Despite these limitations, the study provides a strong initial framework for AI-driven legal obligation extraction and identifies areas, like specifications and pre-conditions, for further refinement and targeted enhancements.

Across the two tasks, it clearly emerges that the LLM comparatively performed better on the AIA than on the GDPR or DSA. This result is somewhat unexpected from a legal perspective, as the AIA is often regarded as being written in a more complex and less straightforward manner than the GDPR. This outcome may, however, depend on the fact that, for large parts, the prompts were designed and optimised for the structure and language of the AIA. Possibly, the use of examples in the prompts, which were predominantly inspired by the regulatory framework in the AIA, was determinant.

All in all, this result suggests that the model’s effectiveness is influenced less by the inherent structure of obligations and more by how the prompts were constructed and optimised. In future work, adjusting the few-shot approach to reflect other regulations like the GDPR or DSA could reveal how example selection impacts the model’s ability to generalise across texts.

Secondly, since our pipeline is automated, errors in earlier stages can propagate through subsequent steps, impacting the accuracy of the overall task. For instance, if a "candidate obligation" is incorrectly classified as a deontic obligation, it will be processed by the second LLM and parsed as if it were deontic. This can lead to misleading extractions, as the model will attempt to identify elements such as addressees, targets, and conditions within a sentence that does not impose an obligation.

Finally, another important factor contributing to lower accuracy scores in some cases (e.g. GDPR preconditions and specifications) was evaluator disagreements (which, by our design, was considered as a proxy for an LLM’s incorrect output) rather than outright misclassification by the model. In many cases, differences in interpretation among evaluators led to inconsistent assessment. This may also depend on the way elements such as “specifications” and “pre-conditions” were defined in our framework. While these definitions provide a general understanding, they may be too broad to consistently capture all relevant elements across different regulations. A more focused and detailed set of definitions, combined with enhanced evaluator training and clearer guidelines, would likely improve consistency in assessments and provide a more accurate measure of the model’s true performance.

5.4 AI Act Knowledge Graphs

In this subsection, we further present the results of the last module (i.e. deontic obligation representation) with regard to the obligations automatically extracted from the AIA.

Following the general methodology explained in Section 4.4, we aimed to build a knowledge graph capturing the relationships between obligations, entities, and conditions specified in the regulation. To ensure consistency and manage variability in terminology, we enriched the ontology by clustering terms such as "provider," "deployer" and "authorities", which showed a high degree of variability in the extracted information under broader, expert-selected categories.

For example, we included the generalised concept of "provider" to encompass specific occurrences in the text (e.g., “providers of high-risk AI system“, “providers of AI system”, “prospective provider” etc.), aligning queries and searches with the language used in the regulation.

The semantic clustering was carried out using Sentence Transformers, using general models (a smaller one `all-MiniLM-L6-v2`, and `nomic-ai/modernbert-embed-base`, a more modern and larger version), as well as models trained on EU legislation[5]. The difference between the models was minimal to none, mainly due to the fact that the terms clustered were short and fairly distinct.

Finally, we exported the resulting knowledge graph as a GraphML file,[6] to enable advanced querying, visualisation, and analysis of the relationships and structures within the graph. This integration supports advanced functionalities, including searching, filtering, and linking obligations to specific entities or categories, making the representation both practical and scalable for compliance and analysis tasks. Figure 2 shows an example of knowledge graphs of some obligations of the provider (addressee) under the AIA.

6 Conclusions and Future Work

This paper presented a structured, automated pipeline that combines advanced natural language processing techniques and legal theory to extract obligations, filter them for relevance, and represent them in a visually navigable and semantically rich knowledge graph. This approach offers practical benefits, such as providing a structured overview that may support stakeholders in identifying relevant obligations more efficiently. It also provides a foundation for future tools that can enhance legal interpretation, obligation tracking, and compliance management.

Our findings reveal that LLMs, particularly when integrated with a robust workflow and enhanced prompts, perform well in identifying and classifying obligations across complex legal texts, achieving high accuracy in filtering and analysing deontic obligations. Furthermore, the use of knowledge graphs adds value by facilitating a structured representation of obligations, enabling better traceability, consistency, and contextual understanding.

Despite these contributions, our work presents certain limitations that warrant further consideration in future research.

One key area for improvement is the refinement of our framework. Some categories, such as specifications and pre-conditions, remain too broad, making it difficult to consistently capture the nuances of legal obligations across different regulations. To address this, future work will explore the introduction of finer-grained subcategories, distinguishing, for example, between procedural benchmarks and substantive criteria for compliance.

