Computer Law & Security Review 58 (2025) 106171

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Towards semi-automating European legislative harmonisation analysis: A harmonised glossary for LLM-based legal concept detection

Davide Audrito[1,*], Ivan Spada[1], Rachele Mignone, Emilio Sulis, Luigi Di Caro

Computer Science Department, University of Torino, via pessinetto 12, 10149 Torino, Italy

ARTICLE INFO

Keywords:
Legal harmonisation
Harmonisation glossary
Large language models
Similarity
Legal concept detection
Concept-based similarity

ABSTRACT

Achieving legislative harmonisation within the European Union (EU) is a multifaceted challenge, hampered by various political, economic, and legal complexities. This article addresses the significant issue of non-compliance by EU member states in transposing EU laws into national frameworks, underscored by numerous infringement procedures. This work introduces a novel methodological framework that combines semantic knowledge modelling and transformer-based language models to address discrepancies in legislative harmonisation. Central to the proposed methodology is the creation of a comprehensive glossary designed to establish correspondences between European legislative concepts and their national counterparts, thus facilitating greater accuracy in legal harmonisation. By deploying Large Language Models (LLMs) for semi-automating concept detection, complemented by legal harmonisation expert's oversight, this research provides an exhaustive, explainable assessment of legislative approximation within the EU. The findings enrich the academic debate on legal harmonisation offering actionable tools designed to decrease the frequency and gravity of infringement procedures, while promoting a more unified and efficient legal framework across the Union. The complete dataset and resources are available at the following link: GitHubrepository.

1. Introduction

The harmonisation of legislative frameworks within the European Union (EU) remains a pivotal yet challenging objective, whose pursuance may have concrete political, economic, and social advantages.

The harmonisation of legislative frameworks within the European Union (EU) remains a pivotal yet challenging objective, whose pursuance may have concrete political, economic, and social advantages. However, this endeavor represents one of the most ambitious legal projects in modern history, yet it remains hindered by significant political and legal complexities. Member states operate within diverse constitutional frameworks, legal traditions, and administrative systems that have evolved over centuries. These differences are not merely technical obstacles but reflect deeper political realities — from varying conceptions of sovereignty and democratic legitimacy to contradictory economic interests and regulatory traditions. The harmonisation process has evolved considerably since the Treaty of Rome, moving through phases of minimal coordination, mutual recognition [1,2], and varying harmonisation [3,4], each reflecting the evolving balance of power between European institutions and national governments. Political factors, including electoral cycles, coalition dynamics, and the rise of eurosceptic movements, have repeatedly modified the commitment to harmonisation, creating a complex landscape where legal, institutional, and political considerations interrelate and sometimes conflict.

As demonstrated in recent studies, including the InfringEye report [5], the inability of EU member states to effectively transpose EU legislative acts into their domestic legal systems has led to a significant number of infringement procedures initiated by the European Commission. By April 2021, an overall amount of 1821 procedures were opened by the European Commission, with the highest non-compliance cases in the legislative areas of environment, energy, mobility, internal market, and justice [6].

There are three primary causes of infringement procedures: violations of regulations and treaties; delays in transposing directives; and incorrect transpositions or applications of directives [7]. To address these challenges, this paper proposes a methodological framework designed to help EU member states correct and prevent errors in the transposition or application of directives and regulations.

The complexity of the research objective mirrors the intricacy of legal harmonisation within the European Union [8]. The persistent

* Corresponding author.
E-mail addresses: d.audrito@unito.it (D. Audrito), ivan.spada@unito.it (I. Spada), rachele.mignone@unito.it (R. Mignone), emilio.sulis@unito.it (E. Sulis), luigi.dicaro@unito.it (L. Di Caro).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.clsr.2025.106171

Available online 2 August 2025

challenges thereof emphasise the need for innovative legal tech solutions to enhance the efficiency and accuracy of consistency assessments. A key challenge is ensuring that legal concepts are consistently interpreted and applied across different jurisdictions, despite linguistic and cultural differences. Indeed, from a mere linguistic perspective, the effective transposition of EU legislation depends on several factors, including the complexity of the language, the multilingual nature of EU law, the type of transposing legislation, and the influence of domestic legal cultures [9]. Accordingly, leveraging semantic knowledge modelling in combination with the cutting-edge, transformers-based large language models (LLMs) does represent a promising approach [10].

The Harmonised Glossary (HG), a novel, interconnected, and taxonomic glossary of European legislative and national corresponding concepts, is proposed as an instrumental tool in analysing legislative harmonisation viewed as a textual phenomenon and through the lenses of legislative approximation. It serves as a foundational tool for standardising terminology, thereby ensuring consistency and clarity across different legal systems. The decision to model fine-grained legislative concepts pursues the aim to overcome subtle, hidden inter-linguistic and intra-linguistic divergences, which are one of the factors that hinder complete and effective harmonisation.

Employing a manually-crafted glossary of concepts extracted from the Council Framework Decision No. 2002/584/JHA, the so-called European Framework Decision on the European Arrest Warrant (FD EAW), a language model was deployed to semi-automatically extract corresponding domestic legislative concepts, relying on its inherent linguistic and semantic knowledge. Besides enriching the glossary with precise legal annotations, this methodology also introduces semi-automation in the analysis of cross-domain legislative texts, without giving up to legal experts' testing and validation of the results. This mixed approach combines semantic- and syntax-based technologies, allowing for a comprehensive, nuanced, and interpretable analysis of the legislative approximation processes within the EU and across its member states.

The HG, in conjunction with LLM-based concept extraction, offers policymakers and legal practitioners a scalable approach to evaluating the effectiveness of European legislation and identifying areas where further harmonisation efforts are needed. This research contributes to the ongoing work of transposing EU law into national legislation, with the aim of reducing both the frequency and severity of infringement procedures. Ultimately, this work seeks to support a more coherent, unified, and efficient legal framework throughout the European Union.

To our knowledge, no LLM-based approach has been applied to European legal harmonisation to computationally address legislative comparison and subsequent automatic or semi-automatic assessment of European legislative alignment. We propose an innovative methodology that advances beyond existing legal alignment tools, which primarily rely on black-box text similarity methodologies [11–14] without explicit conceptual understanding. Our approach generates a semantically informed graph-based conceptual resource that explicitly represents relationships between legal concepts, offering more explainability in the legal harmonisation processes. The enrichment of a curated concept glossary with LLM-driven extraction represents a significant contribution to the field, enabling transparent reasoning about legal alignment that traditional methods cannot achieve. By using this resource computationally, we evaluate state-of-the-art LLMs' capabilities for legal language analysis and reasoning, with preliminary results outperforming conventional NLP techniques in capturing nuanced legal relationships. This methodology automates national legislative analysis while maintaining human oversight, accelerating harmonisation processes and advancing European uniformity objectives through interpretable computational assistance rather than opaque similarity scoring.

As previously discussed [15], numerous tasks within the legal domain necessitate a level of expertise to mitigate the potential risks introduced by automated systems. In the absence of a gold standard, the present work was evaluated through stages of manual assessment.

The article is structured around six sections. Section 2 provides an overview of current approaches to legislative analysis and harmonisation in both academia and industry. Section 3 describes the data used in this study. Section 4 details the methodology for constructing the HG and its cross-member state enrichment, including the use of LLMs for automated concept detection. Section 5 presents the results, encompassing a description of the resource and its evaluation. Lastly, Section 7 discusses the conclusions and future work.

The files containing the manual extraction of European concepts from Council Framework Decision No. 2002/584/JHA, the Harmonised Glossary export, the output of the automatic extraction of national legal concepts, and its manual annotation are available in this GitHub repository.

2. Related work

The phenomenon of legal harmonisation within the European Union is a multifaceted challenge that has garnered considerable scholarly attention from legal, linguistic, and computational perspectives. As previously mentioned, the degree of harmonisation depends on a range of factors, from linguistic and cultural nuances among member states to the nature of EU legislation. According to some scholars, achieving complete harmonisation is likely unattainable [8].

From a legal-linguistic viewpoint, the role of legal translation in legal harmonisation has been extensively analysed [16,17]. EU legislative acts are indeed published in twenty-four official languages and versions — pursuant to Article 4 of Regulation No. 1958/1 and Article 22 of the Charter of Fundamental Rights of the European Union (CFREU) — each possessing the same legal validity. Although multilingualism is crucial for ensuring transparency and equality among EU citizens, it may lead to linguistic divergences and issues with the transposition of laws. Translating legal terminology not only necessitates a deep understanding of legal structures but can also lead to significant shifts in legal concepts due to variations in legal language usage [18]. A similar rationale applies to the political obstacles that impede greater harmonisation [19]. EU member states often wish to maintain a certain discretion in more sensitive legislative areas, such as criminal law and its procedures [20,21]. Conversely, in other domains such as contract law, harmonisation tends to be smoother because functional to the coherence of the European single market [22].

The nature of EU legislative sources also impacts harmonisation. Legal harmonisation in the European Union is deeply influenced by its legislative sources, particularly directives and regulations [3,4,23]. According to Article 288(3) TFEU, the former outlines objectives that all member states are required to achieve, offering them the flexibility to determine the methods for achieving these goals. This flexibility, however, can lead to inconsistent law transposition across countries, impacting the uniformity of harmonisation. Regulations, in contrast, are binding and applicable in all member states, ensuring greater consistency and serving as a more effective tool for harmonisation, without the complexities of national implementation. However, in certain instances, regulations may include provisions that require member states to adopt specific national measures or take actions to comply with the standards set forth in the regulation itself [24].

In the realm of computational legal studies, legal harmonisation was explored in various projects and through different approaches. In the EU-funded [projects], legal experts manually crafted transposition tables composed of, on the left column, EU legislative measures divided into articles and paragraphs, and, on the right side, articles and recitals of the corresponding national implementing legislation from various EU countries. Leveraging these legislative corpora, Sulis et al. [11] successfully implemented vector-based similarity metrics to assess the capability to automatically identify correspondences between EU and related domestic legislation. Following a similar approach, Nanda et al. [12] defined a gold dataset of 43 directives and their corresponding National Implementing Measures (NIMs) in three different

languages (Italian, English, and French). In 2020, the identification of hidden links — connections that are not explicitly referred to within the text between directives and domestic transpositions — have been explored, notwithstanding their changing and dissimilar nature [13]. Overall, in the context of legal harmonisation analysis, identifying domestic transpositions of European Directives has been the most relevant research objective [25]. Adopting a sophisticated approach to develop a scalable and unsupervised dataset, Ferrod et al. [14] assembled a comprehensive dataset of legislative measures in a machine-readable format. The implementing legislation was meticulously cross-referenced with transposing legislation listed on the EUR-Lex website and national official journals. The approach enabled the identification of national laws linked to EU legislative acts.

Analysing the phenomenon of legal harmonisation in the European Union necessitates a multidisciplinary approach. In 2023, Audrito et al. [26] contributed to the research area by creating an analogical lightweight ontology of explicit and implicit definitions identified in six EU directives. This multilingual semantic knowledge modelling highlighted inter-lingual divergences within different linguistic versions of the same directives, responding to the need to explore the impact of multilingualism on harmonisation.

