Retrieval-Augmented Question Answering using the EU AI Act

Purbasha Chowdhury[1], Alberto De Lazzari[2], Teresa Scantamburlo[3,4] and Paolo Falcarin[1,*]

[1]Ca' Foscari University of Venice, Department of Environmental Sciences, Informatics and Statistics Via Torino 155, 30172 Venice, Italy

[2]AKOS AI, Via Porzio 4, Centro Direzionale Isola E2, 80143, Napoli

[3]University of Trieste, via Economo 12/3, 34123 Trieste

[4]European Centre for Living Technology, Ca' Bottacin, Dorsoduro 3911, Calle Crosera, 30123 Venice, Italy

Abstract

Generative Artificial Intelligence systems have significantly improved question answering capabilities; however, applying these systems to legal documents remains challenging due to the complexity of legal semantics and the need for relevant regulatory provisions retrieval accuracy. Retrieval-Augmented Generation (RAG) has emerged as a promising approach to address the challenge by combining language models with external knowledge sources. In this paper, we investigate the use of LightRAG, a graph-based Retrieval-Augmented Generation framework, for question answering over the European Union AI Act. Additionally, we examine the impact of incorporating a domain-specific ontology to guide entity extraction and retrieval, comparing the performance of the framework with and without ontology-based guidance. The evaluation follows two steps: 1) First, it uses ten expert-reviewed legal questions, including their answers and supporting legal references. The retrieved legal references from both configurations are compared against the expert validated legal references. 2) Second, it uses ten other legal queries without answers and regulatory references. The regulatory provisions identified by two LightRAG configurations are compared qualitatively with the outputs of GraphReader to assess the retrieval characteristics of the three approaches.

Keywords

Artificial Intelligence, Retrieval Augmented Generation, Large Language Models (LLMs), Knowledge Modelling, Information Extraction, EU Directives, Ontology, Question Answering

CEUR-WS.org/Vol-4340/paper15.pdf

1. Introduction

The European Union digital strategies have led the publication of many new directives on information and software technologies: the Artificial Intelligence Act (AI Act), the Cyber-Resilience Act, the Network and Information Security Directive 2 (NIS2), and the Digital Service Act. Manual assessments performed by legal auditors are expensive and time-consuming [1]. The increasing deployment of AI systems in public and private organizations has amplified the need for organized, traceable, and reasonable compliance mechanisms. These activities emphasize the principle of information systems engineering tasks such as requirements elicitation, ensuring traceability, validating conformance, and merging legal constraints. AI raises concerns not only for data protection but also transparency [2], fairness [3], information privacy, and freedom of expression [4]. Consequently, the AI Act [5] is the world's first statutory law proposal for regulating AI systems, and it aims to ensure that AI systems in the EU are safe and respectful of fundamental rights and values. Its application extends beyond EU-based organizations and concerns any AI provider, importer, distributor, or authorized representative within the EU. The regulatory frameworks need tools able to deal with the length, cross-referential, complex semantics of legal texts.

The 25th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2026) - Satellite Events (Workshops, Tutorials, Demos, Posters)

*Corresponding author.

purbasha.chowdhury@unive.it (P. Chowdhury); alberto.delazzari@akos-ai.com (A. D. Lazzari); teresa.scantamburlo@units.it (T. Scantamburlo); paolo.falcarin@unive.it (P. Falcarin)

0009-0003-4022-9766 (P. Chowdhury); 0000-0002-6767-7396 (A. D. Lazzari); 0000-0002-3769-8874 (T. Scantamburlo); 0000-0003-1933-5348 (P. Falcarin)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-26

Existing NLP-based search methods [6] and generic Retrieval-Augmented Generation (RAG) systems [7] for legal compliance rely mainly on unstructured text retrieval limiting their ability to support structured regulatory reasoning.
The legal ontology modelling [8, 9] strategy has become a crucial framework[10] for identifying and structuring the legal concepts and relationships included in legal texts including statutory law, judicial decisions, soft-law etc. Legal ontologies are important, in the legal rule modelling, for providing a common vocabulary of predicates and thus to allow inter-connectivity between diverse legal knowledge engineers from different organizations. The Ontology Design Patterns (ODPs) [11, 12] method provides a formal template that structures the automatically extracted knowledge graph [13] to perform question-answering mechanism more effectively.
Retrieval-Augmented Generation (RAG) systems integrate functionalities of the large language models (LLMs) by integrating external knowledge sources. This innovative integration allows LLMs to generate more accurate and contextually relevant responses, significantly improving their utility in real-world applications. By adapting to specific domain knowledge, RAG systems ensure that the information provided is not only pertinent but also tailored to the user's needs. Furthermore, they offer access to up-to-date information, which is crucial in rapidly evolving fields. Chunking plays a vital role in facilitating the retrieval-augmented generation process. By breaking down a large external text corpus into smaller, more manageable segments, chunking significantly enhances the accuracy of information retrieval. This approach allows for more targeted similarity searches, ensuring that the retrieved content is directly relevant to user queries. However, current RAG systems have key constraints that limit their performance. First, conventional retrieval methods primarily depend on text similarity and may fail to capture the relationships between interconnected rigor. Second, the retrieved context may not always be substantial to support factual and well-grounded answers to the queries.
Existing legal QA systems have focused on various national and international regulations [14, 15, 16], including statutory laws and privacy policies, but have not yet directly addressed the recent European Regulation on AI [17]. In this paper, we present an approach for question answering using an European legal directive through legal knowledge extraction and retrieval. The approach incorporates **LightRAG** [18] as a retrieval framework and the European **AI ACT**[1] as case study to apply the discussed methodologies. The evaluation consists of two distinct datasets: 1. The first dataset consists of ten legal queries with their answers and related legal references. 2. The second dataset contains another set of ten legal queries without predetermined answers.
The paper is structured as follows: Section 2 represents related work, section 3 defines methodology, section 4 presents the implementation, section 5 demonstrates the datasets used and results, section 6 concludes the paper with some future directions.

