Scientific Reports

https://doi.org/10.1038/s41598-026-43098-3

Article in Press

Semantic clause retrieval for trademark law using transformer encoders and lexical baselines: a cross-domain agri-robotics compliance case study

Received: 19 September 2025

Accepted: 2 March 2026

Published online: 05 March 2026

Cite this article as: Asfand E Yar M., Hashir Q., Tanveer M.H. *et al.* Semantic clause retrieval for trademark law using transformer encoders and lexical baselines: a cross-domain agri-robotics compliance case study. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-43098-3

Muhammad Asfand E Yar, Qadeer Hashir, M. Hassan Tanveer, Razvan C. Voicu & Adeel Khalid

Semantic Clause Retrieval for Trademark Law Using Transformer Encoders and Lexical Baselines: A Cross-Domain Agri-Robotics Compliance Case Study

Muhammad Asfand-e-Yar*[1], Qadeer Hashir[1], M. Hassan Tanveer*[2], Razvan C. Voicu[2], and Adeel Khalid[3]

[1]Center of Excellence in Artificial Intelligence (CoE-AI), Department of Computer Science, Bahria University, Islamabad, Pakistan
[2]Department of Robotics and Mechatronics Engineering, Kennesaw State University, USA
[3]Department of Industrial and Systems Engineering, Kennesaw State University, USA
*corresponding author: asfandeyar.buic@bahria.edu.pk; mtanveer@kennesaw.edu

ABSTRACT

Clause-level retrieval is a recurring bottleneck in legal research and compliance workflows: relevant obligations, exceptions, procedures, and enforcement conditions are often buried in long statutes and regulatory texts, and users may not know the exact terminology needed for keyword search. We present an application-oriented semantic clause retrieval pipeline that indexes documents at the clause level and ranks candidates using off-the-shelf sentence-transformer encoders with cosine similarity. Standard lexical baselines are included to contextualize performance under the same top-k retrieval and expert relevance judgment protocol. We evaluate the approach in a cross-domain setting spanning, trademark statute retrieval on Trademark Ordinance data and a scoped agri-robotics compliance corpus covering regulatory and standards-oriented requirements. The trademark benchmark serves as the primary quantitative evaluation, while the agri-robotics component is used to assess cross-domain transfer under a bounded query set without overstating generalization. In addition to aggregate ranking metrics, we report query-level analysis to characterize model behavior and common failure modes, including high-similarity but decision-irrelevant matches that arise from procedural or definitional overlap.

Keywords: Semantic search, Information retrieval, Sentence embeddings, Legal NLP, Trademark law, Agri-robotics compliance

Introduction

Legal research is a fundamental part of the legal profession and serves as the basis for informed decision-making and jurisprudential development. It is essential to legal practice because it enables practitioners to locate key legal materials such as statutes, clauses, sections, and prior rulings[1]. As legal landscapes differ across jurisdictions, the ability to understand and access legal information has become important not only for local practitioners but also for the global legal community[2]. In this context, legal systems comprise a wide range of document types, from laws and regulations to constitutions and judgments, each shaped by country-specific structures and traditions[2].

Legal frameworks governing organizations and businesses are similarly diverse and vary widely across countries, reflecting national, cultural, and economic factors. These systems include rules, legislation, and corporate laws that are tailored to local conditions[3]. For example, corporate governance structures, including the roles of directors and shareholders, differ across jurisdictions, influencing how companies operate and defining the rights and obligations of involved parties[4]. In addition, company registration, taxation, and intellectual property regulations vary substantially, creating distinct challenges and opportunities for companies and entrepreneurs worldwide[5].

A central concern of this work is supporting international awareness in legal research through an approach that is inherently transferable across jurisdictions. Historically, legal research relied heavily on manual searching and the handling of bulky information sources[6]. With advances in computing, legal research has undergone significant change. While legal systems have grown more complex, technology has also enabled more sophisticated legal search systems, making research more efficient and accessible.

In this study, we describe the evolution of legal searching from early computer-based systems[7] to modern Natural Language Processing (NLP) and transformer-based methods that improve search accuracy and efficiency. Online legal information portals and digital legal databases have substantially improved access to statutes, case law, and secondary materials by enabling searchable archives and faster retrieval compared with paper-based workflows. Traditional legal search systems were primarily keyword-driven and required users to enter specific terms or phrases to retrieve relevant documents. These systems represented an important stage in legal research practice[8]. More recently, semantic text categorization and ontology modeling have also been adopted in legal settings, including emerging compliance areas such as agriculture and robotics. In precision agriculture, automated systems increasingly rely on semantic analysis to interpret environmental and plant health data. At the same time, legal and compliance requirements for autonomous agricultural robots are complex due to heterogeneous sources, evolving policies, and jurisdiction-specific constraints. As a result, information retrieval mechanisms that can bridge established legal codes and dynamic regulatory requirements are becoming increasingly important for robotics-enabled farming[9–11].

Alongside the growing use of AI and rapid advances in NLP, legal information access has also been reshaped. NLP is a subfield of artificial intelligence that develops computational methods for analyzing, understanding, and generating human language. One key development is semantic search, which differs from conventional keyword-based methods by considering contextual meaning and related concepts rather than relying only on exact term matches. This approach is better aligned with how legal professionals formulate questions in practice. Word embeddings and transformer models such as Bidirectional Encoder Representations from Transformers (BERT) are central to this paradigm because they can capture nuanced semantic information in text[12].