Meanwhile, the rapid evolution of Natural Language Processing (NLP) in recent years can be largely attributed to the emergence of transformer-based architectures [27]. This advancement paved the way for significant breakthroughs, notably the development of BERT (Bidirectional Encoder Representations from Transformers) [28] and the subsequent rise of Large Language Models. LLMs, including GPT (Generative Pre-trained Transformer) [29], LLaMA [30], Gemini [31], and Claude [32] series, have demonstrated remarkable capabilities in handling previously challenging language-based tasks [33]. These advancements are already being applied within the legal field, both in academia [34,35] and industry [36], showcasing a shift in how complex language data is analysed across sectors. Simultaneously, further research works explored methods for evaluating these methodologies, comparing manual and automatic approaches. Recent comparisons have highlighted the essential role of manual evaluation in tasks requiring extensive domain expertise [15].

The EU-funded projects [projects] and academic research [37] demonstrate that current data harmonisation and standardisation methodologies remain mostly manual, inefficient, and susceptible to inconsistencies. Computational language models have been successfully implemented for harmonisation and alignment tasks across large datasets in multiple domains, addressing common challenges such as heterogeneity, unstructured formats, domain specificity, and privacy concerns. Despite shape variations and differing objectives, many alignment challenges remain persistent across domains.

In Environmental Science, previous research [38] utilized a fine-tuned SciBERT model [39] to enhance spatiotemporal data harmonisation in soil analytics. Conversely, in the medical domain, recent experiments address data alignment challenges through the application of state-of-the-art technologies such as Large Language Models.

The capabilities of LLMs for data retrieval, extraction, and generation have been analyzed to create standardised datasets accelerating clinical and translational research in hematological disorders [40]. Higashi et al. [37] combine semantic clustering techniques with LLM-driven natural language processing to extract, interpret, and standardise metadata from diverse sources, including research publications, supplementary tables, and textual data from public repositories. An alternative approach explored LLMs for variable matching across datasets, potentially expediting retrospective data harmonisation used to integrate information from existing datasets or biobanks to create samples [41].

In 2025, Santos et al. [42] introduced Harmonia, a comprehensive system that integrates LLM-based reasoning with an interactive interface and harmonisation primitives library to automate pipeline synthesis for data standardisation. The system's effectiveness was demonstrated in clinical settings, where it enables the creation of reusable transformation workflows for standardising heterogeneous datasets.

Regarding legislative harmonisation, the use of LLMs for legal inquiries raises significant concerns due to requisite expertise and potential impacts [43]. Through workshops with 20 legal experts employing case-based reasoning methodologies, researchers systematically evaluated appropriate boundaries for LLM-generated legal advice [43]. Analysis of realistic case scenarios enabled identification of both granular context-specific considerations and broader technical-legal constraints, yielding actionable guidelines for LLM developers. Their findings suggest the feasibility of translating domain-specific professional knowledge and practices into policies that can guide LLM behaviour properly.

3. A dataset of EU legislative acts and corresponding domestic implementing measures

This study utilises a manually-crafted dataset, specifically designed for both manual and computational analysis, which was created by EU criminal law experts in the framework of the EU-funded [projects]. The dataset comprises parallel corpora of EU legislation and national transpositions, segmented by article and paragraph. To ensure analytical consistency of the multilingual sources, the dataset is presented in English. Specifically, it includes EU legislative acts in their English versions and the official English translation of their corresponding national implementing measures, reflecting the authoritative versions produced by member states.

The data is available in JSON format, enhancing its accessibility for machine-readable applications. The availability of data in JSON format aligns with the European Union's goal of promoting interoperability, as stipulated in the recently adopted EU Interoperable Act (EU Regulation No. 2024/903). The resource comprises 32 articles and their corresponding national implementations, where available. It includes 144 pairs of articles and national implementations in English, distributed as follows: France (27), Germany (26), Italy (31), Portugal (30), and Spain (30).

Fig. 1 illustrates a segment of the transposition table, displaying Art 1 EAW FD segmented into cells on the left side, with the corresponding Spanish legal transpositions on the right side. This visual representation aids in understanding the structural alignment between the EU framework and national implementations.

Despite the detailed nature of the tables, numerous processing steps are required to facilitate effective computational analysis of legislative transposition. For example, textual similarity metrics have been employed recently to verify the semantic correspondence between European and domestic legislation [44]. This crucial phase in legislative harmonisation intentionally omits additional legal sources, such as judicial decisions.

Moreover, past computational methods [25] in legal harmonisation analysis utilised processed transposition tables to evaluate the similarity between EU legislation and corresponding national measures, without accounting for the challenges posed by high semantic similarity within the same legislative areas. Specifically, the task of identifying domestic measures aligned with the EAW FD is complicated by the presence of legislative acts from the same domain — criminal and procedural law — in the corpus.

Embracing the challenge of overcoming the limitations of straightforward, vector-based similarity methods, this research adopts a more in-depth approach, extracting concepts and their definitions from the transposition tables. This manual refinement of the dataset enables the semi-automatic retrieval of concepts, definitions, and corresponding implementing articles, which supports more advanced computational analyses and precise comparative legal studies. The integration of large language models (LLMs) makes this methodology scalable and enhances the models' capability to augment the dataset. This expansion includes the retrieval of domestic concepts and definitions built upon a knowledge base composed of an interconnected, taxonomic glossary of European concepts.

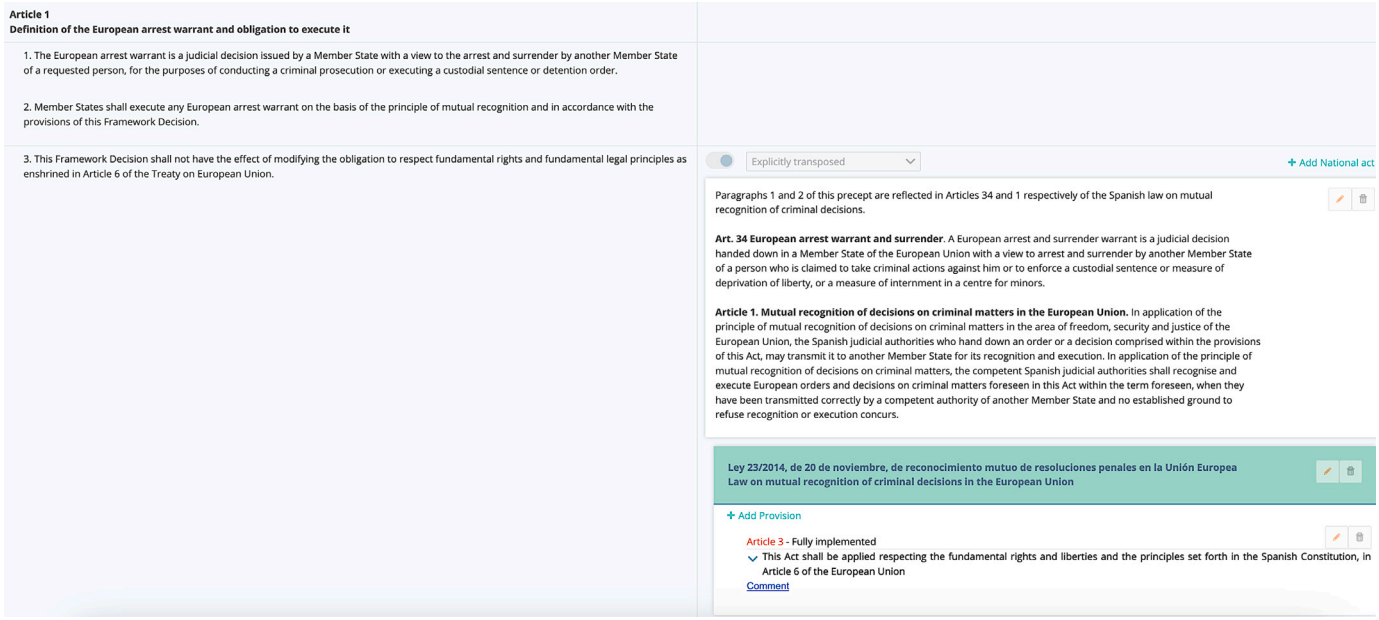


Fig. 1. Excerpt of transposition table for Article 1 FD EAW and corresponding Spanish implementing legislation.

Such investigations are vital for understanding implementation patterns and underlying discrepancies across various legal systems, thus offering insights into the effectiveness of legislative harmonisation within the EU.

4. Methodology

The HG graph is a useful tool for quick and user-friendly identification of EU legal concepts, their national corresponding implementations, and their respective definitions. It consists of a network of EU concepts, articles, and legislative references manually extracted by legal domain experts from the FD EAW. These concepts are interrelated and modelled in a taxonomic, graph-based structure, employing a comprehensive methodology. This initial effort lays the foundation for semi-automatically integrating corresponding concepts enshrined in domestic transposing legislation to provide a data representation suitable for assessing the degree of legislative harmonisation throughout the EU. As a further step, the section outlines the methodology used to automate the analysis of legislative harmonisation, based on a cutting-edge LLM-based extraction.

4.1. Manual concept identification

A legal-tech expert, specialising in legal harmonisation analysis, methodically annotated semantically significant terms within the FD EAW after a thorough review. This process begins with the first normative provision, deliberately excluding recitals. Although recitals serve as interpretative benchmarks that outline the purpose of legislative acts, the main doctrinal approach excludes their legally binding value [45]. A concept is deemed relevant to the extent that its absence would significantly alter the meaning of the entire law or even just one of its paragraphs and is sufficiently specific to be determinate for the comprehension of the essential components of the law under scrutiny.

By way of illustration, Fig. 2 portrays the annotation of concepts in the text of Art 1 FD EAW. The concept of European Arrest Warrant (EAW) is fundamental because the term refers to the general objective of the legislative act. Likewise, the principle of mutual recognition of judgments and judicial decisions, which lays the foundation of the European judicial cooperation in criminal matters pursuant to Article 82 of the Treaty on the Functioning of the European Union (TFEU) is

CHAPTER 1

GENERAL PRINCIPLES

Article 1

Definition of the European arrest warrant and obligation to execute it

1. The European arrest warrant is a judicial decision issued by a Member State with a view to the arrest and surrender by another Member State of a requested person, for the purposes of conducting a criminal prosecution or executing a custodial sentence or detention order.

2. Member States shall execute any European arrest warrant on the basis of the principle of mutual recognition and in accordance with the provisions of this Framework Decision.

3. This Framework Decision shall not have the effect of modifying the obligation to respect fundamental rights and fundamental legal principles as enshrined in Article 6 of the Treaty on European Union.