2. Related Work

The advancement of legal question answering [19] has transformed from traditional information retrieval approaches to retrieval-augmented [20] and graph-based methods that integrate language models with structured legal knowledge. This evolution seeks to improve answer grounding and support the retrieval of relevant legal provisions.
Ontology-based compliance models have also been considered as a means of converting legal requirements into a machine-readable one [21]. Chennu et al. [22] extracted legal entities, relationships, and constraints using natural language processing (NLP) and ontology engineering [9] which are organized into an ontology and converted into a machine-readable knowledge graph, allowing manufacturers and stakeholders to visualize and get clear information about legal requirements easily. Legal ontologies [23] provide structured vocabularies and conceptual models for expressing normative concepts, obligations, actors, and exceptions.
Prior work emphasizes the need for ontological grounding to support interoperability and compliance automation, but constructing comprehensive legal ontologies remains resource-intensive and requires

[1] https://artificialintelligenceact.eu/

specialized expertise. Information extraction using machine learning and deep NLP has been applied to identify entities, obligations, definitions, and relationships within legal texts. The use of Large language models simplify the extraction of the context from the legal texts. BERTopic [24] creates dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions. Legal-Bert, a language model specialized for legal text [25], is a significant development in this field.[26] applied Bert-based approach to assess GDPR compliance in the web service policies, demonstrating how transformer models can enhance accuracy in legal text classification.

Retrieval-Augmented Generation (RAG) systems enrich LLM inputs by retrieving relevant information from external sources, leveraging accurate, domain-specific knowledge to inform responses [27, 28]. The existing RAG approaches [29, 30, 31] usually combine queries in a vector space to find the nearest context vectors. However, many of these techniques depend on fragmented text chunks and only retrieve a specific number of contexts, constraining their ability to capture comprehensive global information needed for effective responses. A recent work [32] explores graph-based knowledge structures, two major challenges remain. First, many graph-based approaches do not support dynamic updates or scalable expansion, making it difficult to incorporate new information over time. LightRAG [18], overcomes these limitations by supporting rapid and flexible graph updates, enabling the system to remain accurate and up-to-date.

Legal question answering (QA) [33, 34] systems aim to automatically retrieve relevant legal information in response to natural language queries. Traditional QA systems often follow IR-based systems that leverage classical information retrieval models, such as TF-IDF or Latent Semantic Analysis (LSA) [35], to produce vector representations of legal documents and rank them based on query similarity. While efficient, these models struggle with complex legal queries, as they prove to be ineffective in contextual understanding. Besides, Ontology-based systems ([36], [37], [38]) use manually crafted legal ontologies to improve query interpretation and retrieval accuracy. Although effective in capturing legal domain knowledge, they require significant expert input and struggle to adapt to latest or evolving regulatory texts. Neural systems use deep learning architectures such as Convolutional Neural Networks (CNNs, [39]) and transformer-based models (BERT, RoBERTa, GPT-based solutions, [40]) to model semantic relationships in legal texts. While these models demonstrate strong performance, they require substantial amounts of annotated data and typically lack interpretability.

Our work explores the use of Retrieval-Augmented Generation for legal question answering over the EU AI Act by comparing standard and ontology-guided retrieval configurations. We evaluate both configurations using expert-reviewed references and a qualitative comparison with GraphReader to analyze their ability to retrieve relevant Articles and Recitals from the regulation.

3. Methodology

3.1. Ontology

The widespread adoption of Artificial Intelligence (AI) has created new opportunities across industry, public services, and society. At the same time, concerns related to fairness, transparency, privacy, safety, and accountability have led to the development of regulatory frameworks governing the design, deployment, and use of AI systems. Among these, the European Union AI Act establishes a comprehensive legal framework based on a risk-oriented approach, defining obligations for providers, deployers, and other stakeholders according to the potential impact of AI systems. As the regulation contains numerous interconnected articles, recitals, and definitions, efficiently locating the relevant legal provisions remains a challenging task for practitioners, auditors, and researchers seeking to understand and apply its requirements. An ontology provides a formal organized representation of concepts, entities, and relationships within a specific domain, allowing the knowledge to be modelled in a logically explicit and machine-interpretable form. Ontologies play a crucial role in enabling interconnected and semantically linked data by assigning unique identities and comprehensive semantic meanings to domain concepts within the broader aspect of semantic web. Through this formalization, ontologies simplify interoperability, automated reasoning, knowledge integration, and semantic querying across