This study proposes a semantic search approach for trademark-focused legal research to enhance efficiency, comprehensiveness, and usability. At the same time, the underlying retrieval formulation is intended to be applicable to legal research in other jurisdictions and to compliance texts beyond trademark law.

The main contributions of this paper are as follows:

- We present an application-oriented clause retrieval pipeline that uses off-the-shelf legal sentence encoders (no fine-tuning) and cosine similarity ranking to retrieve relevant statutory clauses for practitioner-style queries.
- We strengthen evaluation by expanding the query set to cover both trademark law (Q1–Q10) and agri-robotics compliance (Q11–Q13), and by adding standard lexical baselines (TF–IDF, BM25) under the same top-k protocol ($k = 5$).
- We provide quantitative benchmarking on the Pakistan Trademark Ordinance (2001), reporting AP per query and MAP across Q1–Q10 for TF–IDF, BM25, and three transformer encoders, and we report cross-domain baseline MAP on Q11–Q13.
- We add analytical insights through query-level error analysis (including false-positive patterns where procedural/legal phrasing yields high similarity but fails decision relevance), and we explicitly separate current validated results from future work on broader cross-domain transformer benchmarking and reliability reporting.

The remainder of this paper is organized as follows. Section 2. reviews related work and summarizes key developments in legal search and semantic retrieval. Section 3. describes the methodology used to build the search system, including how queries and statutory segments are represented. Section 4. details the proposed retrieval pipeline, covering preprocessing, embedding generation, ranking, and evaluation. Section 5. also reports evaluation results and performance across methods. Finally, Section 6. summarizes the main findings, discusses limitations, and outlines directions for future work.

Related Work

In the 1970s, two names became more popular in legal research and they were Lexis Legal Search and Westlaw[7]. Though such machines initially did not allow for complex questions to be answered, they were an incredible feat of technology given the fact that lawyers and legal professionals could then search for all relevant cases and legal documents using keywords and other search parameters rather than having to printed (hard-copy) examine volumes of legal materials. Such systems initiated a revolution in the legal domain because they created an environment that featured dependent and accessible legal research. Prior to the development of the later systems, legal research was both a time-consuming and an expensive process, and it required lawyers to manually browse through the books and other printed (hard-copy) legal materials literature[13].

Citation Analysis in Legal Research

The evaluation of the references of cases being examined is a very important technique in the assessment of the connections between law cases through the observation of the citations made to the previous cases in the given document. This method is based on the assumption that one may assess the opportunity of the legal issue of one case to be relied on by others. The application of citation analysis as a legal research method enables the identification of pivotal legal cases reported in subsequent cases. This technique is particularly valuable for finding situations that caught little attention in their initial merely precedent-setting but at a later stage got a sufficiently wide acceptance because they continuously influenced other ensuing legal cases ([14];[15]). The other study looks into the Canadian legal cases network using the methods of statistics and networks.

Thom Neale looks at the discrepancy in coverage between CanLII viz two private vendors and then compares the figures showing the number of cases from various provinces that have been cited in different years over time. The study reveals that indegree centrality and PageRank scores of cases adjudged on CanLII's webpage tend to be factors that contribute to the popularity of those cases[16]. While early research in semantic search and ontology modeling was predominantly situated in legal and judicial contexts facilitating the retrieval and analysis of complex statutory and case law texts[17] there has been a notable expansion into agricultural and robotic domains in recent years. Kamilaris et al.[9] provide a comprehensive review of big data and semantic approaches in agriculture, highlighting how automated systems leverage ontology and semantic categorization for interpreting environmental and plant health data. In parallel, Bechar and Vigneault[10] examine the integration of robotics in field operations, underscoring the emerging regulatory and data management challenges associated with autonomous agricultural machinery. The ethical and legal implications of these technologies have also been systematically addressed; van Wynsberghe and Donhauser[11] detail the evolving frameworks needed to ensure responsible deployment of autonomous robots in agriculture. Policy and standards organizations, such as the FAO[18] and IEEE[19], have responded by publishing global guidelines to support safe, ethical, and effective adoption of digital and robotic systems in farming. Collectively, these works underscore the growing importance of advanced information retrieval and semantic analysis systems capable of bridging established legal structures with the dynamic regulatory landscape in agri-robotics.