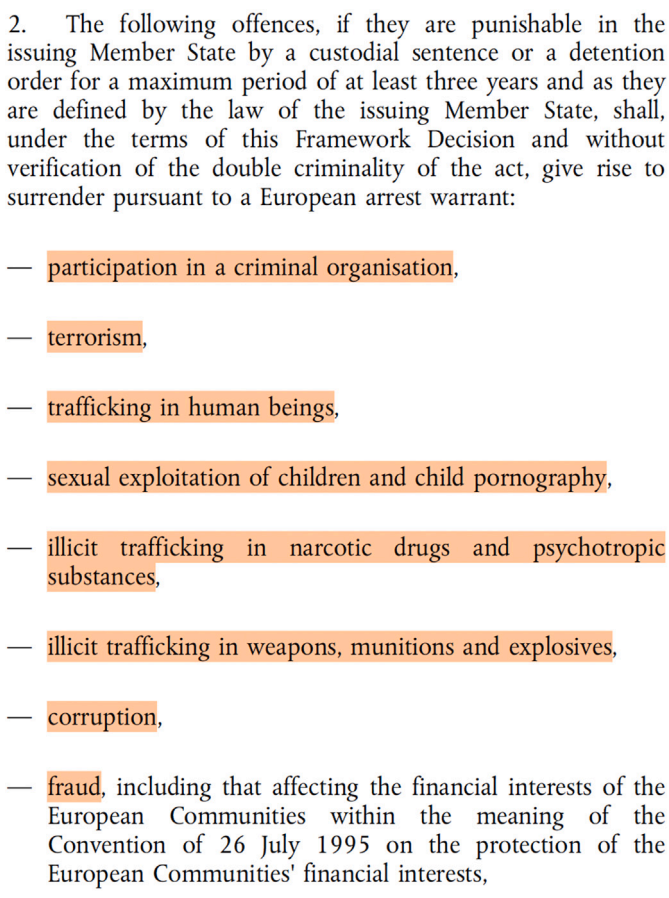
Fig. 2. Manual identification of concepts under Article 1 FD EAW.

essential for the functioning of the mechanism of the EAW. On the contrary, the concept of "entry into force" is not considered semantically relevant. This is indeed a permanent legislative arrangement indicating the scope of application *ratione temporis* of each legislative act and lacking a meaning able to alter the semantic content of enunciates and legal provisions, despite its importance for the terms of validity of the legislative act.

In contrast, concepts such as "terrorism", "human trafficking", or "corruption" are relevant to the FD EAW. These criminal offenses may result, occurring some further conditions and in the absence of any other grounds for refusal, in the surrender of a sought-after person by the competent authorities of one member state to another even if the offence is not provided in the same terms in the legal system of the executing member state. The absence of any of these offenses would significantly alter the meaning of Article 2 FD EAW.

Likewise, the terms "fundamental rights" and "fundamental legal principles" are incorporated into the glossary pursuant to Article 1(3) FD EAW. This inclusion underscores the essential role these concepts play within the context of the FD EAW, emphasising that the operational provisions embedded in the EAW must not contravene the obligations to uphold fundamental rights as stipulated by EU law. This provision clarifies that the legal framework of the FD EAW is subordinate to overarching EU treaties. This subordination ensures that all actions taken under the FD EAW align with, and do not restrict, European fundamental rights. Such integration not only reinforces the legal robustness of the EAW but also safeguards it against potential misuse, aligning its application with the fundamental principles of European law.

The decision to include terms that are not explicitly defined in the FD EAW, along with its implicit concepts, was motivated by the choice to comprehensively model the essential semantic constituents of the law under scrutiny at a very high level of granularity. This methodological choice avoids the risk of omitting relevant concepts and leads to a significant expansion of the annotated corpus. By mapping both explicit and implicit concepts, the semantic modelling of the text is enhanced. This results in a more detailed and accurate glossary of interconnected concepts, which not only improves the glossary's precision but also facilitates easier consultation of the resource. Overall, the identification of concepts and relations required around thirty-two hours, including a subsequent eight-hours revision of the first draft.



2. The following offences, if they are punishable in the issuing Member State by a custodial sentence or a detention order for a maximum period of at least three years and as they are defined by the law of the issuing Member State, shall, under the terms of this Framework Decision and without verification of the double criminality of the act, give rise to surrender pursuant to a European arrest warrant:

— participation in a criminal organisation,

— terrorism,

— trafficking in human beings,

— sexual exploitation of children and child pornography,

— illicit trafficking in narcotic drugs and psychotropic substances,

— illicit trafficking in weapons, munitions and explosives,

— corruption,

— fraud, including that affecting the financial interests of the European Communities within the meaning of the Convention of 26 July 1995 on the protection of the European Communities' financial interests,

Fig. 3. Manual identification of concepts under Article 2(1) FD EAW.

4.2. Interconnecting concepts around a taxonomic structure

The second analytical step organizes concepts within a taxonomic structure and identifies relations between them. In doing so, concepts are linked to broader categories, in a taxonomic structure, that facilitate the use of the glossary for both direct human exploration and NLP purposes, specifically for semi-automatic classification. The broader categories of concepts, which underpin the creation of the taxonomic structure, are further identified as "implicit concepts". As will be illustrated in the next section, also implicit concepts will be incorporated into the system prompt as input for the language model. This second methodological step enhances the identification of both explicit and implicit concepts, as illustrated in Fig. 3. In the HG, definitions of concepts correspond to the sections of text where these concepts are situated within the legislative act, with the exception of the explicit definition of EAW that was manually included in the dataset. The legal-tech expert categorised concepts and highlighted references to the articles and paragraphs where they are located, thereby facilitating the semi-automatic retrieval of definitions.

As illustrated in Fig. 3, the portion of text from Article 1(2) FD EAW lists the categories of offenses that, without verification of the double criminality of the act and provided that certain other conditions are met, give rise to the surrender of the requested person according to the EAW. The list of concepts is fundamental in the fabric of the judicial cooperation mechanism because EU member states' competent authorities will have the obligation to surrender the requested person if charged with one of the listed offenses, notwithstanding the absence of double criminality, unless other grounds for refusal apply. These concepts are linked to the broader concept-category "offense giving rise to surrender pursuant to the European Arrest Warrant". This decision was made to enhance the semantic modelling of the FD EAW and to simplify its consultation, while introducing the broader taxonomic categories of implicit concepts. This approach also enables assessment of the language model's ability to semi-automatically retrieve the nature and location of implicit concepts within the legislative act.

The list of concepts should be understood as the result of a selection process designed to comprehensively reflect the semantic content of legislative enunciates in each article of the FD EAW. It is not intended to represent the normative content but rather stems from the textual content alone. In other words, the concepts are used to evaluate the ability of LLMs to identify semantically co-occurring concepts in EU and national transpositions, with the aim of reducing misalignments and facilitating comparisons between legislative texts. It should be noted that LLMs are not tasked with identifying concepts that may have legally binding value or influence legal practitioners.

In conclusion, the procedure resulted in the creation of two tabular artifacts. The Concept Table, displayed in Table 1, contains the identified concepts along with their locations within the FD EAW. This table serves as a foundation for subsequent efforts at semi-automatic definitions and relations' extraction and indexing. Alongside the Concept-Relation Table that encompasses binary relations between concepts identified by legal experts — as shown in Table 2 — the concept Relation Table is also employed in constructing the graph outlined in Section 5. This data structure enables interconnecting concepts on both a topological and semantic basis.

4.3. Between syntax and semantics: the harmonised glossary as a knowledge base for LLM-based concept-detection

While the extraction of concepts FD EAW was manually performed by a legal-tech expert, the task of extracting concepts from the corresponding domestic implementing legislation was accomplished using an

Table 1
Example of concept table comprising manually annotated concepts and links to legislative articles.

Article	EU extended concepts	EU shorter concepts
Art 1(1), 2(1)	European arrest warrant	European arrest warrant
Art 1(3)	Fundamental rights ex article 6 TUE	Fundamental rights
Art 1(3)	Fundamental legal principles ex article 6 TUE	Fundamental legal principles
Art 2(1)	Act punishable by the law of the issuing member state	Punishable act
Art 2(1)	Custodial sentence for a maximum period of at least 12 months	12-month max custodial sentence
Art 2(1)	Detention order for a maximum period of at least 12 months	12-month max detention order
Art 2(1)	Sentence of at least four months	Sentence of at least four months
Art 3(3)	Passing or making a detention order	Detention order issuance
Art 1(1)	Arrest and surrender of a requested person	Arrest and surrender
Art 2(2)	Offence giving rise to surrender pursuant to an european arrest warrant	Offence for surrender
Art 4(3)	Participation in a criminal organisation	Criminal organisation
Art 2(2)	Terrorism	Terrorism
Art 2(2)	Trafficking in human beings	Human trafficking
Art 2(2)	Sexual exploitation of children and child pornography	Child exploitation
Art 2(2)	Illicit trafficking in narcotic drugs and psychotropic substances	Narcotics trafficking

Table 2
Example of concept-relation table manually annotated that represents the oriented relations between legal concepts.

EU extended concepts (from)	Relation	EU extended concepts (to)
European arrest warrant	Is issued by	Issuing judicial authority
European arrest warrant	Is issued for	Act punishable by the law of the issuing member state
European arrest warrant	Has mandatory compliance	Fundamental rights ex article 6 TUE
European arrest warrant	Has mandatory compliance	Fundamental legal principles ex article 6 TUE
Act punishable by the law of the issuing Member State	Is punished by	Detention order for a maximum period of at least 12 months
Sentence of at least four months	Has condition	Passing or making a detention order
European arrest warrant	Is executed by	Executing judicial authority
European arrest warrant	Has purpose	Arrest and surrender of a requested person
European arrest warrant	Has mandatory execution	Offence giving rise to surrender pursuant to an european arrest warrant
Offence giving rise to surrender pursuant to an european arrest warrant	Has instance	Sexual exploitation of children and child pornography
Offence giving rise to surrender pursuant to an european arrest warrant	Has instance	Illicit trafficking in narcotic drugs and psychotropic substances

LLM-based approach. Creating the initial HG was a meticulous manual process requiring substantial effort. Extending the glossary to all five member states in the dataset would have demanded a similar investment of time. Therefore, the exploration of automating the extraction of European legal concepts within the national implementations, using the initial glossary as grounding knowledge, was opted for.

Close collaboration between legal and computer science experts facilitated the development of both the system prompt and the prompt designed to elicit the requisite legal knowledge and identify the most effective methodologies for the task at hand. Appendix A details the system prompt, which comprises the following components: Role, Awareness of Resources, Context, Example, and Goal. The LLM is introduced as an expert in European legislation and member state implementations, aware of a pre-existing ontology of European legal concepts. It is instructed to operate within the context of the Framework Decision on the European Arrest Warrant and is provided with a concrete example of European legislative text and its constituent concepts. Finally, the LLM is informed of transposition tables, which it should use to analyse European harmonisation and identify correspondences between European and member state legislation.

Several leading generalist large language models (LLMs), distinguished by their scale (parameter count and training data volume) and prominence in current research, were considered for this task. Preliminary experiments were conducted with a subset of these models — LLaMA-3.1-405B [46], GPT-4-Turbo [29], Claude-3.5-Sonnet [32], and Gemini-1.5-Pro [31] — to determine the most suitable candidate. Claude-3.5-Sonnet-200k was ultimately selected due to its superior accuracy and lower propensity for generating fabricated information ("hallucinations"), exhibiting a preference for non-response over misleading output.

Concerning the methodology of prompt design, the Tabular Chain of Thoughts strategy was opted for [47,48]. This prompt engineering approach requires the model to reason in a "step-by-step" manner, performing a series of intermediate calculations before arriving at the final objective. This method also allows the output to be formatted in a tabular form, which is excellent for aligning European and corresponding domestic concepts for comparative purposes.