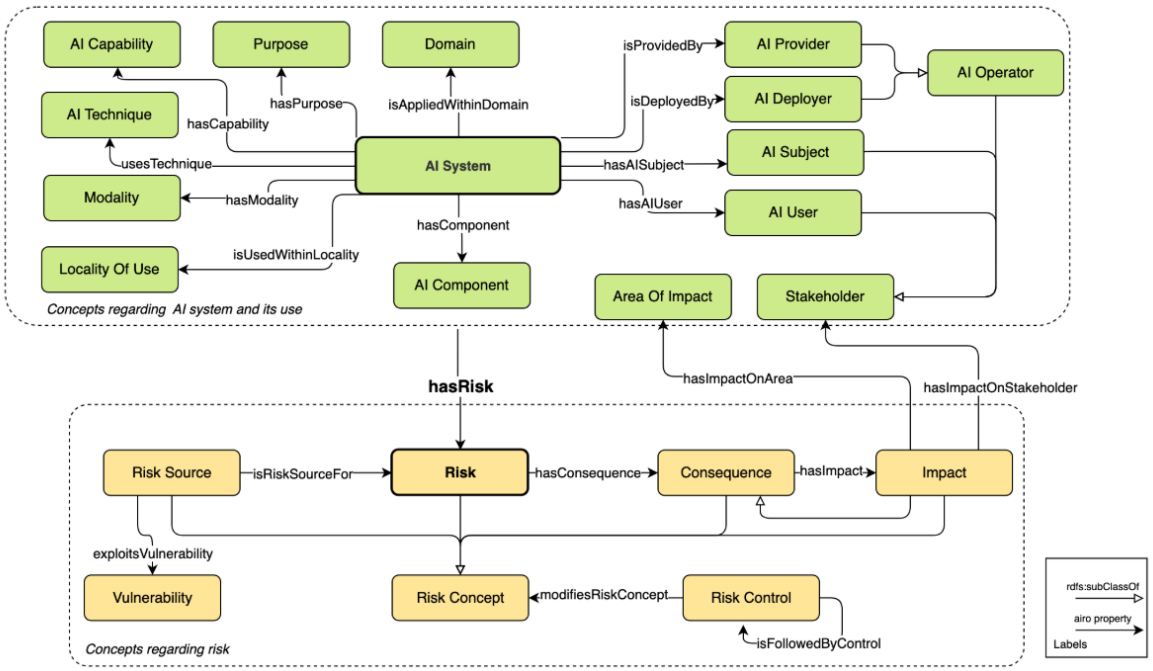


Figure 1: AIRO core concepts and relations. Reproduced from Golpayegani et al.[41], licensed under CC BY 4.0.

diverse systems.

However, ontology construction from scratch is a time-consuming and cumbersome task that requires substantial domain expertise, conceptual modelling, and validation. Consequently, we first started to investigate the availability of existing ontologies that could serve as a foundational semantic framework for extracting entities, relationships, and legal predicates related to the European Union AI Act. In this context, we identified the AIRO (AI Risk Ontology) [41, 42], which partially captures the regulatory domain addressed by the AI Act. AIRO is mainly designed based on the risk-oriented perspective which is a risk based-framework similar to the GDPR, the European privacy regulation. According to the ontology creators, "The purpose of AIRO is to express AI risks to enable organisations to represent their AI systems and the associated AI risks." This objective closely corresponds to one of the foundational dimensions of the EU AI Act, namely its risk-based classification and governance framework for artificial intelligence systems. Despite its relevance, AIRO does not comprehensively represent the full semantic scope of the AI Act.

3.2. Question-Answering

We leverage LightRAG [18] that facilitates rapid retrieval of relevant information from the graph through dual-level retrieval paradigm, significantly enhancing both retrieval efficiency and response speed. LightRAG enhances the retrieval system by segmenting documents into smaller, more manageable pieces. This strategy allows for quick identification and access to relevant information without analysing entire documents.

Compared with GraphRAG, the framework has demonstrated stronger performance on large and diverse datasets, including large legal corpora, particularly for complex queries that require connecting information across multiple entities and documents. Its retrieval strategy integrates entity-level and graph-level exploration, allowing the system to retrieve both corresponding legal provisions and broader contextual information. These characteristics make LightRAG a suitable framework for evaluating question answering over the EU AI Act, where legal concepts and regulatory provisions are closely interconnected.

The legal text is first preprocessed and divided into manageable chunks for indexing within the LightRAG

framework. These chunks are then processed to construct a knowledge graph and a vector index that together support hybrid retrieval. In addition to the standard LightRAG configuration, we evaluate a second configuration in which a domain-specific ontology (AIRO) is used to guide the entity extraction process during indexing. The ontology provides predefined legal concepts and relationships that assist the LLM in identifying entities and their semantic associations from the legal text to finally generate the responses with the relevant provisions. During question answering, LightRAG performs hybrid retrieval by combining semantic similarity search over the vector index with graph-based exploration of related entities and their connections.

4. Implementation

The approach is implemented in Python 3.10 leveraging the **LightRAG**[2]. The workflow starts from a collection of 113 articles and 180 recitals of the **EU AI Act explorer**[3]. By systematically inspecting [43] the HTML structure of the webpage, the articles and recitals are separated into individual files, where each file contains the complete textual content of a single article or recital. Each article and recital is stored as a separate JSON document, and the corresponding file path is preserved so that retrieved answers can be linked back to the relevant legal provision. The codes and generated answers including the retrieved legal references of both the frameworks are available in this **GitHub repository**[4]

4.1. Standard Configuration

The standard configuration is implemented using the LightRAG framework without any ontology-based semantic guidance. The corpus consists of the individual JSON documents of 113 articles and 180 recitals of AI Act. During indexing, LightRAG divides the legal documents into chunks of 512 tokens with an overlap of 256 tokens. The text chunks are embedded using the OpenAI embedding model text-embedding-3-small[5] and indexed in NanoVectorDB[6]. The extracted entities and their relationships are leveraged to construct the knowledge graph using Neo4j[7] storage. The legal queries were passed individually to both the standard and ontology guided framework through the *get_answer()* function. For each query, LightRAG first retrieves relevant document chunks from the vector index together with related entities and relationships from the knowledge graph. GPT-4o-mini[8] (gpt-4o-mini-complete) is employed both during the indexing stage for entity and relation extraction and during query processing for answer generation. At query time, the program operates in mix retrieval mode, combining semantic retrieval from the vector database with graph-based exploration before generating the final response.