Advances in Semantic Legal Search

The technological advances in software, hardware, and true NLP, apart from making more unique legal research methods, have been doing so. This process may be performed by utilizing data mining and machine learning in which strengths and weaknesses of search components play an essential role to this end. Such technique, when considering vocabulary, context, and the premise taken, is a guarantee of good decision-making. This helps to make the utterance of these questions well understood through the systems. Having this powerful computing engine opens the gates to a wide range of possibilities to group features like numbers of term frequencies, text length, and one-time words to obtain rich vector representations for juridical texts. The use of Q-grams similarity space is a key element in the modeling of consistencies[20]. To further elaborate, the system takes on an investigation that utilizes NLP techniques to collect cases that entail law and find past cases that can be supportive to law practitioners and those who are doing research on law. Additionally, another research proved that the traditional project undertaking always takes a long time until this study came out, which suggests an online platform that automatically imports the items in a person's wish list into their website based on their behavior[15]. Exploring LLMs Applications in Law[21] provides a systematic review of 61 studies on Transformer-based LLMs for tasks such as legal document analysis, case prediction, contract review, and legal research, highlighting emerging trends, performance, and methodological gaps in current legal NLP. Another recent survey[22] uses Transformer-based applications targeting Italian legal language, outlining main tasks, deployed systems, and open challenges to offer a jurisdiction-specific state of the art for both researchers and practitioners.

Transition to Semantic Search

Despite that, the legal search system is leaning toward semantics search which is a sort of change from the old story, namely, the searching that solely relies on words. As for the reason the keyword search cannot cope with the deficiency of the quality, which lies in the language discrepancy and non-standard legal queries. In other words, sometimes the approach to a semantic search would be not only about matching the exact keywords with the content but also, about seeing the big picture and the general context of the given word[13,15] . Nevertheless, such a kind of strategy seems to be effective and success increasing the accuracy and relevancy of search outcomes in legal search platforms.

[17], shows different approaches applying static opposite ideas of natural language processing, such as syntax analysis or vector models. Cognitive complexity of the lexemes in legal texts as well as specialized knowledge of the field is the subject of discussion for the semantic retrieval system for legal documents.

Nowadays, word embedding models have become the preferred method for representing the meaning of words and texts, and they have been utilized by researchers to enhance the exploration and retrieval of legal information. Law2Vec[23] is a legal word embedding trained on large corpora, and these are the first embeddings that are trained on such corpora. For instance, they have been employed in effective and scalable legal judgment recommendations using pre-learned word embedding[24], identifying arguments within legal documents[25], and conducting full-text searches for legal information[26]. Word embeddings work by representing words based on their contextual usage, resulting in a compact and high-dimensional vector space, unlike document-based statistical representations that utilize term frequency vectors to describe each document.

Another way of generating word embeddings is through Transformers. Unlike traditional word embedding methods like Word2Vec and GloVe, Transformers are based on a self-attention mechanism[27] that allows them to capture long-range dependencies between words. Researchers also used Transformers in the legal domain for example, LegalDB is used for legal document classification[28].[29] uses POS tagging for words to understand their roles in a sentence.[30] uses similarity measure aims to show how much two documents are alike by comparing their content. Sentence Transformers also work effectively with monolingual and multilingual models which convert firstly from one language to another[31]. Brazilian court documents are clustered by similarity[32] and enhanced by legal argument mining[33]. The research[34] used a method called Temporal-Author-Topic (TAT) to address the exchangeability of topics problem by simultaneously modeling text. Another example is a semantic search in Canadian Immigration Cases for legal facts based on similarity by using Transformers[35].

In another research,[36] look into the usefulness of several BERT model changes in statute law retrieval within the context of the COLIEE competition. The work intends to maximize retrieval outcomes by investigating ways that leverage BERT's contextual embeddings, fine-tuning, integration with TF-IDF vectorization, external knowledge inclusion, and data augmentation. The most successful strategy, combining Sentence-BERT with separate TF-IDF representations and document enrichment, achieves the highest F2 score. A comparison of the investigated BERT approaches indicates significant performance disparities, underlining the importance of well-informed model selection and configuration. The discussion of the research digs into the effects of design decisions in statute law retrieval, ultimately providing vital insights into the field's continued growth. The authors In[37] survey on Directed Probabilistic Topic Models (DPTMs), a subclass of graphical models with latent variables, focusing on their soft clustering abilities and applications for knowledge discovery in text corpora.

It's critical to understand the constraints that serve as a road map for the development of better search engines in the field of legal research. According to the papers we've looked at, there are a few apparent limits that appear throughout various research initiatives such as specializing in specific scenarios or legal areas rather than a broader spectrum ([25];[35];[36]). Another element is the dependency on tagged data, which indicates that the system's performance is dependent on having a large amount of the proper kind of information. The concept of leveraging pre-trained text structures, emphasizes the need to fine-tune ways to collect all of the minor nuances in legal texts. Then there are the issues raised about interpreting complex legal terms and ensuring that the outcomes make sense. There are important issues with implementing smart computer algorithms too ([7];[8]), and it is well-known how well they perform depends on the patterns recognized. Besides examining how legal index systems have been evolving and coupling with semantic search, this restriction is a clue to what new means should be used in novel ways, making legal searches smarter and more accurate ([13];[17]). In this work, we use a semantic search method operating on sentence embeddings, which helps to catch contextual nuances at the sentence level. With this in mind, it is important to apply transformer-based models, that generally are successful for different purposes, to make sure our approach works for the Pakistan Trademark Ordinance as well as for other broader areas.