In addition, improving prompt design is crucial to increasing the model’s adaptability across different regulations. As discussed earlier, the model performed significantly better on the AI Act compared to other regulations, which is likely due to the examples used in the prompt, primarily based on the AI Act’s structure and terminology. Future research will investigate whether modifying the examples to better reflect the linguistic and structural diversity of regulations such as the GDPR and DSA can enhance the model’s generalisability. Moreover, expanding the evaluation to include other EU digital regulations will allow for a more comprehensive assessment of the framework’s scalability.

Another aspect that requires attention is the use of contextual information. While our results indicate that contextual information improves performance, references beyond nearby paragraphs (e.g., citations to other sections) were found to be less critical. However, these structural elements may still hold value in specific legal contexts. Future iterations of the framework will focus on developing more refined methods for integrating cross-referenced information, ensuring that contextual dependencies are considered appropriately without introducing unnecessary complexity.

[5]All the models used were fetched from HuggingFace.

[6]For example in Neo4j, available at https://neo4j.com.

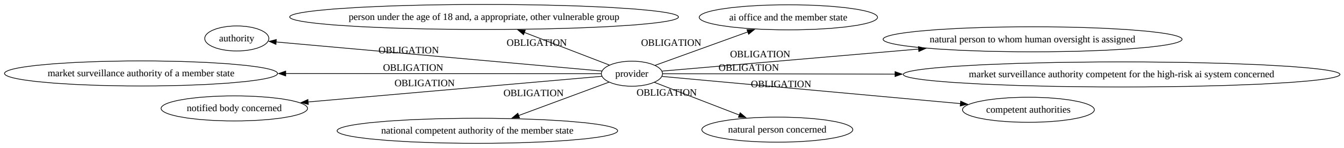

Figure 2: AIA's "Provider" Knowledge Graph.

To further strengthen our findings, experimenting with alternative LLMs will be an essential next step. This will help determine whether certain models perform better in legal text processing.

Further evaluation will be carried out in future developments on the output of the first module (Obligation Detection), to identify issues that may propagate through the pipeline. We consider however that the impact of error propagation from this phase is arguably limited: False Negatives may result in some relevant sentences being excluded from further analysis, while False Positives are handled by the downstream LLM for filtering, which is expected to filter them out for the subsequent stages.

Another limitation relates to the evaluation procedure, particularly in how disagreements between evaluators are handled. Currently, these disagreements are treated as incorrect predictions, which may not accurately reflect the model's performance. To improve this aspect, we plan to implement an ex-post agreement procedure, allowing for a more systematic resolution of evaluator disagreements.

Beyond these refinements, extending the framework will be necessary to accommodate more complex aspects of legal obligations. One promising direction is incorporating logic-based reasoning, exceptions, and temporal constraints as distinct elements. This will enable a deeper analysis of how obligations interact, particularly in cases where compliance depends on sequential conditions or exceptions. To enhance specialisation and accuracy, we may explore a multi-agent LLM approach, where different models are dedicated to specific tasks within the framework.

Finally, an important avenue for future work is the integration of legal ontologies. Linking obligations to structured legal concepts within EU regulations will allow us, in the long run, to identify overlaps between obligations and enhance the interpretability of compliance requirements.

Acknowledgments

The work has been partially supported by the "CompuLaw" project, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 833647).

References

[1] Vijay K Bhatia. 2005. *Vagueness in normative texts*. Vol. 23. Peter Lang.

[2] Mauro Dragoni, Serena Villata, Williams Rizzi, and Guido Governatori. 2016. Combining NLP approaches for rule extraction from legal documents. In *1st Workshop on MIning and REasoning with Legal texts (MIREL 2016)*.

[3] European Commission. 2025. English Style Guide. A handbook for authors and translators in the European Commission. https://commission.europa.eu/document/download/c45f5b70-2d0e-4da7-b181-b5fe3a16c4bb_en

[4] Annarita Felici. 2012. 'Shall' Ambiguities in EU Legislative Texts. *Comparative Legilinguistics* 10, 1 (2012), 51–66.

[5] Xibin Gao and Munindar P Singh. 2014. Extracting normative relationships from business contracts. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 101–108.

[6] Inge Graef and Bart Van der Sloot. 2024. *The Legal Consistency of Technology Regulation in Europe*. Bloomsbury Publishing.