4.2. Ontology Guided Configuration

The ontology-guided configuration incorporates the ontology during the indexing stage. Before initializing LightRAG, the AI Risk Ontology (AIRO) is processed using a module, which extracts the ontology-defined entity types. These entity types are supplied to LightRAG enabling the LLM to use domain-specific categories during entity extraction. Consequently, the generated knowledge graph is organized according to AIRO concepts rather than relying solely on the default entity categories. The remaining components remain unchanged: GPT-4o-mini is used for entity extraction and answer generation, OpenAI embeddings are used for semantic indexing, Neo4j stores the ontology-guided knowledge graph, and NanoVectorDB stores the vector embeddings. During question answering, LightRAG performs hybrid retrieval, combining semantic similarity retrieval with graph exploration over

[2]https://github.com/HKUDS/LightRAG
[3]https://artificialintelligenceact.eu/ai-act-explorer/
[4]https://github.com/PurbashaChowdhur/LightRAG
[5]https://openai.com/index/new-embedding-models-and-api-updates/
[6]https://pypi.org/project/nano-vectordb/
[7]https://neo4j.com/
[8]https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

the ontology-guided knowledge graph before passing the retrieved context to GPT-4o-mini to generate the final response.

5. Analysis of Results

5.1. First Evaluation

Due to the unavailability of datasets for evaluating QA systems on the AI Act, a small dataset was created consisting of 10 questions. Each question is paired with an expected answer and related legal provisions reviewed by a panel of lawyers with expertise in the AI Act. While the number of questions is limited, this dataset has been used for both the standard and ontology-guided LightRAG experiments to evaluate the performance of the corresponding approaches.

1.1 **Question:** What risk categories are provided for AI systems?
Answer: The AI Act establishes a risk-based framework, categorising AI systems based on their potential for harm. Here are the risk categories: Unacceptable Risk, High Risk, Limited Risk and Minimal Risk.
References: Articles: 3, 5, 6, 50. Recitals: 26, 50, 52, 54, 55, 56, 57, 58, 59, 61, 64, 66, 132, 165.

1.2 **Question:** What types of AI systems or practices are explicitly prohibited by the AI Act?
Answer: Subliminal techniques beyond a person's consciousness to materially distort behaviour in a way that causes harm. Exploiting vulnerabilities of specific groups (e.g., age, disability) to materially distort behaviour and cause harm. Social scoring systems can lead to detrimental treatment in unrelated social contexts or disproportionate treatment. Real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes, except for specific exceptions. AI systems used by or on behalf of law enforcement to make risk assessments for predicting criminal offenses based solely on profiling or personality traits.
References: Articles: 5.

1.3 **Question:** What is the definition of high risk?
Answer: The AI Act defines 'high risk' in the context of AI systems that pose significant risks to the health and safety or the fundamental rights of persons. This determination considers both the severity of the possible harm and the probability of its occurrence. The Act specifically outlines two main categories of high-risk AI systems: AI Systems Integrated into Products and Stand-alone High-Risk AI Systems. See Annex III.
References: Articles: 6.

1.4 **Question:** Which sectors or applications are considered high-risk under the regulation?
Answer: Irrespective of whether an AI system is placed on the market or put into service independently of the products referred to in points (a) and (b), that AI system shall be considered to be high-risk where both of the following conditions are fulfilled: (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I; (b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I. 2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall be considered to be high-risk.
References: Articles: 6.

1.5 **Question:** How should users be informed when interacting with AI systems?
Answer: When interacting with an AI system, natural persons should be informed that they are interacting with an AI system unless it is obvious to a reasonably well-informed person. This obligation does not apply to AI systems used for law enforcement purposes, unless those systems are available to the public for reporting criminal offences. Deployers of emotion recognition or biometric categorisation systems should inform natural persons exposed to the system of its

operation. This obligation does not apply to AI systems used for biometric categorisation and emotion recognition that are permitted to detect, prevent or investigate criminal offences, subject to safeguards for the rights and freedoms of third parties, and in accordance with Union law. Deployers of an AI system that generates or manipulates text, which is published to inform the public on matters of public interest, shall disclose that the text has been artificially generated or manipulated. Information should be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. This information should conform to the applicable accessibility requirements.
References: Articles: 50.

1.6 **Question:** What are AI systems with minimal risks?
Answer: While the regulation is centred on high-risk AI systems and general-purpose AI models with systemic risks, it specifies that AI systems may be developed for purposes that do not pose significant risks to health, safety, or fundamental rights. These include: An AI system that transforms unstructured data into structured data; An AI system that classifies incoming documents into categories; An AI system used to detect duplicates among a large number of applications; An AI system that improves the language used in documents without changing their substance; An AI system that detects decision-making patterns or deviations from prior patterns; An AI system that performs tasks preparatory to an assessment.
References: Recitals: 53.

1.7 **Question:** What are the guiding principles established by the AI Act?
Answer: Protection of fundamental rights, including democracy, the rule of law, environmental protection, health, and safety.
References: Recitals: 1, 7, 48.

1.8 **Question:** What is the definition of a "system of artificial intelligence"?
Answer: A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments
References: Articles: 3.

1.9 **Question:** What are the penalties for violating the provisions of the regulation for high-risk AI systems?
Answer: Non-compliance with Obligations for High-Risk Systems and transparency obligations for providers and deployers: Breaches of other provisions concerning high-risk AI systems can result in administrative fines of up to 15,000,000 EUR or 3% of the offender's total worldwide annual turnover.
References: Articles: 99.

1.10 **Question:** What is the difference between the deployer and the provider?
Answer: The Provider is a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge. The Deployer is a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity". The provider bears the primary responsibility for ensuring the AI system complies with the requirements of the AI Act, especially if it is classified as a high-risk system. The deployer is the entity that uses the AI system for its intended purpose.
References: Articles: 3, 16.

Table 1
Comparison of expert-reviewed legal references and the references retrieved by Standard and Ontology-guided LightRAG across ten legal queries, including the correctly retrieved legal references with expert verified references.