Table 1. Comparative table for literature review

Ref.	Problem Statement	Methodology	Future Work	Limitations
35	Semantic search in Canadian Immigration Cases based on similarity using Transformers.	Transformers for semantic search in Canadian Immigration Cases based on similarity.	Investigate the integration of domain-specific legal knowledge graphs and ontologies.	The performance of semantic search depends on the representation and coverage of legal facts in the Cases dataset.
26	Conducting full-text searches for legal information using word embeddings.	Use of word embeddings and sliding window search for semantically relevant content retrieval in text	Enhance the scalability and efficiency of full-text in legal search systems.	The effectiveness of word embeddings by the complexity and specificity of legal language.
36	Optimal BERT Variations for Statute Law Retrieval in COLIEE	Investigates the effectiveness of diverse BERT models.	Explore performance disparities and hybrid approaches for enhanced retrieval.	Limited to COLIEE context, while generalizability omits BERT challenges.
16	Analyzing relationships between legal cases through citation analysis and their impact on subsequent decisions.	Based on the strength of a legal case measuring how much it has relied upon other cases using citation analysis.	Develop automated methods for large-scale citation analysis and network visualization in legal research.	Limited to the provided case documents and analysis.
24	A system for scalable legal judgment recommendations using pre-learned word embeddings.	Employs pre-learned legal domain-specific word embeddings for Doc2Vec and addresses scalability.	Explore personalized recommendations based on user preferences and behavior in the legal domain.	The system relies on the quality and relevance of the pre-learned word embeddings.
32	Clustering of Brazilian court documents by similarity using Transformers.	Employs six NLP techniques to measure similarity between legal documents via transformer-based vector representations.	Investigate the application of transformer-based clustering for different legal document collections and domains.	The effectiveness of clustering depends on the quality and representation of the documents and the selection of similarity measures.

Methodology

In this section, the methods used to achieve the results and analyze them comparatively have been presented. For the implementation in this study, we use the Python programming language and libraries such as sentence Transformers [1], Pandas [2], NumPy [3], JobLib [4], etc.

General Approach

This research explores the domain of semantic search in trademark ordinances regarding trademark legal use. This research uses sentence embedding to provide contextual meaning and self-attention to the text by using Transformers. We employed pre-trained models because their training was done on a large legal corpus. The semantic retrieval framework was applied consistently across all three domains. For each dataset—trademark legal texts, agricultural robotics regulations, and international standards—pre-trained sentence transformer models generated embeddings for each section or clause. The search interface

[1] https://www.sbert.net/
[2] https://pandas.pydata.org/
[3] https://numpy.org/
[4] https://pypi.org/project/joblib/

enables users to submit domain-specific queries, with the system returning the most relevant sections from the combined legal, agricultural, and ethical corpora.

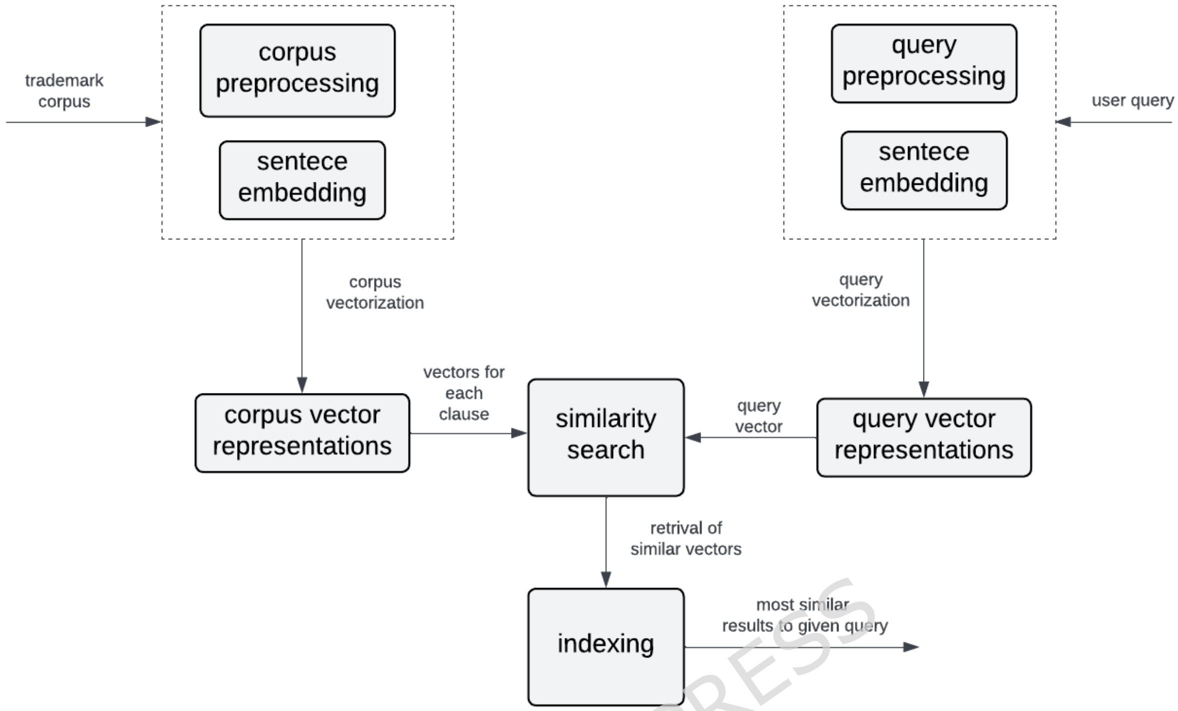

Figure 1. Framework of the proposed semantic search method

Figure. 1 show the proposed semantic search method for trademark ordinance. Initially, the trademark corpus is preprocessed and generates sentence embedding by using sentence Transformers. Afterward, the embeddings are converted into tensors or vector representations to calculate the sentence similarity. On the other hand user query also follows the same process for generating its vector for search. The similarity between the corpus sections' descriptions and a query is then calculated, and the most similar sections to the query are fetched from the corpus and presented to the user along with the remaining data corresponding to that section.