The extraction prompt is detailed in Appendix B. It details the topic of legal harmonisation in Europe and the task of legal concept extraction. To specify the expected output format, it details the columns of the table to be constructed: *european concept* (a manually annotated concept from the Harmonised Glossary), *corresponding national concept* (a related national concept found by the model), *nature of concept* ("explicit" or "implicit"), *type of concept match* ("semantic exact match", "semantic partial match", or "not applicable"), *corresponding national definition* (a list of national implementation excerpts containing the identified concept), and *notes* (free-text explanation of the results). Following this, it provides two examples illustrating explicit and implicit concepts. Finally, it requests the computation of the table, given a list of candidate European concepts and the national legislative implementations where these concepts might appear. It concludes by using standard markdown notation to specify the desired tabular output format: "| European concept | corresponding national concept | nature of concept | type of concept match | corresponding national definition | notes |". This approach leverages the model's ability to identify corresponding national concepts based on both the provided text and the concepts extracted by legal experts. Appendix C provides a sample LLM input prompt containing EU concepts from the HG and national legislative sections from transposition tables manually created by legal experts within the EU-funded project to facilitate experiment replication. All data appears in English, having been manually translated by legal experts with reference to official translations.

Analyzing the project's transposition tables described in Section 3, which cover three legal acts across five member states to enhance data generalisation, we conducted structural data analysis to evaluate the model's context window requirements. Each European legislative

article query to the LLM comprised 699 tokens for the prompt instruction, plus between 906 and 16,229 tokens representing domestic legislative segments. The token count for European concepts from the HG was negligible. The total token count remained within the selected model's maximum context window capacity and, given the performance detailed in Section 5, no additional processing for smaller context windows was necessary as no context management degradation was observed despite the prompt size. Initially, we deliberately chose large prompts to test the model on complex tasks. Token counts were calculated using the "tiktoken cl100k_base" [49] tokenizer.

An interesting aspect of the prompt is the inclusion of both implicit and explicit concepts. This does not merely evaluate the model's ability to align explicitly codified concepts, whether they be matching words or synonyms, but also its capacity to infer implicit content that mostly represents the taxonomic categories around which the glossary is modelled. Between explicit and implicit concepts, mostly hierarchical relationships exist, which, although not explicitly indicated in the prompt, appear useful in accomplishing the final task.

As will be detailed in the following section, the model achieves exceptionally high results in the retrieval and alignment of both explicit and implicit concepts. Moreover, the output concept table includes additional columns where the model identifies whether a concept is explicit or implicit, the degree of semantic match, partial or total, between European and national concepts, the legislative text segment where the concepts are located, and also provides reflections on each alignment within the "notes" column.

4.4. Evaluating LLM-based concept detection: a domain expert assessment

This subsection describes the process implemented to assess the model's ability to semi-automatically extract legal concepts from legislative texts. Detailed annotation guidelines, provided in Appendix D, ensure the experiment's replicability.

Initially, two evaluation approaches were considered: employing either two or more annotators, or a single annotator. After careful consideration, the latter approach was chosen, entrusting the entire validation process to a single expert who also compiled the initial version of the glossary containing the concepts of the FD EAW. This decision deviates from common practices in the state of the art, favouring a more focused and direct approach given the specificity of the context. The annotator's profound expertise enabled precise identification of both explicit and implicit concepts, the latter often corresponding to taxonomic categories, and established a consistent degree of total or partial semantic correspondence between concepts. Furthermore, the expert was capable of quickly associating concepts with the specific text segments where they were located. Despite the unconventional approach of using a single annotator, it was decided that the unique peculiarities of this evaluation exercise justify this choice. This approach allows to evaluate whether the model's annotations are consistent with those of the annotator who initially created the glossary for the European legislative texts.

Nevertheless, Section 5 provides an example of annotation and the reasoning behind the validation of the results, focusing on the criteria of accuracy, explainability, and consistency. Moreover, to promote transparency and encourage further analysis, the results of the annotation have been published and made freely available to the scientific community in this GitHub repository. The annotation guidelines (see Appendix D) describe a binary classification task, performed by the annotator on model-produced results, to assess concept alignment.

Further, the annotator performed two additional tasks: validating returned concepts and validating omitted concepts. While the prompt (see Appendix B) requests tabular output, the model may omit specified European concepts. The second task determines if these omissions reflect genuine absence from national implementations or model oversight. Omissions reflecting genuine absence do not introduce errors into the HG.

In **Task 1**, annotators were asked the following questions:

1. *Q1: Is the "nature of concept" field correct?* Yes/No
2. *Q2: Is the "type of concept match" field correct?* Yes/No
3. *Q3: Is the "corresponding national definition" field correct and comprehensive?* Yes/No

Task 2 consisted of the following question:

1. *Q4: Is the omitted concept present in the national segment?* Yes/No

For questions Q1, Q2, and Q3, a higher number of "Yes" responses indicates greater accuracy of the automated system. Conversely, for question Q4, a higher number of "No" responses is desirable, signifying that the model omitted fewer concepts that should have been considered due to their presence in the text. Annotation questions are explained in detail in the Appendix D.

4.5. The implementing tool: Neo4J

The choice of Neo4J [50] for implementing the interconnected concept graph derived from the FD EAW is underpinned by various reasons, reflecting its widespread adoption in the state of the art for complex data structures [51]. This graph database excels in handling the intricate data structures essential for mapping the complex relationships between legal concepts and legislative articles.

Firstly, Neo4J offers extensive visualisation capabilities that facilitate the visual interpretation of complex networks of nodes and relationships. This feature is crucial in the legal domain, where clarity and comprehensibility of information are paramount. Secondly, it supports the querying of complex networks, enabling users to perform searches within the graph. This functionality is valuable in the legal context, where users may need to explore the relationships between various legislative concepts to correctly interpret laws or identify relevant precedents.

Thirdly, the scalability of this tool ensures that the database can grow with the addition of more concepts and relationships without a degradation in performance. This is vital in legal research projects, where data volume can expand rapidly. Lastly, it boasts robust transaction management capabilities, essential for maintaining data integrity in an environment where modifications to concepts or relationships can have significant implications.

Furthermore, the graph database can be exported in standard ontological resource formats such as .rdf, .rdfs, and .ttl.

5. Results

As previously clarified, the final version of the glossary is composed of concepts, relationships, and attributes manually identified by a legal expert. Subsequently, the glossary was enriched with corresponding national legislative concepts from various EU member states. Therefore, the first subsection presents the results from the initial version, and the second subsection showcases those from the complete glossary, with meticulously validated results.

5.1. The EU concept-based glossary

The Neo4J graph, derived from manually extracting concepts from the FD EAW, incorporates 124 articles, 198 EU concepts, 212 'found in' relationships, and 225 inter-concept relationships. Fig. 4 below depicts the overall structure of the graph-based glossary.

The graph's complexity, characterised by a high density of hubs, arises from several reasons. High granularity in concept identification results in many sub-classes of broader concepts, such as those categorising crimes that automatically trigger the issuance of a EAW. Furthermore, the arrangement of concepts and relationships enhances information retrieval efficiency. Legal experts have emphasised the roles and functions of the primary institutions in the EAW process—specifically, the issuing and executing national authorities—and have

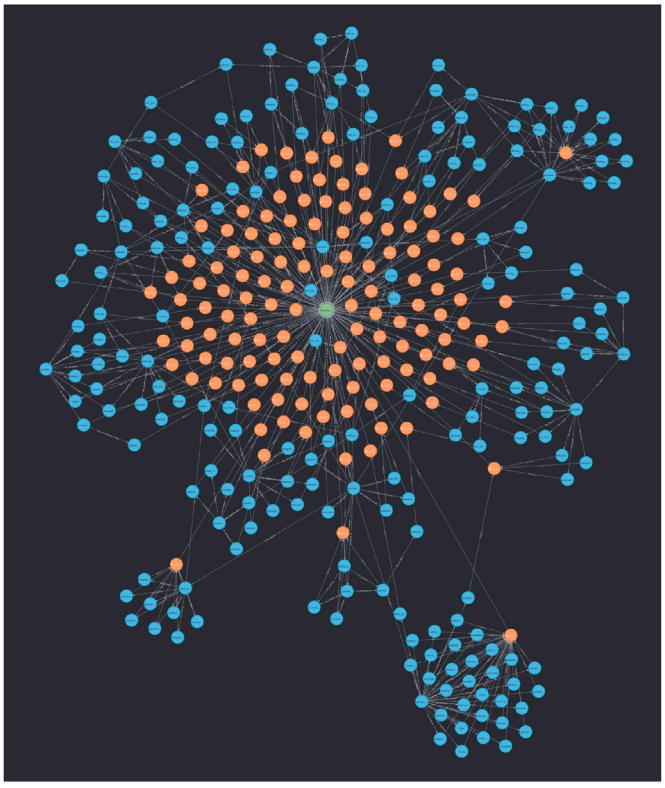


Fig. 4. The HG of concepts from the FD EAW.

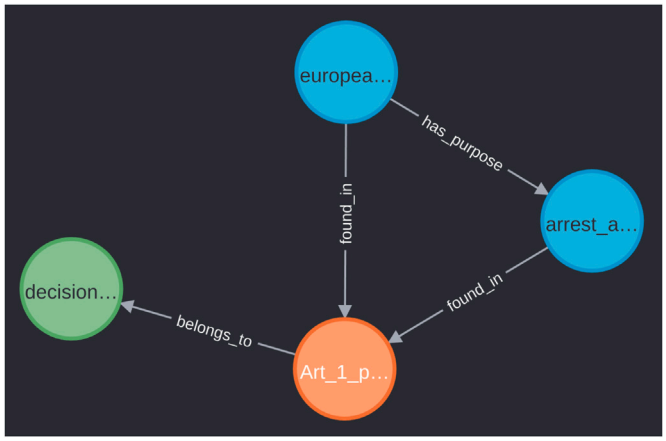


Fig. 5. The modelling of the EAW concept and relations within the FD EAW.

detailed the responsibilities of the member states. As a result, nodes representing these elements have a higher degree and, hence are more interconnected than others. Additionally, while some articles are rich in concepts, others are less so, leading to varying degrees of node content representation across the graph.

The graph-based glossary is composed of entities, i.e. nodes and attributes, and their relations, which are detailed as follows:

Entities

- **EUconcept**: Represents the extracted concepts, interconnected through relations manually defined by legal domain experts. Attributes for each **EUconcept** node include:
  - **id**: Item identifier.
  - **name**: Textual identifier.
  - **long_name**: Full name of the concept as manually extracted by domain experts.
  - **short_name**: Shortened name for display on the node for readability.
  - **definition**: Segment of the European directive defining the concept (if present).
  - **eu_source**: A list of European legislative segments manually annotated that contains the EUconcept.
  - **[nation]_source**: For each nation, it contains a list of legislative portions of text automatically extracted from the transposition tables.
- **Article**: Specifies the location in which each EUconcept was found, expressed as an article, paragraph, and sub-paragraph (if present). Attributes include:
  - **id**: Item identifier.
  - **name**: Full name of the concept as manually extracted by domain experts.
- **EUlegislation**: Specifies the document to which the Article belongs. Attributes include:
  - **id**: Item identifier.
  - **name**: Full name of the concept as manually extracted by domain experts.

Relations

- **Inter-concept relation(s)**: Custom relations between concepts obtained after manual annotation.
- **Found_in**: Links the concept to the related segment of the EU Directive.
- **Belongs_to**: Links the segment of the EU Directive to the EU Directive itself.

The initial version of the glossary, crafted by hand, was solely based on concepts extracted from the FD EAW. Accordingly, the item [nation]_source is empty at this first stage. The subsequent sub-section will present an example of entities pertaining to national concepts and sources, focusing particularly on the most significant concept of the FD EAW, which is the European Arrest Warrant, whose graphical representation is illustrated in Fig. 5.