Questions	Expert-reviewed references	Standard LightRAG	Ontology-guided LightRAG	Correctly retrieved references
Q1.1	**Article(s)**: 3,5,6,50 **Recital(s)**: 26,50,52,54,55,56, 57,58,59,61,64,66,132,165	**Article(s)**: 7,40,15 **Recital(s)**: 110,64	**Article(s)**: 7,14,36 **Recital(s)**: 62,110	**Standard LightRAG**: Rec.64
Q1.2	**Article(s)**: 5	**Article(s)**: 5 **Recital(s)**: 29,31,44,58	**Article(s)**: 1,8 **Recital(s)**: 57,97	**Standard LightRAG**: Art.5
Q1.3	**Article(s)**: 6	**Article(s)**: 3 **Recital(s)**: 58,126,5,66	**Article(s)**: 80,3,82,17 **Recital(s)**: 50	–
Q1.4	**Article(s)**: 6	**Article(s)**: 1,4,58,17,19	**Article(s)**: 74 **Recital(s)**: 4,58,59,130	–
Q1.5	**Article(s)**: 50	**Article(s)**: 50 **Recital(s)**: 132,31,44,137	**Article(s)**: 50 **Recital(s)**: 132,91,72,66	**Standard LightRAG**: Art.50 **Ontology-guided LightRAG**: Art.50
Q1.6	**Recital(s)**: 53	**Article(s)**: 6 **Recital(s)**: 53,54,6,61	**Article(s)**: 6 **Recital(s)**: 53,25,7,165	**Standard LightRAG**: Rec.53 **Ontology-guided LightRAG**: Rec.53
Q1.7	**Recital(s)**: 1,7,48	**Article(s)**: 1,40 **Recital(s)**: 27,6,53	**Article(s)**: 1,2 **Recital(s)**: 27,130,165	–
Q1.8	**Article(s)**: 3	**Article(s)**: 3,7,68 **Recital(s)**: 12,97	**Article(s)**: 80,17,112 **Recital(s)**: 82,51	–
Q1.9	**Article(s)**: 99	**Article(s)**: 99,101,100 **Recital(s)**: 168	**Article(s)**: 99,100,112	**Standard LightRAG**: Art.99 **Ontology-guided LightRAG**: Art.99
Q1.10	**Article(s)**: 3,16	**Article(s)**: 20,26,43 **Recital(s)**: 65,13	**Article(s)**: 60,27,80,26,94	–

Table 1 compares the legal references identified by legal experts with those retrieved by the Standard LightRAG and Ontology-guided LightRAG configurations for the ten evaluation queries. The analysis considers only the retrieved legal provisions (articles and recitals) allowing the retrieval capability of both configurations to be assessed independently of response generation. The final column reports the legal references that exactly match the expert-reviewed references.

The Standard LightRAG configuration retrieves correct references for Q1.1, Q1.2, Q1.5, Q1.6, and Q1.9, whereas the Ontology-guided LightRAG configuration retrieves correct references for Q1.5, Q1.6, and Q1.9. No correct reference is retrieved by either configuration for Q1.3, Q1.4, Q1.7, Q1.8, and Q1.10. These results indicate that the standard configuration retrieves a greater number of expert-reviewed legal provisions for this evaluation dataset. Although fewer exact matches are obtained with the ontology-guided configuration, both retrieval strategies return several common legal references. In

particular, the two configurations retrieve common references for Q1.1 (Article 7 and Recital 110), Q1.5 (Article 50 and Recital 132), Q1.6 (Article 6 and Recital 53), Q1.7 (Article 1 and Recital 27), Q1.9 (Articles 99 and 100), and Q1.10 (Article 26). This overlap suggests that both configurations frequently identify similar regions of the legal corpus, while the ontology-guided configuration explores additional semantically related legal concepts during graph traversal, it does not consistently increase the number of expert-reviewed legal references retrieved in this experiment. Overall, the comparison shows that both configurations are capable of retrieving relevant legal provisions for several compliance-related queries. While the Standard LightRAG configuration achieves a higher correspondence with the expert-reviewed references, the ontology-guided configuration demonstrates that incorporating domain knowledge changes the retrieval behaviour by directing the search towards semantically related legal entities and provisions.

5.2. Second Evaluation

Aggio et al. developed two customized implementations [44] of GraphReader and applied to question answering on the AI Act. Unlike the original approach by [45], the GraphReader (base) model selects extracted statements from a specific chunk of the legal directives based on the associated parts-of-speech components such as verbs, nouns, adjectives extracted from the input question. In the other side, they have implemented another variant of GraphReader that removes the decomposed parts-of-speech selection phase and instead directly selects the corresponding chunks based on similarity to the legal queries. The authors used the same set of legal queries for their experimental analysis.
In our work, we considered the Graphreader (base) produced legal references to compare with those generated by both the LightRAG approaches. While both automatic and human-based evaluation methods were considered in their work, we relied on the qualitative analysis based on the similar legal materials of the discussed methods. The future work aims to conduct a quantitative analysis based on standard and scalable measurements. However, these methods might not be effective for complex responses, so human evaluation will also be included to provide a more subtle assessment of response quality.
The second evaluation is conducted using 10 legal queries without predefined answers. The legal references retrieved by the two LightRAG configurations exhibit more diverse outcomes than the legal provisions observed for the expert-reviewed evaluation dataset.

2.1 **Question:** What criteria exist for assessing the risk of an AI system?
2.2 **Question:** What are the sanctions for violating the provisions of the regulation for transparency-risk systems?
2.3 **Question:** What obligations exist for deployers of high-risk AI systems?
2.4 **Question:** What requirements must AI systems classified as high-risk meet?
2.5 **Question:** What obligations do providers of high-risk AI systems have in terms of transparency and technical documentation?
2.6 **Question:** What does a conformity assessment consist of?
2.7 **Question:** What does systemic-risk mean?
2.8 **Question:** What is the definition of General-purpose AI?
2.9 **Question:** What are the components of a quality management system?
2.10 **Question:** What are the requirements for documenting bias mitigation measures in AI models?