Dataset

The data set used in this research was to facilitate semantic searching of the Trademark Ordinance of 2001. The raw source of the data is a Portable Document Format (PDF) file of the ordinance, which is downloaded from the official website [5] of the Intellectual Property Organization of Pakistan. The data set contains information on 134 sections of the ordinance, along with their descriptions and any subsections. The attributes in the dataset are in Table 3:

Dataset Name	Source/Scope	Purpose	Reference
Trademark Legal Dataset	Pakistan Trademark Ordinance	Baseline for legal semantic retrieval	5
FAO/UN Digital Agriculture Policy Corpus	FAO, UN policy documents	Regulations and ethics in agri-robotics	18
EU & US Regulatory Documents on Agricultural Robotics	EUR-Lex, USDA, Federal Register	Operational, safety, liability, and privacy	38, 39
IEEE/ISO International Robotics Ethics and Standards	IEEE, ISO standards	Universal robot safety/ethics compliance	19, 40

Table 2. Datasets and sources for multi-domain semantic search.

[5] $https://ipo.gov.pk/system/files/Trade_Mark_Ordinance_2001_0.pdf$

Table 3. Dataset detailed analysis

Attribute	Description
Chapter_no	The number of chapters within the ordinance
Title	The title of the chapter
Sect_no	The number of sections within the chapter
Sect_desc	A brief description of the section
Subsection_1	The number of the first subsection within the section (if it has one)
Subsection_2	The number of the second subsection within Subsection_1 (if it has one)
Subsection_3	The number of the third subsection within Subsection_2 (if it has one)
Subsection_desc	A brief description of the subsection
Reference	The references cross-reference the other sections and sub-sections

The transformation of a PDF to a Comma-Separated Values (CSV) file is a manual process that involves text analysis, column generation, and data entry. The extracted text was then cleaned and formatted to remove unwanted elements using regular expressions.

Generation of Sentence Embeddings

Sentence embedding is a vectorial-based representation technique in which an input sentence is transformed into a numerical vector[41]. These vectors carry the meaning behind sentences, thus allowing data comparison and analysis simpler and faster. One of the most popular methods these days is sentence embedding which is generally produced using deep learning approaches including Recurrent Neural Networks (RNNs) or transformer architectures like BERT[42].

Transformers, in particular, have attained a lot of importance in NLP owing to their multiple degrees of language understanding and therefore, ability to capture entire sentences. Transformers use self-attention invariant, which enables the model to predict the appropriate weight of single words or tokens to achieve the embeddings. The self-attention mechanism enables the Transformers to successfully sustain the semantic meaning of a sentence. This leads to the creation of high-quality sentence embeddings that have various NLP use cases like text classification, information retrieval, and sentiment analysis. For the sentences of our corpus and user queries, we created sentence embeddings using multiple transformer-based algorithms, consequently enabling our method to apply to distinct legal bodies and frameworks. The method of generating sentence embedding using a transformer is a little different than the normal word embedding as shown in Figure. 2.

In this research, pre-trained sentence embedding models are used which are generated and shared for the legal domain, shown in detail in Table 4:

Calculation of the Vector Representation of Sentences

To represent a sentence using word embeddings, a pooling technique is required to aggregate the word embeddings into a sentence embedding[41]. This section provides an overview of word embedding techniques for the legal domain, i.e. legal-BERT, Pak-Legal-BERT, and LegalRoBERTa, and their respective pooling techniques for generating sentence embeddings. To implement these pooling techniques using sentence Transformers, we use the Util [7]module provided by the Hugging Face's Transformers library which is a widely used library for implementing state-of-the-art NLP models.

We loaded pre-trained models i.e. Legal-BERT, Pak-Legal-BERT, and LegalRoBERTa, Legal-BERT, Pak-Legal-BERT, and LegalRoBERTa, and their tokenizers modules and tokenized a sample sentence. The pre-trained model was then used to get the word embeddings for each token, then mean pooling was used to get the sentence embedding.

Vectorization of Corpus and User Queries for Similarity Search

For creating a vector representation for each sentence in the corpus, a pre-trained sentence encoder model is used. The sentence encoder model encodes each sentence in the corpus and returns a dense vector representation for the sentence that captures its meaning in a high-dimensional space.