As an example, Fig. 5 illustrates the concept of the EAW, although the node labels are not entirely legible due to length constraints. At the center and top of the image, the concept of the EAW is linked to Article 1(1) via a 'found_in' relationship, as this concept is initially contained within that specific text segment. Article 1(1), in turn, is connected to the entire FD EAW legislative act through a 'belongs_to' relationship. Finally, the concept of EAW is linked to 'arrest and surrender of the requested person' through a 'has_purpose' relationship, reflecting one of the primary objectives of issuing a EAW: the arrest and surrender of the targeted individual by the 'executing' state to the 'issuing' state.

In contrast to the representation of the EAW concept with its attributes and relationships, the depiction of the criminal offenses outlined in Article 2(2) FD EAW appears more complex, as illustrated in Fig. 6. Each criminal offense, represented as a concept, is connected to the node labelled "offence giving rise to surrender", which serves as a taxonomic category encompassing all the offenses listed in Article 2(2) of the FD EAW. These offenses indeed represent "instances" of this broader category, connected through a "has_instance" relationship. Finally, each concept is linked to the article in which it is contained, displayed in orange, which in turn is embedded within the FD EAW shown in green.

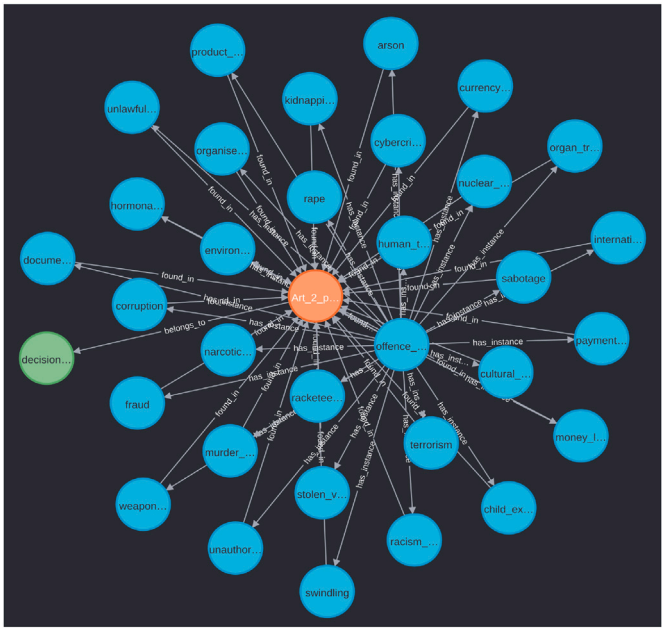


Fig. 6. The modelling of the concept and relations of EAW within the FD EAW. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.2. Incorporating corresponding national concepts: an LLM-based extraction

The graph's structure remains unchanged despite the inclusion of national concepts corresponding to the identified European ones. However, there has been a significant alteration in the number of attributes associated with the nodes, as can be seen in the details pertaining to the EAW concept below. This concept is uniquely defined within the law, which is the reason why its explicit definition was manually included. For sources referencing various EU member states, the corresponding national legislative measures where the concept is embedded are also detailed. This method ensures that the graph accurately reflects the variations and specifics of how the EAW is incorporated and interpreted across various domestic legal systems. Below are the attributes of the EAW concept, presented as follows:

- **Name**: European Arrest Warrant
- **Long Name**: European arrest warrant
- **Short Name**: EAW
- **Definition**: The European arrest warrant is a judicial decision issued by a Member State with the aim of the arrest and surrender by another Member State of a requested person, for the purposes of conducting a criminal prosecution or executing a custodial sentence or detention order.
- **EU Source**:
  - Article 1, Paragraph 1
  - Article 2, Paragraph 1
- **France Source**: Under the law of the issuing Member State, offences may give rise to a European arrest warrant. Refusal of execution may occur if the offence does not constitute an offence under French law.
- **Germany Source**: Section 81 Extradition for prosecution or for enforcement. No double criminality check is carried out if the offence is punishable by deprivation of liberty of at least three years and belongs to the categories listed in Article 2(2) of the Framework Decision.
- **Italy Source**: Execution of the European arrest warrant must not violate specific conditions. It shall be executed regardless of double criminality for offences punishable by at least three years as described in Article 2(2) of the Framework Decision.
- **Portugal Source**: The European arrest warrant may be issued for acts punishable by a custodial sentence or a detention order for at least twelve months, or for sentences of at least four months.
- **Spain Source**: When a European arrest and surrender warrant has been issued.

Beyond the development of the HG, examining the model's output through various iterations of system prompts and prompts is intriguing. This analysis reveals the capabilities of the Claude-3.5-Sonnet-200k model to not only identify corresponding national concepts given European concepts but also to discern whether a concept is explicit or implicit. Furthermore, the model can detect either a total or partial semantic correspondence, identify the text segment containing the concept, and propose "notes" regarding the actual overlap of European and national concepts. This level of detail in the model's output underscores its sophistication in handling complex legal terminology across different legislative frameworks.

Table 3 showcases exemplified outcomes generated by the Claude-3.5-Sonnet-200k model, focusing on the alignment between four European concepts and their counterparts in French legislation. It highlights various instances, such as the EAW concept, which exhibits an exact semantic match and is explicitly mentioned in both legislative frameworks. In contrast, the concept of a "detention order for a maximum period of at least 12 months" finds no direct counterpart in French law, being implicit and possibly inferable from other custodial sentence regulations. Concepts like "illicit trafficking in narcotic drugs and psychotropic substances" and "computer-related crime" demonstrate partial matches, with the national legislation omitting specific qualifiers or adopting modern but equivalent terminologies, e.g. cybercrime. Furthermore, "environmental crime, including the trafficking of endangered species" is articulated with slightly different but substantively identical terms. The model further demonstrates its ability to identify the corresponding national definitions of concepts, specifying the segments of legislative text where they are located. Most interestingly, the column titled "notes" reveals the argumentative capacity of LLMs, containing considerations and reasoning on the degree of semantic overlap between concepts. Examples highlighted in Table 3 serve as a perfect illustration of the overall results of semi-automatic concept retrieval and alignment.

The results of the LLM-based alignment process between European and national concepts were manually evaluated. Fifty pairs of European legislative paragraphs and their corresponding domestic transpositions were randomly selected, from which the corresponding European concepts contained in the HG were extracted. This process generated a list of 200 concept instances for annotation, of which only 135 were unique (non-duplicate) entries. Although the model was explicitly tasked with reporting all European concepts in the output table, the LLM failed to retrieve 57 out of 200 instances. The following discussion examines the annotation results (Fig. 7) by separately addressing the four questions outlined in the guidelines D and elaborating on the examples available in Table 4.

The model demonstrated 98% accuracy in classifying concepts as explicit or implicit (Q1). The model successfully distinguished between concepts explicitly enshrined in legislative sources and those serving primarily taxonomic functions within the glossary's semantic framework. For instance, as illustrated in Table 4, the right to a properly qualified interpreter is explicitly transposed in the Portuguese implementing measure, which also includes additional specifications regarding language proficiency and the free provision of the service. On the contrary, the concept of "temporary transfer", which is explicitly laid down in the EU legislative text, lacks transposition in the corresponding national legislative segment. Despite this, the model erroneously categorises the concept as implicit.

Table 3
Alignment of European and national corresponding concepts from Article 2 FD EAW and French implementing legislation.

European concept	Corresponding national concept	Nature of concept	Type of concept match	Corresponding national definition	Notes
European arrest warrant	European arrest warrant	Explicit	Semantic exact match	[“Under the law of the issuing Member State, the following offences may give rise to a European arrest warrant:”, “The execution of a European arrest warrant may also be refused if the offence for which the warrant is issued does not constitute an offence under French law”.]	The concept is directly mentioned in the national legislation.
Detention order for a maximum period of at least 12 months	No	Implicit	Not applicable	[ ]	The Concept is not explicitly mentioned in the national legislation, but may be implied in the custodial sentence provisions.
Illicit trafficking in narcotic drugs and psychotropic substances	Trafficking in narcotic drugs and psychotropic substances	Explicit	Semantic partial match	[“5° Trafficking in narcotic drugs and psychotropic substances;“]	The national legislation omits the word “illicit”, but the meaning is implied.
computer-related crime	Cybercrime	Explicit	Semantic partial match	[“11°Cybercrime”;]	The national legislation uses a more modern term, but the concept is essentially the same.
Environmental crime, including illicit trafficking in endangered animal species and in endangered plant species and varieties	Crimes and offences against the environment, including trafficking in endangered animal species and endangered plant species and essences	Explicit	Semantic partial match	[“12°Crimes and offences against the environment, including trafficking in endangered animal species and endangered plant species and essences”;]	The national legislation uses slightly different wording but covers the same concept.

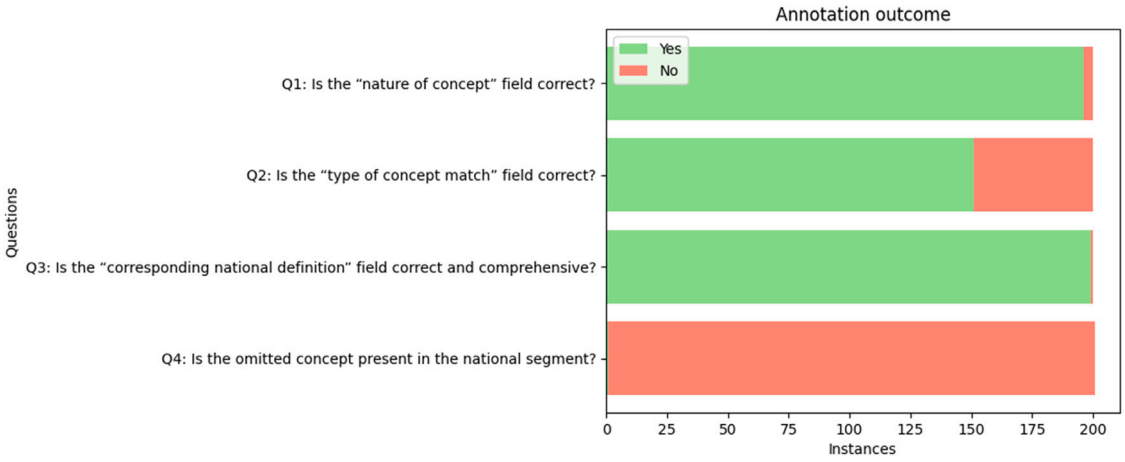


Fig. 7. Results of the annotations on questions Q1, Q2, Q3 and Q4.

The model showed suboptimal performance in identifying the “type of concept match” (Q2), achieving an accuracy of 75.5%. This task implicitly requires the model to determine the presence or absence of European concepts within national implementing measures, where “Not applicable” indicates concept absence, while “Semantic exact match” and “Semantic partial match” indicates its transposition. Manual annotation confirmed correct recognition of 151 out of 200 EU concept instances. Among the 49 errors committed by the model, 45 instances were incorrectly classified as “Semantic partial match”, 3 as “Semantic exact match”, and 1 as “Not applicable”. More precisely, the model accurately identified 97,12% of exact matches, 32,84% of partial matches, and 96,55% of non-matches. The model tends to misclassify partial matches, indicating that errors mostly occur in ambiguous rather than in straightforward cases, suggesting minimal hallucination behaviour.