Table 2 compares the legal references retrieved by an existing approach GraphReader (base) with both standard LightRAG and ontology-guided LightRAG for the second evaluation dataset, which consists of ten legal queries without predefined answers.
The comparison primarily focuses on the consistency and diversity of the retrieved legal provisions across the three retrieval approaches.
The results show that all three methods retrieve different combinations of articles and recitals, reflecting the influence of their respective retrieval strategies. There are common legal provisions

Table 2
Legal references retrieved by GraphReader, Standard LightRAG, and Ontology-guided LightRAG for second dataset legal queries including common references in all three frameworks.

Questions	GraphReader (base)[44]	Standard LightRAG	Ontology-guided LightRAG	Common references in all
Q2.1	**Article(s)**: 7,79,72 **Recital(s)**: 66,64,65	**Article(s)**: 9,7 **Recital(s)**: 74,66,40	**Article(s)**: 7,10,16 **Recital(s)**: 66,53	**Article(s)**: 7 **Recital(s)**: 66
Q2.2	**Article(s)**: 99,112 **Recital(s)**: 46,52,49,51	**Article(s)**: 99,100,85 **Recital(s)**: 179,168	**Article(s)**: 99,74,112 **Recital(s)**: 168,45	**Article(s)**: 99
Q2.3	**Article(s)**: 26,27,49 **Recital(s)**: 65,92	**Article(s)**: 26 **Recital(s)**: 91,93,96,178	**Article(s)**: 1,26,9 **Recital(s)**: 66,177	**Article(s)**: 26
Q2.4	**Recitals**: 66	**Article(s)**: 7,16,9 **Recital(s)**: 66,159	**Article(s)**: 1,9,25 **Recitals**: 71,171	–
Q2.5	**Article(s)**: 11,21,18,13,99 **Recitals**: 65	**Article(s)**: 19,16,73 **Recital(s)**: 72,66	**Article(s)**: 16,74,9 **Recital(s)**: 66,171	–
Q2.6	**Article(s)**: 43,16,45,46 **Recital(s)**: 53	**Article(s)**: 31,44,16 **Recital(s)**: 125,173	**Article(s)**: 25,39,38 **Recital(s)**: 125,173	–
Q2.7	**Article(s)**: 3,56	**Article(s)**: 55,53 **Recital(s)**: 97,110,139	**Article(s)**: 3,55 **Recital(s)**: 110,136,120	–
Q2.8	**Article(s)**: 3 **Recital(s)**: 12,6,110	**Article(s)**: 3 **Recital(s)**: 97,99,100,112	**Article(s)**: 3,94,54 **Recital(s)**: 110,111	**Article(s)**: 3
Q2.9	**Article(s)**: 17 **Recital(s)**: 27,66	**Article(s)**: 17,16,31 **Recital(s)**: 81,139	**Article(s)**: 16,17,63,8 **Recital(s)**: 81	**Article(s)**: 17
Q2.10	**Article(s)**: 26,9,21 **Recital(s)**: 65,66,61,27	**Article(s)**: 10,53 **Recital(s)**: 70,25,114	**Article(s)**: 10,9 **Recital(s)**: 70,65,67	–

retrieved by all the three approaches. For example, **Article 7**[9] and **Recital 66**[10] (related articles 10,11,12,13,14,15) for Q2.1 which are closely related to the classification of high-risk AI systems and the associated risk assessment criteria, although the ontology-guided configuration further retrieves **Article 10**[11] concerning data and data governance, indicating that ontology-guided retrieval expands the search towards semantically related compliance requirements beyond the primary risk classification provisions, **Article 99**[12] for Q2.2 which specifies administrative fines and penalties under the AI Act, **Article 26**[13] related to the obligations of deployers of high risk AI systems for Q2.3.

For Q2.5, concerning transparency and technical documentation obligations of providers, no common provision is retrieved by all three approaches. GraphReader identifies Articles 11, 21, 18, 13, and 99, which relate to technical documentation, obligations, transparency, and enforcement. Standard LightRAG retrieves Articles 19, 16, and 73, whereas the ontology-guided configuration returns Articles

[9] https://artificialintelligenceact.eu/article/7/
[10] https://artificialintelligenceact.eu/recital/66/
[11] https://artificialintelligenceact.eu/article/10/
[12] https://artificialintelligenceact.eu/article/99/
[13] https://artificialintelligenceact.eu/article/26/

16, 74, and 9 which mostly address the **obligations of the providers of high-risk AI systems**[14]. For Q2.6, which focuses on conformity assessment procedures, GraphReader retrieves Articles 43, 16, 45, and 46, together with **Article 53**[15], while the Standard and Ontology-guided LightRAG configurations retrieve different Articles but both include Recitals 125 and 173. The common retrieval of these Recitals suggests that both LightRAG configurations identify similar explanatory provisions, whereas GraphReader outperforms the procedural Articles governing conformity assessment.
For Q2.7, addressing the definition of systemic risk, all three approaches retrieve different Articles, although both LightRAG configurations identify **Recital 110**[16]. This overlap indicates that the two LightRAG variants rely on similar contextual information despite selecting different supporting Articles.
For Q2.8, which asks for the definition of general-purpose AI, all three frameworks consistently retrieve **Article 3**[17], where the relevant definitions are provided.
For Q2.9, concerning the components of a quality management system, all three methods retrieve **Article 17**[18], which defines the quality management system for high-risk AI systems. Finally, Q2.10, which concerns documenting bias mitigation measures, produces the greatest divergence among the three approaches. None of the retrieved Articles or Recitals are shared across all frameworks. GraphReader retrieves Articles 26 and 21, Standard LightRAG retrieves Articles 10 and 53, while the ontology-guided configuration retrieves Articles 10 and 9. Therefore, unlike GraphReader, both the LightRAG retrieved the related reference since Article 10 addresses data and data governance also for the bias mitigation (point 2(f) & 2(g) of Article 10). These differences suggest that the concept of bias mitigation is addressed through several interconnected regulatory requirements rather than a single explicit provision, resulting in substantially different retrieval paths. Compared with both LightRAG configurations, GraphReader often retrieves a different set of supporting provisions, demonstrating an alternative graph traversal strategy. Overall, the comparison demonstrates that ontology guidance changes the set of legal references explored by the retrieval process rather than simply reproducing the retrieval way of the standard LightRAG configuration.