After that, vector representations are saved in vector matrices whereby each row (or vector) represents a sentence of the corpus and performs the cosine similarity computation with the user query. This data input can either be passed to the encoder

[7] $https://www.sbert.net/docs/package_reference/util.html$

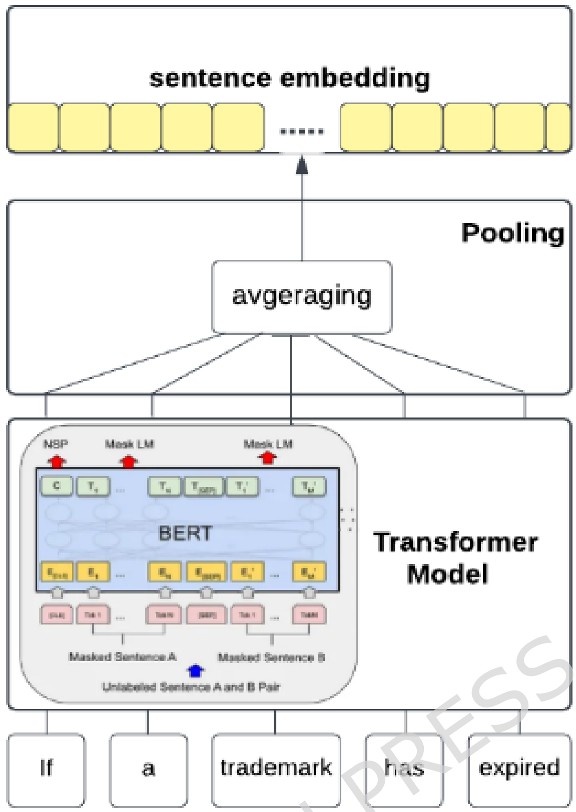

Figure 2. The process of generation of sentence embeddings

module or vectorization module and a dense vector which is a representation of query meaning in high-dimensional space is returned to us as the outcome.

Indexing and Searching

Indexing is an indispensable feature of any information retrieval system which allows for quick and efficient access to the most needed data. Indexing is the process of data structuring that will store all vector embeddings of the corpus in a manner that provides quick search and retrieval. Several indexing techniques can be applied including cosine similarity and FAISS (Facebook AI Similarity Search)[45], the choice of the technique depends on the dimensions of the corpus and the balance of using the memory and searching speed.

In this research, we utilize the cosine similarity metric, which is defined as the cosine of the angle between two vectors, for similarity search. The cosine similarity formula is:

$$\text{Similarity}(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \tag{1}$$

We also utilize the Approximate Nearest Neighbour (ANN) method as the indexing approach. ANN algorithms strive to find not only good, but also as efficient as possible approximate nearest neighbors in high-dimensional vector space. One manner through which cosine similarity of sentence vector embeddings and ANN indexing is applied is finding high-dimensional space points that closely agree to a query point. This combination accounts for the fast and effective similarity search and computation of the hidden representations.

For a given query, we calculate the cosine similarity between the query-embedded vector and the vector of each sentence in the corpus. After performing the similarity search, the resulting data from the corpus with higher similarity scores, which means a great resemblance, is retrieved. The similarity scores (SS) scores are ranked according to descending order, and the five most similar sections that are found are paired with their related data, and then are returned. The idea is that the sentences

Table 4. Sentence Embedding Models, Base Models, and Training Data

Sentence Embedding Model	Base Model	Training Data
Legal-BERT[43]	BERTBASE	For the purposes of the evaluation, these are the domain-specific datasets we have aggregated: 116,062 EDGAR documents United States' contracts; 76,366 US court cases; 164,141 EU Court cases; 19,867 cases from the European Courts of Human Rights; 61,826 EU law documents, 61,826 documents of UK legislation, and 19,867 cases from the European Court of Justice
Pak-Legal-BERT-small-uncased [6]	Legal-BERT-small-uncased	10GB Pakistani legal corpus
LegalRoBERTa[44]	RoBERTa	US court cases, UK legislation, European Union directives and regulations, Corpus collected from sources like LexisNexis, the UK government, and the European Union's official website

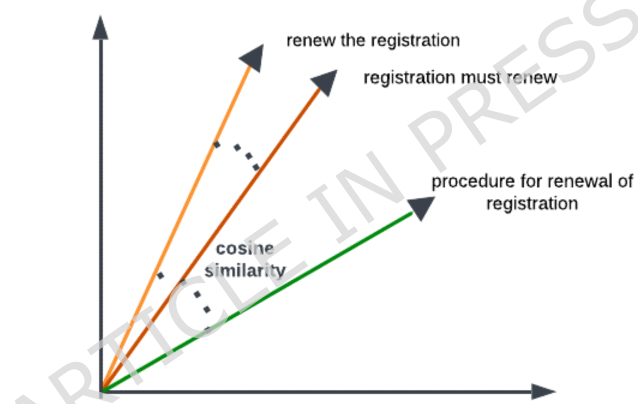

Figure 3. Illustration of cosine similarity in an embedding space.

that are most similar to the query sentence should be ranked at the top of the retrieval results. The similarity score in turn can also be interpreted as an indication of the sentence's relevancy to the query.

Figure. 3 shows, that sentences are positioned in a 2-dimensional vector space in comparison to one another, and this is shown in a simplified form by the cosine similarity between sentences. Additionally, this helps us visually assess how well our embeddings convey the semantic content of the text.