Despite the model’s ability to identify nuanced semantic overlaps and discrepancies, results show failure to meet legal analytical standards, as semantic correspondence evaluation produced inconsistent and inexplicable outcomes. To address this shortcoming, the annotator implemented the category of partial semantic match. Partial matches, showcasing potential inconsistent legal interpretations across jurisdictions, were classified as positive instances of partial matches. On the contrary, mere semantic variations were designated as negative instances and categorised as full matches. For instance, the concept “expenses of the executing member state” under Article 30 FD EAW was classified to only partially match its Portuguese correspondence “expenses incurred in the national territory”. The Portuguese version leverages “national territory” to contextualise the concept, representing a necessary adaptation of EU legislative terminology to national legal frameworks without generating legally divergent interpretations. On the contrary, the concept “possible revocation of the right to consent to surrender” from Article 13 FD EAW was classified as a partial match with the Portuguese “consent to surrender may not be revoked”, which denotes “non-revocation of the consent to surrender”. In this case, the national formulation explicitly prohibits revocation of consent, substantially differing from the EU version, which allows for possible revocation.

Regarding the identification of corresponding national transpositions (Q3), the model accurately identified the textual sections containing each concept in 99.5% of cases, confirming the precision of the manually-developed version of the HG. The first case use-case example shows correspondence between the national the European concepts. The EU concept outlined in Article 29 of the FD EAW, "property as evidence", is indeed clearly mirrored in the German implementing measure found in Article 38 of the Act on International Cooperation in Criminal Matters, which uses the phrase "property which may serve as evidence". On the contrary, in a second example, the model failed to identify a corresponding national concept. The concept "purpose of prosecution of a custodial sentence or detention order" is indeed codified in Article 16 of the FD EAW, whereas the Italian implementing legislation mentions "criminal proceedings" and "execution of a penalty or security measure involving deprivation of liberty", and does not explicitly use the terms "custodial sentence" or "detention order". This partial alignment prevented the model to correctly identify the corresponding national concept.

Despite the prompt instruction to report both transposed and absent concepts in the output, the annotation showed that all of the omitted 57 concepts (Q4) in the model's outputs were not actually present in national transpositions. This indicates that the model recognised non-existing concepts in national sources, thus avoiding alignment errors during HG enrichment. The evaluation showed that the concepts were effectively omitted because of their absence from the text of the source material under scrutiny.

In conclusion, the manual evaluation yielded promising results for the HG enrichment, suggesting opportunities for future research, including case study expansion, increased annotator participation, and multilingual extension incorporating original-language national legislation. These findings facilitate deeper analysis of European and national legislation, possibly supporting policymakers and enabling new experiments through the provision of a gold standard. This informed, semi-automated resource delivered temporal advantages since the alignment with member states occurred automatically.

6. Limitations

This section highlights important considerations regarding our research methodology and findings, providing context for the practical application and future development of our approach.

Our approach, while demonstrating promising results, faces several limitations. Processing extensive legal texts with large context window models requires substantial computational resources, creating challenges when scaling to handle the full corpus of EU and member states' legislation. As this legislative corpus continues to expand across the EU's 27 member states and increasingly diverse domains, computational demands grow exponentially. Future implementations could benefit from selective prompt creation methodologies that optimise resource utilisation through prompting mechanisms identifying the most relevant textual segments, thereby reducing computational overhead while maintaining analytical integrity.

The current implementation relies on human expert involvement for validation and refinement. While this complementary relationship between AI systems and human legal expertise represents a strength in addressing complex legal concepts requiring interpretative judgment and in increasing the overall reliability of the system, it also indicates possible scalability challenges. This dependency becomes particularly pronounced when dealing with complex legal concepts requiring jurisprudential context or novel legal domains with limited precedent, highlighting the need for domain specialists across multiple jurisdictions and languages.

The linguistic and jurisdictional diversity of the EU presents challenges since legal terminology carries different procedural or substantive implications across national legal frameworks, requiring our approach to be considered language- or jurisdiction-dependent. This diversity reflects fundamental differences in legal traditions and interpretative approaches across member states, potentially resulting in inconsistencies in concept alignment accuracy. This necessitates tailored approaches that respect jurisdiction-specific nuances while working towards harmonisation, making the system's evaluation even more reliant on specialized human oversight.

From a computational perspective, our current implementation leverages a proprietary model, limiting research reproducibility. The "black box" nature of LLMs makes it difficult to understand precisely how legal concepts are being identified and aligned, potentially limiting trust in legal applications. Integrating open weights LLMs presents an opportunity to enhance both reproducibility and specialized performance, potentially improving understanding of how these models identify and align legal concepts thus increasing the system's reliability.

Our evaluation methodology, though rigorous within its scope, focuses primarily on a limited subset of legal domains and languages, which may not generalise across all EU member states. The lack of standardised benchmarks for legal concept alignment further complicates comparative assessment against other methodologies. Developing a more comprehensive evaluation framework with larger datasets spanning more jurisdictions would enhance generalisability, as suggested by our preliminary testing with expanded annotator groups. Furthermore, while proving the ever-increasing LLMs' performance in complex domain-specific tasks, the high scores obtained in our experiments can be partly attributed to the fact that the same domain expert who created the HG also conducted the evaluation task.

An important aspect that requires further consideration concerns the public perception of automated systems in the legal domain. The trust of legal practitioners and citizens towards AI technologies applied to law could influence the adoption and success of our methodology. Recent studies highlight how transparency in automated decision-making processes and understanding the complementary role of AI relative to human expertise are determining factors for the acceptance of these tools [52]. Furthermore, implementing this technology across diversified legal and cultural contexts within the EU presents challenges. Differences in national legal systems, interpretative traditions, and cultural attitudes towards technological innovation may require ad hoc approaches for each member state, further complicating the scalability of the proposed solution. Continuous dialogue with national authorities and European institutions will be necessary to ensure our methodology respects the peculiarities of each legal system while maintaining the objective of conceptual harmonisation at the European level.

These limitations show promising directions for continued research to refine our methodology, expand our evaluation frameworks, and develop increasingly efficient approaches to legal concept extraction and alignment within the European Union legal context.

7. Conclusions and future work

This study demonstrated that large-scale models, specifically Claude 3.5 Sonnet 200k, possess remarkable abilities to extract and align legislative concepts from national implementation norms to those contained in European legislative acts. These results represent a significant advance towards achieving a primary research goal, which is the semi-automatic alignment between European legislative articles and their corresponding national implementations, a key component in the computational analysis of legal harmonisation in the European Union.

To further advance this research and maintain a multidisciplinary perspective, several future research lines are proposed, focusing on enhancing multilingual legal analysis. A key step is addressing the multilingualism of European and national legislation by applying the research methodology to a multilingual dataset and carefully evaluating the results. Central to this is expanding the HG to facilitate broader cross-lingual comparisons that transcend traditional translation-dependent methods. The expanded HG will enable the exploration of concept-based similarity measures between legal texts, allowing for the

Table 4
Manual annotation examples of the LLM output.

Question	Instance type	Example
Q1	Correct	EU concept in Art. 11 FD EAW: "right to be assisted by an interpreter" Corresponding Portuguese concept found in Art. 17, Law on the European Arrest Warrant (Law no. 65/2003, of 23 August): "right to a properly qualified interpreter" Corresponding Portuguese text: ["If the arrested person does not adequately understand or speak Portuguese, a properly qualified interpreter shall be provided free of charge".] Nature of concept: explicit
	Wrong	EU concept in Art. 18 FD EAW: "temporary transfer" Corresponding German concept: "no" Corresponding German text: [] Nature of concept: implicit
Q2	Correct	EU concept in Art. 13 FD EAW: "possible revocation of the right to consent to surrender" Portuguese concept found in Art. 20, Law on the European Arrest Warrant (Law no. 65/2003, of 23 August): "consent to surrender may not be revoked" Type of concept match: Semantic partial match
	Wrong	EU concept in Art. 30 FD EAW: "expenses of the executing member state" Portuguese concept found in Art. 35, Law on the European Arrest Warrant (Law no. 65/2003, of 23 August): "expenses incurred in the national territory" Type of concept match: Semantic partial match
Q3	Correct	EU concept in Art. 29 FD EAW: "property as evidence" Corresponding German concept found in Art. 38, Act on International Cooperation in Criminal Matters: "property which may serve as evidence" Corresponding German text: ["property which may serve as evidence in the foreign proceedings"]
	Wrong	EU concept in Art. 16 FD EAW: "purpose of prosecution of a custodial sentence or detention order" Corresponding Italian concept found in Art. 2, Provisions to adapt national legislation to the Council Framework Decision 2002/584/JHA of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States: "warrants were issued in the course of criminal proceedings or for the execution of a penalty or security measure that involves a deprivation of liberty" Corresponding Italian text: ["1. When two or more Member States have issued a European arrest warrant against the same person, the Court of Appeal shall decide which arrest warrant shall be executed, taking into account all relevant factors of assessment and, in particular, the seriousness of the offences for which the warrants were issued, the place where the offences were committed and the dates on which the arrest warrants were issued and considering, in this context, whether the warrants were issued in the course of criminal proceedings or for the execution of a penalty or security measure that involves a deprivation of liberty".]

identification of conceptual convergence or divergence across jurisdictions, even in the absence of direct translations. Further exploration of semi-automatic concept extraction through LLMs, benchmarked against a "gold standard" of manually extracted concepts from diverse legislative areas, will be crucial for refining the HG and ensuring its accuracy. These advancements will empower researchers and legal professionals to gain a more nuanced understanding of legal harmonisation within multilingual settings like the European Union.

In conclusion, it is important to acknowledge that the analysis of legal harmonisation extends beyond the mere semantic-textual matching of legislative texts. The effectiveness of implementation may also depend on non-legislative legal sources, such as jurisprudential decisions. Although effective methodologies already exist for network analysis and semantic modelling of jurisprudential content, integrating these with large language models represents an important new frontier of research yet to be explored.

While Hicks et al. [53] argue that LLMs produce useless outputs due to their inherent indifference to truth, this perspective overlooks the potential for these models to be effectively employed with appropriate prompting, grounding techniques, and human oversight. The authors focus on the tendency of LLMs to generate plausible-sounding yet inaccurate information, but they fail to adequately address emerging research demonstrating that careful guidance and integration with external knowledge sources can significantly enhance the reliability and usefulness of LLM outputs. Therefore, as this research article also aims to suggest, the blanket assertion that LLMs are inherently "bullshit" is an oversimplification that ignores the evolving landscape of LLM capabilities and their potential for productive application.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the EU-Next Generation EU, Mission 4, Component 2, CUP: J33C22002830006 (ExtrAI).

Appendix A. System prompt

You are an expert in European legislation and the corresponding national implementing legislation of member states. You have a deep

knowledge of legal concepts that appear both in European and national legislation. Notably, you are familiar with a non-formal ontology of European legal concepts that has been manually developed by legal experts. These experts have linked these concepts to specific segments of European legislation, including the Framework Decision on the European Arrest Warrant (EU Decision 2002/584).