[7] Herbert L.A. Hart. 1961. *The Concept of Law*. Oxford University Press.

[8] Shabnam Hassani. 2024. Enhancing Legal Compliance and Regulation Analysis with Large Language Models. *arXiv preprint arXiv:2404.17522* (2024).

[9] Wesley Newcomb Hohfeld. 1917. Fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal* 26, 8 (1917), 710–770.

[10] Hans Kelsen. 1967. *The Pure Theory of Law*. University of California Press.

[11] Nadzeya Kiyavitskaya, Nicola Zeni, Travis D Breaux, Annie I Antón, James R Cordy, Luisa Mich, and John Mylopoulos. 2008. Automating the extraction of rights and obligations for regulatory compliance. In *Conceptual Modeling-ER 2008: 27th International Conference on Conceptual Modeling, Barcelona, Spain, October 20-24, 2008. Proceedings 27*. Springer, 154–168.

[12] Nadzeya Kiyavitskaya, Nicola Zeni, Luisa Mich, James R Cordy, and John Mylopoulos. 2006. Text mining through semi automatic semantic annotation. In *Practical Aspects of Knowledge Management: 6th International Conference, PAKM 2006, Vienna, Austria, November 30-December1, 2006. Proceedings 6*. Springer, 143–154.

[13] Bimal Kumar and Dmitri Roussinov. 2024. NLP-based Regulatory Compliance–Using GPT 4.0 to Decode Regulatory Documents. *arXiv preprint arXiv:2412.20602* (2024).

[14] Jinqi Lai, Wensheng Gan, Jiayang Wu, Zhenlian Qi, and S Yu Philip. 2024. Large language models in law: A survey. *AI Open* (2024).

[15] Davide Liga and Livio Robaldo. 2023. Fine-tuning GPT-3 for legal rule classification. *Computer Law & Security Review* 51 (2023), 105864.

[16] Kumar Manas, Stefan Zwicklbauer, and Adrian Paschke. 2024. TR2MTL: LLM based framework for Metric Temporal Logic Formalization of Traffic Rules. *arXiv preprint arXiv:2406.05709* (2024).

[17] Zach Meyers. 2025. If the "Brussels Effect" Fades in Tech Markets, the EU Will Only Have Itself to Blame. *Competition Policy International* (3 January 2025).

[18] James O' Neill, Paul Buitelaar, Cecile Robin, and Leona O' Brien. 2017. Classifying sentential modality in legal language: a use case in financial regulations, acts and directives. In *Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law*. 159–168.

[19] Ha-Thanh Nguyen, Wachara Fungwacharakorn, and Ken Satoh. 2023. Enhancing logical reasoning in large language models to facilitate legal applications. *arXiv preprint arXiv:2311.13095* (2023).

[20] Vagelis Papakonstantinou and Paul De Hert. 2024. *The Regulation of Digital Technologies in the EU: Act-ification, GDPR Mimesis and EU Law Brutality at Play*. Taylor & Francis.

[21] Henry Prakken and Marek Sergot. 1996. Contrary-to-duty obligations. *Studia Logica* 57 (1996), 91–115.

[22] Annalisa Sandrelli et al. 2020. A corpus-based study of deontic modality in English Eurolect. *ESP Across Cultures* 17 (2020), 113–128.

[23] Stéphanie Laulhé Shaelou and Yulia Razmetaeva. 2023. Challenges to Fundamental Human Rights in the age of Artificial Intelligence Systems: shaping the digital legal order while upholding Rule of Law principles and European values. In *ERA Forum*, Vol. 24. Springer, 567–587.

[24] George H. von Wright. 1963. *Norm and Action*. Routledge and Kegan Paul.

[25] Bernhard Waltl, Johannes Muhr, Ingo Glaser, Georg Bonczek, Elena Scepankova, and Florian Matthes. 2017. Classifying legal norms with active machine learning. In *Legal Knowledge and Information Systems*. IOS Press, 11–20.

[26] Edwin B. Wilson. 1927. Probable Inference, the Law of Succession, and Statistical Inference. *J. Amer. Statist. Assoc.* 22, 158 (June 1927), 209–212. https://doi.org/10.1080/01621459.1927.10502953

[27] Adam Wyner and Wim Peters. 2011. On rule extraction from regulations. In *Legal knowledge and information systems*. IOS Press, 113–122.