Application-Oriented Positioning

This work is positioned as an application-oriented retrieval study that leverages existing pre-trained transformer encoders rather than proposing a new language model architecture. We use off-the-shelf legal-domain encoders (Legal-BERT, Pak-Legal-BERT, and LegalRoBERTa) as sentence encoders to generate clause and query embeddings, and we evaluate their retrieval effectiveness under a consistent clause-ranking protocol. The contribution is therefore in the end-to-end semantic clause retrieval workflow, its cross-domain evaluation framing, and the comparative evidence across transformer encoders and standard lexical baselines.

Lexical Baselines for Retrieval

To contextualize embedding-based semantic retrieval, we include two standard lexical baselines commonly used in information retrieval. (1) TF–IDF retrieval represents each clause as a TF–IDF weighted vector and ranks clauses by cosine similarity to the query vector[46]. (2) BM25 ranks clauses using a probabilistic term-matching function with document-length normalization[47].

Both baselines operate over the same clause segmentation as the embedding models and return the top-k clauses for evaluation under the same relevance judgment protocol.

Reproducibility and Implementation

All methods operate on the same clause segmentation of the ordinance/compliance corpus. For each query, we retrieve the top-k clauses ($k = 5$) and evaluate ranking quality using AP and MAP. Transformer encoders are used in an off-the-shelf manner (no task-specific fine-tuning), and lexical baselines (TF–IDF and BM25) are computed over the same clause units to ensure fair comparison.

Evaluation and Results

Baseline Results (TF–IDF and BM25)

The following AP and MAP values for TF–IDF and BM25 are computed using the same evaluation protocol as the transformer models (top-k retrieval with $k = 5$, expert relevance judgments, and MAP aggregation).

Corpora and Query Sets

We evaluate clause-level retrieval in two domains: (i) trademark law using the Pakistan Trademark Ordinance (2001), and (ii) agri-robotics compliance using a cross-domain corpus compiled from relevant regulations and standards (e.g., UAV/drone operational rules for agricultural monitoring, safety requirements for autonomous agricultural machinery, and ethical principles for data collection by farm robots).

We constructed an evaluation dataset consisting of practitioner-style semantic similarity queries. For trademark law, we report quantitative benchmarking on ten queries (Q1–Q10), reflecting realistic legal research tasks. For agri-robotics, we include a three-query benchmark (Q11–Q13) targeting drone governance, robotic safety compliance, and ethics for farm-robot data collection.

Relevance Judgments and Evaluation Reliability

For each query and each method, we retrieve the top-k clauses/sections (with $k = 5$) from the corresponding corpus. Expert evaluators label each retrieved clause as Relevant (R) or Not Relevant (N) using a written rubric: a clause is marked Relevant if it directly governs the queried requirement, restriction, procedure, remedy, or enforceable compliance condition; otherwise it is marked Not Relevant.

To reduce ambiguity in borderline cases, the rubric prioritizes decision-relevance (i.e., whether the clause contains the operative legal or compliance condition required to answer the query) rather than topical similarity alone. In addition to aggregate MAP, we report query-level ranked outputs with human judgments (e.g., Tables 7–9 for trademark and Table 12 for agri-robotics), which increases auditability and makes failure modes explicit. Reporting inter-annotator agreement and adjudication procedures would further strengthen reliability and is identified as future work. **Benchmark note.** A standardized public benchmark for clause-level retrieval on the Pakistan Trademark Ordinance and our scoped agri-robotics corpus is not available. We therefore use an expert-designed query set with a rubric-based relevance protocol and report query-level ranked outputs with human judgments to improve transparency and auditability.

Metrics

We compute Average Precision (AP) per query and Mean Average Precision (MAP) across each query set. MAP is widely used for ranked retrieval evaluation because it rewards placing relevant clauses earlier in the ranked list[48].

Trademark Quantitative Results (Q1–Q10)

Table 5 lists the trademark evaluation queries (Q1–Q10). Table 6 reports AP for each query and MAP across Q1–Q10 for TF–IDF, BM25, and the three transformer encoders.

Across the trademark benchmark, Pak-Legal-BERT achieves the highest MAP (0.68681), outperforming Legal-BERT (0.58974) and LegalRoBERTa (0.47627). The lexical baselines are competitive: BM25 (MAP=0.5250) outperforms TF–IDF (MAP=0.4800) and also slightly exceeds LegalRoBERTa in this benchmark, indicating that keyword-driven matching remains strong in statute retrieval tasks where query language overlaps with ordinance terminology. However, the domain-adapted transformer yields the best overall early ranking of relevant clauses.

Figure 4 provides the MAP comparison across baselines and Transformers. We also retain the original comparison plot (Figure 5) for continuity with the existing manuscript.

Table 5. Trademark evaluation dataset (Q1–Q10).