This is an example: European legislation segment: "1. A European arrest warrant may be issued for acts punishable by the law of the issuing Member State by a custodial sentence or a detention order for a maximum period of at least 12 months or, where a sentence has been passed or a detention order has been made, for sentences of at least four months". European legal concepts found: ['european arrest warrant', 'act punishable by the law of the issuing member state', 'custodial sentence for a maximum period of at least 12 months', 'detention order for a maximum period of at least 12 months', 'sentence of at least four months']

In a European research project, tables have been manually crafted to show the alignment between the provisions of the Framework Decision on the European Arrest Warrant and the corresponding national laws implemented in various European countries. The goal is to automate the analysis of legislative harmonisation across Europe, defined as the lexical and semantic alignment between European legislation and its national implementations. Your task is to identify correspondences between concepts in European laws and their equivalents in national legislation. Given articles and concepts from European laws, you will match them with the appropriate national legislation, noting the presence of identical concepts or their synonyms.

Appendix B. Prompt for legal concept detection

Create a mapping table of all European concepts manually extracted by legal experts to assess how accurately they have been transposed into national law. More precisely, identify possible concepts in national legislation that correspond to European concepts. Output in this following format: "| european concept | corresponding national concept | nature of concept | type of concept match | corresponding national definition | notes |".

The "european concept" column contains concepts that are extracted from the European legislative texts. Use the example provided above to identify concepts under scrutiny.

The "corresponding national concept" column can have as a specification either no or the corresponding concept found in the national transposing legislative segments.

The "nature of concept" column can have values either "explicit" or "implicit". It refers to the presence, if any, explicit or implicit of the concept in national legislation

The "type of concept match" column could have the following values that explains the type of match between "european concept" and "corresponding national concept" columns:

- *Not applicable*
- *Semantic exact match*: all meanings of the european legal concept are present in the national legislation, differences such as verb conjugations, stopwords, numbers, or other textual components with weak meaning are to be ignored
- *Semantic partial match*: some of the European legal concept meanings are present in the national legislation but substantial differences were omitted from the perspective of legal language, e.g. the presence of "illicit", "may", "obligation", "right", etc.

The "corresponding national definition" column contains a list of national sentences, which are national legislative portions, that contain the "corresponding national concept". A concept could be found in more legislative portions.

The "notes" column contains an explanation of the concept match found.

I give you two examples of table rows.
Example 1:

- *"european concept"*: custodial sentence for a maximum period of at least 12 months
- *"corresponding national concept"*: custodial sentence of one year or more
- *"nature of concept"*: explicit
- *"type of concept match"*: Semantic partial match
- *"corresponding national definition"*: ["2° Offences punishable by a custodial sentence of one year or more or, where a custodial sentence has been pronounced, where the sentence pronounced is four months' imprisonment or more".]

Example 2:

- "european concept": offence giving rise to surrender pursuant to an european arrest warrant
- *"corresponding national concept"*: offence giving rise to execution of european arrest warrant
- *"nature of concept"*: implicit
- *"type of concept match"*: Semantic partial match
- *"corresponding national definition"*: ["By way of derogation from the first paragraph, a European arrest warrant shall be executed without verification of the double criminality of the offences for which it is issued when those offences are, under the law of the issuing Member State, punishable by a custodial sentence of at least three years' imprisonment or by a detention order of a similar duration and fall within one of the categories of offences provided for in Article 694-32".]

Now compute the table given the following European legal concepts and the national legislative segments.

Concepts manually extracted by legal experts in the European legislative segment:
"""
[concept 1, concept 2, concept 3, ...]
"""

Corresponding legislative segments of national implementation in list format:
"""
[national implementation portion 1, national implementation portion 2, national implementation portion 3, ...]
"""

"| european concept | corresponding national concept | nature of concept | type of concept match | corresponding national definition | notes |".

Appendix C. Prompt input with instance example

Create a mapping table of all European concepts manually extracted by legal experts to assess how accurately they have been transposed into national law. More precisely, identify possible concepts in national legislation that correspond to European concepts. Output in this following format: "| european concept | corresponding national concept | nature of concept | type of concept match | corresponding national definition | notes |".

The "european concept" column contains concepts that are extracted from the European legislative texts. Use the example provided above to identify concepts under scrutiny.

The "corresponding national concept" column can have as a specification either no or the corresponding concept found in the national transposing legislative segments.

The "nature of concept" column can have values either "explicit" or "implicit". It refers to the presence, if any, explicit or implicit of the concept in national legislation

The "type of concept match" column could have the following values that explains the type of match between "european concept" and "corresponding national concept" columns:

- *Not applicable*
- *Semantic exact match*: all meanings of the european legal concept are present in the national legislation, differences such as verb conjugations, stopwords, numbers, or other textual components with weak meaning are to be ignored
- *Semantic partial match*: some of the European legal concept meanings are present in the national legislation but substantial differences were omitted from the perspective of legal language, e.g. the presence of "illicit", "may", "obligation", "right", etc.

The "corresponding national definition" column contains a list of national sentences, which are national legislative portions, that contain the "corresponding national concept". A concept could be found in more legislative portions.

The "notes" column contains an explanation of the concept match found.

I give you two examples of table rows.
Example 1:

- *"european concept"*: custodial sentence for a maximum period of at least 12 months
- *"corresponding national concept"*: custodial sentence of one year or more
- *"nature of concept"*: explicit
- *"type of concept match"*: Semantic partial match
- *"corresponding national definition"*: ["2° Offences punishable by a custodial sentence of one year or more or, where a custodial sentence has been pronounced, where the sentence pronounced is four months' imprisonment or more".]

Example 2:

- "european concept": offence giving rise to surrender pursuant to an european arrest warrant
- *"corresponding national concept"*: offence giving rise to execution of european arrest warrant
- *"nature of concept"*: implicit
- *"type of concept match"*: Semantic partial match
- *"corresponding national definition"*: ["By way of derogation from the first paragraph, a European arrest warrant shall be executed without verification of the double criminality of the offences for which it is issued when those offences are, under the law of the issuing Member State, punishable by a custodial sentence of at least three years' imprisonment or by a detention order of a similar duration and fall within one of the categories of offences provided for in Article 694-32".]

Now compute the table given the following European legal concepts and the national legislative segments.

Concepts manually extracted by legal experts in the European legislative segment:
"""
['property acquired through the offence', 'rights acquired in the property', 'death or escape', 'seizing and handing over property', 'property liable to seizure or confiscation', 'return of property', 'property as evidence']
"""
Corresponding legislative segments of national implementation in list format:
"""
['At the request of the issuing judicial authority or on the initiative of the competent authorities, the competent court for the execution of the European arrest warrant shall order the seizure and handing over to the issuing judicial authority of property which:', 'May be required as evidence;', 'At the request of the issuing judicial authority or on the initiative of the competent authorities, the competent court for the execution of the European arrest warrant shall order the seizure and handing over to the issuing judicial authority of property which:', 'Has been acquired by the requested person as a result of the offence'., 'At the request of the issuing judicial authority or on the initiative of the competent authorities, the competent court for the execution of the European arrest warrant shall order the seizure and handing over to the issuing judicial authority of property which:', 'The property referred to in the preceding paragraph shall be handed over to the issuing judicial authority even if the European arrest warrant cannot be carried out owing to the death or escape of the requested person'., 'Where the property referred to in paragraph 1 is liable to seizure or confiscation, it may, if it is needed in connection with pending criminal proceedings in Portugal, be temporarily retained or handed over to the issuing Member State, on condition that it is returned'., 'The rights acquired by the Portuguese State or third parties over the property referred to in paragraph 1 shall be preserved'., 'In the situation mentioned in the preceding paragraph, the issuing Member State shall return the property that was seized and handed over without charge to the executing Member State as soon as the criminal proceedings have been terminated'.]
"""

"| european concept | corresponding national concept | nature of concept | type of concept match | corresponding national definition | notes |".

Appendix D. Annotation guidelines

This document outlines the guidelines for a manual annotation task designed to evaluate the performance of an automated system for detecting European legal concepts within Member State legislation. The results of this annotation will be used to assess the potential of this automation to assist policymakers and legal practitioners in evaluating the effectiveness of European legislation.

D.1. Task context

European legislation is implemented at the national level by each Member State. To facilitate analysis, a matching table has been created by legal experts. This table links specific articles in European legislation to their corresponding implementing articles within the national laws of each Member State. This pre-existing mapping table constitutes the ground truth for national legislative implementations and will be provided to support the annotation and evaluate the automatically generated results.

D.2. Task objective

An automated system has been developed to identify specific legal concepts within legislative texts. Using the expert-created matching table, your task is to evaluate the output of this automated system. This evaluation will help us understand the strengths and weaknesses of the automated legal concept detection process and its potential utility in real-world scenarios.

D.3. Annotation procedure

In this section, you can find information about the inputs, tasks, and tools to use to complete the annotation.

D.3.1. Input

You will be provided with the following information:

1. A specific article from European legislation.
2. The corresponding national implementing articles (as identified by the expert matching table).
3. The automatically generated output table contains the following:

   (a) *European Concept*: A legal concept manually annotated from the Council Framework Decision 2002/584/JHA on the European Arrest Warrant.
   (b) *Corresponding National Concept*: The corresponding national concept, if found.
   (c) *Nature of Concept*: Whether the concept is implicit or explicit.
   (d) *Type of Concept Match*: Indicates whether the match is a semantic exact match, a semantic partial match, or not applicable.
   (e) *Corresponding National Definition*: Text segments within the national implementing articles where European legal concepts should be present.
   (f) *Notes*: A generated comment explaining the reasoning behind the row's generation.

D.3.2. Task 1

Your task is to assess the accuracy of the automated system's output for each provided pair of European and national implementing articles. Specifically, you will answer three questions for each table row with "Yes" or "No" values:

1. *Q1: Is the "nature of concept" field correct?*
   - *Yes*: after reading the national implementation, determine whether the European concept is present explicitly, implicitly, or not at all. Then, compare your assessment with the automated system's output. If your assessment matches the system's output, the answer is considered correct.
   - *No*: otherwise
2. *Q2: Is the "type of concept match" field correct?*
   - *Yes*: the automatic process identifies European concepts in national implementations as either a "semantic match" (partial or exact) or "not applicable". A "semantic match" is correct if the concept is present; "not applicable" is correct if it's absent.
   - *No*: otherwise
3. *Q3: Is the "corresponding national definition" field correct and comprehensive?*
   - *Yes*: the system's output is considered correct if it accurately identifies the presence or absence of the European legal concept derived from the matched European article. Furthermore, if the concept is present, the list of corresponding national legislative portions must be accurate and complete for the output to be considered correct.
   - *No*: otherwise

D.3.3. Task 2

As the automatic system sometimes ignored some European legal concepts, your task is to answer the following question for those omitted:

1. *Q4: Is the omitted concept present in the national segment?*
   - *Yes*: after reading the national implementation, you found that the concept omitted by the system is actually present in the national implementation
   - *No*: otherwise

D.3.4. Annotation tool

You will be provided with two spreadsheets, the first for Task 1 and the second for Task 2. Respond with "yes" or "no" in the cells corresponding to the European concept and the question you are answering.

D.3.5. Important considerations

1. *Ambiguity*: If you encounter any ambiguity in the automated system's output or the expert matching, please document it clearly in the provided space for comments.
2. *Consistency*: Maintain consistency in your annotations throughout the task. If you are unsure about a specific case, refer back to these guidelines or consult with the research team.