6. Conclusion and Future Work

In this paper, we investigated the application of a graph-based Retrieval-Augmented Generation framework for question answering over the European Union AI Act. Two retrieval configurations were evaluated using the same set of expert-reviewed legal queries: the standard LightRAG framework and an ontology-guided variant that incorporates AIRO-derived entity types during knowledge graph construction. The experiments focused on comparing the legal provisions retrieved by both configurations against expert-reviewed and another RAG framework generated legal references. The results demonstrate that LightRAG is capable of retrieving relevant articles and recitals for the legal queries, although the retrieval accuracy varies depending on the characteristics of the question. The study highlights the importance of evaluating the retrieval stage independently from the generated answers, particularly in legal question answering where the correctness of supporting legal references is essential for transparency and traceability. By providing a comparison between standard and ontology-guided graph retrieval over the EU AI Act, this work offers insights into the practical use of graph-based RAG systems for legal information retrieval. As future work, we aim to employ different large language models functionalities to create richer ontology expanding the existing ones with more RDF triples, classes, relations to evaluate larger collections of legal questions covering different regulatory domains, build a baseline dataset to conduct the quantitative analysis between different other retrieval methods. Besides, the future work aims to involve human validation to generate questionnaires on the contextual foundation and validate the results from the models.

[14] https://artificialintelligenceact.eu/article/16/
[15] https://artificialintelligenceact.eu/article/53/
[16] https://artificialintelligenceact.eu/recital/110/
[17] https://artificialintelligenceact.eu/article/3/
[18] https://artificialintelligenceact.eu/article/17/

Acknowledgments

This work was supported by a PNRR-funded PhD fellowship under the National Recovery and Resilience Plan (PNRR).

Declaration on Generative AI

The authors did not use generative AI tools in the research or preparation of this manuscript, except for language polishing.

References

[1] T. Scantamburlo, P. Falcarin, A. Veneri, A. Fabris, C. Gallese, V. Billa, F. Rotolo, F. Marcuzzi, Software systems compliance with the AI act: Lessons learned from an international challenge, in: Proceedings of the 2nd International Workshop on Responsible AI Engineering, RAIE '24, ACM, 2024, p. 44–51. URL: https://doi.org/10.1145/3643691.3648589. doi:10.1145/3643691.3648589.

[2] J. Burrell, How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society 3 (2016) 2053951715622512. doi:10.1177/2053951715622512.

[3] O. Amaral, M. Azeem, S. Abualhaija, L. Briand, Nlp-based automated compliance checking of data processing agreements against gdpr, IEEE Transactions on Software Engineering PP (2023) 1–23. doi:10.1109/TSE.2023.3288901.

[4] T. Scantamburlo, A. Charlesworth, N. Cristianini, Machine decisions and human consequences, CoRR abs/1811.06747 (2018). URL: http://arxiv.org/abs/1811.06747. arXiv:1811.06747.

[5] European union: Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

[6] O. A. Cejas, M. I. Azeem, S. Abualhaija, L. C. Briand, Nlp-based automated compliance checking of data processing agreements against gdpr, IEEE Transactions on Software Engineering 49 (2023) 4282–4303. doi:10.1109/TSE.2023.3288901.

[7] N. Malali, The role of retrieval-augmented generation (rag) in financial document processing: Automating compliance and reporting, Int. J. Manag. Technol 12 (2025) 26–46.

[8] M. Fernández-Barrera, G. Sartor, The legal theory perspective: doctrinal conceptual systems vs. computational ontologies, in: Approaches to Legal Ontologies: Theories, Domains, Methodologies, Springer, 2010, pp. 15–47.

[9] M. Avgerinos Loutsaris, Z. Lachana, C. Alexopoulos, Y. Charalabidis, Legal text processing: Combing two legal ontological approaches through text mining, in: Proceedings of the 22nd Annual International Conference on Digital Government Research, 2021, pp. 522–532.

[10] N. Casellas, Legal ontology engineering: Methodologies, modelling trends, and the ontology of professional judicial knowledge, volume 3, Springer Science & Business Media, 2011.

[11] P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, V. Presutti, Ontology engineering with ontology design patterns: foundations and applications, volume 25, IOS Press, 2016.

[12] V. Presutti, G. Lodi, A. Nuzzolese, A. Gangemi, S. Peroni, L. Asprino, The role of ontology design patterns in linked data projects, in: International Conference on Conceptual Modeling, Springer, 2016, pp. 113–121.

[13] F. Sovrano, M. Palmirani, F. Vitali, et al., Legal knowledge extraction for knowledge graph based question-answering, Frontiers in Artificial Intelligence and Applications 334 (2020) 143–153.

[14] A. Ravichander, A. W. Black, S. Wilson, T. Norton, N. Sadeh, Question answering for privacy policies: Combining computational and legal perspectives, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 4947–4958.