By carefully following these guidelines, you will contribute valuable data to this research project, ultimately helping us to better understand how automated systems can support the analysis of European legislation and its implementation across Member States.

We thank you for participating in this experiment.

References

[1] Janssens C. The principle of mutual recognition in EU law. OUP Oxford; 2013.
[2] Regan E. The role of the principles of mutual trust and mutual recognition in EU law. Il Dirit Unione Eur 2018;2:231.
[3] Weatherill S. Maximum versus minimum harmonization: Choosing between unity and diversity in the search for the soul of the internal market. In: From Single Market to Economic Union: Essays in Memory of John A Usher. 2012, p. 175–99.
[4] De Cecco F. Minimum harmonization and the limits of union fundamental rights review: TSN and AKT. Common Mark Law Rev 2021;58(1).
[5] EdjNet and Openpolis. Understanding EU infringement procedures through data. 2023, Available online: https://www.openpolis.it/infringeye-il-monitoraggio-delle-infrazioni-europee-con-i-dati/. [Accessed 19 September 2024].
[6] European Data Journalism Network. Infringeye. Pending infringement procedures by country and policy area. 2021, https://www.europeandatajournalism.eu/resources-for-journalists/infringeye/. [Accessed 30 September 2024].
[7] European Commission. 2023 annual report on monitoring the application of EU law. 2023, URL https://commission.europa.eu/law/application-eu-law/implementing-eu-law/infringement-procedure/2023-annual-report-monitoring-application-eu-law_en.
[8] Bratanić M, Lončar M. The myth of EU terminology harmonization on national and EU level. In: Language and culture in EU law. Routledge; 2016, p. 207–18.
[9] Andenas M, Andersen CB. Theory and practice of harmonisation. Edward Elgar Publishing; 2012.
[10] Chen Z, Mao H, Li H, Jin W, Wen H, Wei X, Wang S, Yin D, Fan W, Liu H, Tang J. Exploring the potential of large language models (LLMs)in learning on graphs. ACM SIGKDD Explor Newsl 2024;25:42–61. http://dx.doi.org/10.1145/3655103.3655110.
[11] Sulis E, Humphreys LB, Audrito D, Di Caro L. Exploiting textual similarity techniques in harmonization of laws. In: Bandini S, Gasparini F, Mascardi V, Palmonari M, Vizzari G, editors. AIxIA 2021 – advances in artificial intelligence. Cham: Springer International Publishing; 2022, p. 185–97.
[12] Nanda R, Siragusa G, Di Caro L, Boella G, Grossio L, Gerbaudo M, Costamagna F. Unsupervised and supervised text similarity systems for automated identification of national implementing measures of European directives. Artif Intell Law 2019;27:199–225.
[13] Nanda R, Humphreys L, Grossio L, Kolawole John A. Multilingual legal information retrieval system for mapping recitals and normative provisions. In: Legal knowledge and information systems. IOS Press; 2020, p. 123–32.
[14] Ferrod R, Bondarenko DA, Audrito D, Siragusa G. Pairing EU directives and their national implementing measures: A dataset for semantic search. Comput Law Secur Rev 2023;51:105862.
[15] Magesh V, Surani F, Dahl M, Suzgun M, Manning CD, Ho DE. Hallucination-free? Assessing the reliability of leading AI legal research tools. 2024, http://dx.doi.org/10.48550/ARXIV.2405.20362, CoRR abs/2405.20362 arXiv:2405.20362.

[16] Baaij CJ. The role of legal translation in legal harmonization. Kluwer Law International BV; 2012.
[17] Ruiz-Cortés E. Legal-linguistic analysis of EU law and its transpositions: A useful approach to explore harmonisation problems? J Lang Law/Rev Lleng I Dret 2020;(74).
[18] European Commission. Communication to the Commission. Translation as part of the Commission's decision-making process. 2016, https://www.asktheeu.org/en/request/4878/response/15921/attach/6/C%202016%202000.PDF.pdf?cookie_passthrough=1. [Accessed 30 September 2024].
[19] Ghio E. Harmonisation and European integration in times of crisis. In: Are legal systems converging or diverging? lessons from contemporary crises. Springer; 2024, p. 255–76.
[20] Satzger H. The harmonisation of criminal sanctions in the European Union - A new approach. Eucrim 2019. http://dx.doi.org/10.30709/eucrim-2019-007.
[21] Schroeder W. Limits to European harmonisation of criminal law. Eucrim 2020. http://dx.doi.org/10.30709/eucrim-2020-008.
[22] Vogenauer S, Weatherill S. The Harmonisation of European Contract Law: Implications for European Private Laws, Business and Legal Practice. Bloomsbury Publishing; 2006, http://dx.doi.org/10.1111/j.1468-0386.2007.00376_4.x.
[23] Weatherill S. Harmonisation: How much, how little. Eur Bus L. Rev 2005;16:533.
[24] Voigt P, von dem Bussche A, Voigt P, von dem Bussche A. Practical implementation of the requirements under the GDPR. EU Gen Data Prot Regul (GDPR) A Pr Guid 2017;245–9.
[25] Nanda R, Di Caro L, Boella G, Konstantinov H, Tyankov T, Traykov D, Hristov H, Costamagna F, Humphreys L, Robaldo L, et al. A unifying similarity measure for automated identification of national implementations of European union directives. In: Proceedings of the 16th edition of the international conference on articial intelligence and law. 2017, p. 149–58.
[26] Audrito D, Sulis E, Humphreys L, Di Caro L. Analogical lightweight ontology of EU criminal procedural rights in judicial cooperation. Artif Intell Law 2023;31(3):629–52.
[27] Vaswani A. Attention is all you need. Adv Neural Inf Process Syst 2017.
[28] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: North American chapter of the association for computational linguistics. 2019, URL https://api.semanticscholar.org/CorpusID:52967399.
[29] OpenAI, et al. GPT-4 technical report. 2024, arXiv:2303.08774.
[30] META. Introducing meta llama 3: The most capable openly available LLM to date. 2024, URL https://ai.meta.com/blog/meta-llama-3/.
[31] Gemini Team, et al. Gemini: A family of highly capable multimodal models. 2024, arXiv:2312.11805 URL https://arxiv.org/abs/2312.11805.
[32] The Claude 3 model family: Opus, Sonnet, Haiku, URL https://api.semanticscholar.org/CorpusID:268232499.
[33] Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, Barnes N, Mian AS. A comprehensive overview of large language models. 2023, ArXiv abs/2307.06435 URL https://api.semanticscholar.org/CorpusID:259847443.
[34] Sun Z. A short survey of viewing large language models in legal aspect. 2023, arXiv preprint arXiv:2303.09136.
[35] Lai J, Gan W, Wu J, Qi Z, Yu PS. Large language models in law: A survey. 2023, arXiv preprint arXiv:2312.03718.
[36] Lexroomai. Revolutionizing legal research: Lexroom.ai's AI-Driven solution. 2024, https://go-dive.net/revolutionizing-legal-research-lexroom-ais-ai-driven-solution/. [Accessed 28 September 2024].
[37] Higashi K, Nakagawa Z, Yamada T, Mori H. Automated harmonization and Large-Scale integration of heterogeneous biomedical sample metadata using large language models. BioRxiv 2024. URL https://api.semanticscholar.org/CorpusID:273776393.
[38] Dhavale C, Bharambe U, Shinde S, Murkute A. Enhancing spatiotemporal data harmonization through semantic enrichment using SciBERT LLM for soil analytic. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. 2024, p. 6871–4, URL https://api.semanticscholar.org/CorpusID:272432543.
[39] Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. In: EMNLP. Association for Computational Linguistics; 2019, URL https://www.aclweb.org/anthology/D19-1371.
[40] Delleani M, D'amico S, Sauta E, Asti G, Zazzetti E, Campagna A, Lanino L, Maggioni G, Grondelli MC, Barrero AF, Morandini P, Ubezio M, Todisco G, Russo A, Tentori CA, Buizza A, Bonometti A, Lancellotti C, di Tommaso L, Rahal D, Bicchieri M, Savevski V, Santoro A, Santini V, Solé F, Platzbecker U, Fenaux P, Díez-Campelo M, Komrokji RS, Garcia-Manero G, Haferlach T, Kordasti S, Zeidan AM, Castellani GC, Porta MGD. The "David Vs Goliath" study: Application of large language models (LLM) for automatic medical information retrieval from multiple data sources to accelerate clinical and translational research in hematology. Blood 2024. URL https://api.semanticscholar.org/CorpusID:274565282.
[41] Li Z, Prabhu SP, Popp Z, Jain SS, Balakundi V, Ang TFA, Au R, Chen J. A natural language processing approach to support biomedical data harmonization: Leveraging large language models. 2024, ArXiv abs/2411.02730 URL https://api.semanticscholar.org/CorpusID:273821518.
[42] Santos ASR, Pena EHM, López R, Freire J. Interactive data harmonization with LLM agents. 2025, ArXiv abs/2502.07132 URL https://api.semanticscholar.org/CorpusID:276259466.
[43] Cheong I, Xia K, Feng KJK, Chen QZ, Zhang AX. (A)I am not a lawyer, But...: Engaging legal experts towards responsible LLM policies for legal advice. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. 2024, URL https://api.semanticscholar.org/CorpusID:267413187.
[44] Nanda R, Di Caro L, Boella G. A text similarity approach for automated transposition detection of European union directives. In: International conference on legal knowledge and information systems. 2016, URL https://api.semanticscholar.org/CorpusID:32838837.
[45] Klimas T, Vaiciukaite J. The law of recitals in European community legislation. ILSA J Int Comp Law 2008;15:61–93, URL https://api.semanticscholar.org/CorpusID:143144202.
[46] Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, Mathur A, Schelten A, Yang A, Fan A, et al. The llama 3 herd of models. 2024, arXiv preprint arXiv:2407.21783.
[47] Jin Z, Lu W. Tab-CoT: Zero-shot tabular chain of thought. In: Rogers A, Boyd-Graber JL, Okazaki N, editors. Findings of the association for computational linguistics: ACL 2023, toronto, Canada, July 9-14, 2023. Association for Computational Linguistics; 2023, p. 10259–77. http://dx.doi.org/10.18653/V1/2023.FINDINGS-ACL.651.
[48] Wei J, Wang X, Schuurmans D, Bosma M, Chi E, Xia F, Le Q, Zhou D. Chain of thought prompting elicits reasoning in large language models. 2022, ArXiv abs/2201.11903.
[49] OpenAI. How to count tokens with Tiktoken. 2022, Available online: https://cookbook.openai.com/examples/how_to_count_tokens_with_tiktoken. [Accessed 10 April 2025].
[50] Neo4J website. 2019, https://neo4j.com/. [Accessed 30 September 2024].
[51] Sartor G, Santin P, Audrito D, Sulis E, Di Caro L. Automated extraction and representation of citation network: a CJEU case-study. In: International conference on conceptual modeling. Springer; 2022, p. 102–11.
[52] Gillespie N, Lockey S, Curtis C, Pool J, Akbari A. Trust in artificial intelligence: A global study, vol. 10, The University of Queensland and KPMG Australia; 2023.
[53] Hicks M, Humphries J, Slater J. ChatGPT is bullshit. Ethics Inf Technol 2024;26:1–10. http://dx.doi.org/10.1007/s10676-024-09775-5.