[15] F. Sovrano, M. Palmirani, S. Sapienza, V. Pistone, Discolqa: zero-shot discourse-based legal question answering on european legislation, Artificial Intelligence and Law (2024) 1–37.
[16] A. Louis, G. van Dijck, G. Spanakis, Interpretable long-form legal question answering with retrieval-augmented large language models, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 2024, pp. 22266–22275.
[17] European Parliament, Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act), 2024. URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.
[18] Z. Guo, L. Xia, Y. Yu, T. Ao, C. Huang, Lightrag: Simple and fast retrieval-augmented generation, arXiv preprint arXiv:2410.05779 (2024).
[19] N. Aggio, A graph-enhanced llm-based question answering system for the ai act (2023).
[20] H. Jelodar, S. Bai, M. Meymani, P. Hamedi, R. Razavi-Far, A. Ghorbani, Integrating graphs, large language models, and agents: Reasoning and retrieval, arXiv preprint arXiv:2604.15951 (2026).
[21] I. Oranekwu, L. Elluri, G. Batra, Automated knowledge framework for iot cybersecurity compliance, in: 2024 IEEE International Conference on Big Data (BigData), IEEE, 2024, pp. 6336–6345.
[22] S. S. Chennu, L. Elluri, G. Batra, Bridging ai and legal compliance: Knowledge graphs for iot cybersecurity regulations (2025).
[23] T. Van Engers, A. Boer, J. Breuker, A. Valente, R. Winkels, Ontologies in the legal domain, in: Digital government: E-government research, case studies, and implementation, Springer, 2008, pp. 233–261.
[24] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, arXiv preprint arXiv:2203.05794 (2022). URL: https://maartengr.github.io/BERTopic/index.html.
[25] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos, Legal-bert: The muppets straight out of law school, arXiv preprint arXiv:2010.02559 (2020).
[26] L. Elluri, S. S. L. Chukkapalli, K. P. Joshi, T. Finin, A. Joshi, A bert based approach to measure web services policies compliance with gdpr, IEEE Access 9 (2021) 148004–148016.
[27] O. Ram, Y. Levine, I. Dalmedigos, D. Muhlgay, A. Shashua, K. Leyton-Brown, Y. Shoham, In-context retrieval-augmented language models, Transactions of the Association for Computational Linguistics 11 (2023) 1316–1331.
[28] W. Fan, Y. Ding, L. Ning, S. Wang, H. Li, D. Yin, T.-S. Chua, Q. Li, A survey on rag meeting llms: Towards retrieval-augmented large language models, in: Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, 2024, pp. 6491–6501.
[29] L. Gao, X. Ma, J. Lin, J. Callan, Precise zero-shot dense retrieval without relevance labels, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 1762–1777.
[30] C.-M. Chan, C. Xu, R. Yuan, H. Luo, W. Xue, Y. Guo, J. Fu, Rq-rag: Learning to refine queries for retrieval augmented generation, arXiv preprint arXiv:2404.00610 (2024).
[31] Y. Yu, W. Ping, Z. Liu, B. Wang, J. You, C. Zhang, M. Shoeybi, B. Catanzaro, Rankrag: Unifying context ranking with retrieval-augmented generation in llms, Advances in Neural Information Processing Systems 37 (2024) 121156–121184.
[32] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitansky, R. O. Ness, J. Larson, From local to global: A graph rag approach to query-focused summarization, arXiv preprint arXiv:2404.16130 (2024).
[33] J. Martinez-Gil, A survey on legal question–answering systems, Computer Science Review 48 (2023) 100552.
[34] A. Abdallah, B. Piryani, A. Jatowt, Exploring the state of the art in legal qa systems, Journal of Big Data 10 (2023) 127.
[35] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman, Indexing by latent semantic analysis, Journal of the American society for information science 41 (1990) 391–407.
[36] G. Lame, Using nlp techniques to identify legal ontology components: Concepts and relations,

Artificial Intelligence and Law 12 (2004) 379–396.
[37] K. Xu, S. Reddy, Y. Feng, S. Huang, D. Zhao, Question answering on freebase via relation extraction and textual evidence, arXiv preprint arXiv:1603.00957 (2016).
[38] B. Fawei, J. Z. Pan, M. Kollingbaum, A. Z. Wyner, A methodology for a criminal law and procedure ontology for legal question answering, in: Semantic Technology: 8th Joint International Conference, JIST 2018, Awaji, Japan, November 26–28, 2018, Proceedings 8, Springer, 2018, pp. 198–214.
[39] M.-Y. Kim, Y. Xu, R. Goebel, Applying a convolutional neural network to legal question answering, in: JSAI International Symposium on Artificial Intelligence, Springer, 2015, pp. 282–294.
[40] N.-N. Zhang, Y. Xing, Questions and answers on legal texts based on bert-bigru, in: Journal of Physics: Conference Series, volume 1828, IOP Publishing, 2021, p. 012035.
[41] D. Golpayegani, H. J. Pandit, D. Lewis, Airo: An ontology for representing ai risks based on the proposed eu ai act and iso risk management standards, in: Towards a Knowledge-Aware AI: SEMANTiCS 2022—Proceedings of the 18th International Conference on Semantic Systems, 13–15 September 2022, Vienna, Austria, SAGE Publications Pvt. Ltd 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP, 2022, pp. 51–65.
[42] D. Golpayegani, H. J. Pandit, D. Lewis, To be high-risk, or not to be—semantic specifications and implications of the ai act's high-risk ai applications and harmonised standards, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, pp. 905–915.
[43] E. Carbone, A. G. Berton, P. Chowdhury, T. Scantamburlo, P. Falcarin, Topic modelling on the european ai act, in: TRUST-AI: The European Workshop on Trustworthy AI. Co-located with the European Conference of Artificial Intelligence ECAI, 2025, pp. 200–208. URL: https://ceur-ws.org/Vol-4132/.
[44] N. Aggio, A. de Lazzari, T. Scantamburlo, A graph-enhanced llm-based question answering system for the ai act, 2025.
[45] S. Li, Y. He, H. Guo, X. Bu, G. Bai, J. Liu, J. Liu, X. Qu, Y. Li, W. Ouyang, et al., Graphreader: Building graph-based agent to enhance long-context abilities of large language models, arXiv preprint arXiv:2406.14550 (2024).