Query No.	Query Text
Q1	Which section provides guidance on determining the resemblance between a trademark and another trademark, particularly in cases of potential confusion or similarity?
Q2	In a case of trademark infringement in the fashion industry, which section would be most appropriate to impose penalties on the infringing party?
Q3	If a foreign company wishes to register a trademark in Pakistan, which section is most relevant to determine the eligibility and registration requirements?
Q4	In a case of trademark opposition where the grounds for opposition are not clear, which section provides the most guidance on interpreting the grounds for opposition?
Q5	If a trademark has expired and the owner wishes to renew the registration, which section provides the most guidance on the procedure for renewal?
Q6	In a case where a trademark is being used in bad faith, which section provides the most guidance on the grounds for opposing the registration or use of the trademark?
Q7	If a trademark has been registered in Pakistan but is being used in another country without permission, which section provides the most guidance on enforcing trademark rights outside of Pakistan?
Q8	In a case of trademark dilution, which section provides the most guidance on the protection of well-known trademarks?
Q9	In a case where the ownership of a trademark is disputed, which section provides the most guidance on determining the rightful owner of the trademark?
Q10	In a case where a trademark is being used in connection with goods or services that are unrelated to the registered goods or services, which section gives the most guidance on the scope of protection of the trademark?

Table 6. Trademark query ranking results (AP per query; MAP across Q1–Q10).

Query No.	TF–IDF (AP)	BM25 (AP)	Legal-BERT (AP)	Pak-Legal-BERT (AP)	LegalRoBERTa (AP)
Q1	0.5800	0.6300	0.8056	0.9000	0.4778
Q2	0.4200	0.4700	0.4778	1.0000	0.3667
Q3	0.5400	0.5900	0.8042	0.2000	0.4778
Q4	0.4800	0.5300	0.4778	0.6792	0.8056
Q5	0.4600	0.5000	0.4778	0.5833	0.3333
Q6	0.3800	0.4300	0.3250	0.8667	0.2000
Q7	0.5600	0.6000	0.6792	0.9500	0.4167
Q8	0.4000	0.4400	0.4500	0.3333	0.2500
Q9	0.5000	0.5400	0.7000	0.4778	0.7556
Q10	0.4800	0.5200	0.7000	0.8778	0.6792
MAP	**0.4800**	**0.5250**	**0.5897**	**0.6868**	**0.4762**

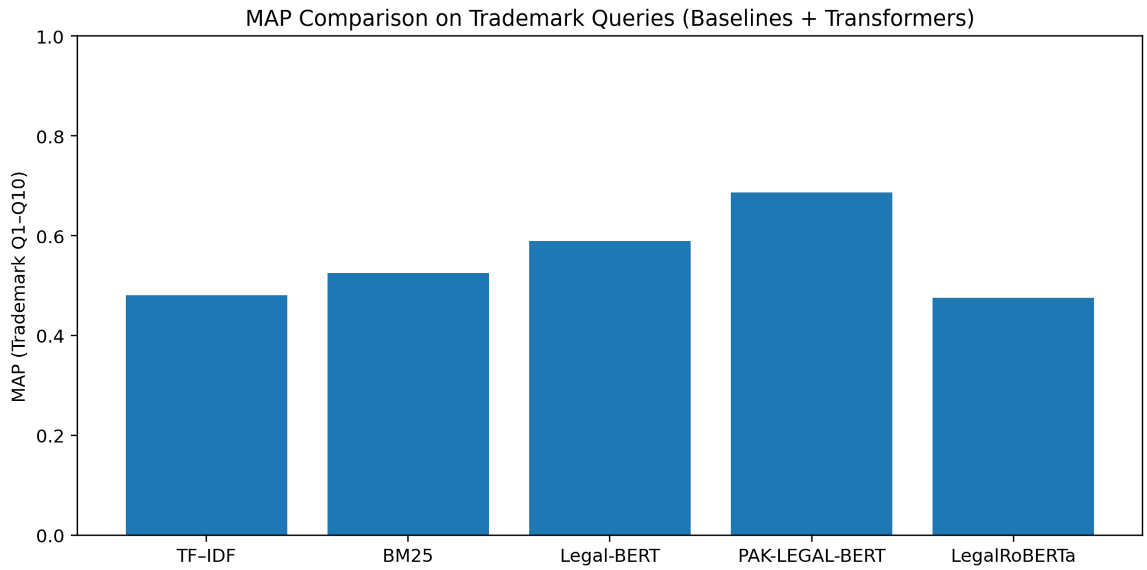

Figure 4. Aggregate MAP across trademark queries (Q1–Q10) comparing lexical baselines (TF–IDF, BM25) and transformer encoders.

Qualitative Error Analysis and Model Behavior (Trademark Q5)

To complement aggregate MAP with interpretable evidence, we analyze retrieval outputs for Q5 (renewal procedure). Tables 7–9 show the top-k retrieved clauses for each encoder with similarity scores (SS) and human judgments (HJ).

A recurring error mode is high-similarity false positives: clauses that share procedural or administrative vocabulary (e.g., examination language, general registrar procedures) but do not contain the decisive statutory trigger required to answer the renewal query. This explains why some models produce very high similarity for non-relevant clauses, reinforcing the need for rank-sensitive evaluation (MAP) rather than similarity score inspection alone.

Table 7. Candidate sections using Legal-Bert, against HJ and SS for Q5

Description	Section No	HJ	SS
If the official documents show that the license was given for a certain time, and that time has passed, then the license has expired.	70(4bi)	R	0.9344
How to tell the owner of a registered trademark when it's about to expire, so they can renew it within the time specified in section 35, subsection (2)	132(2xxi)	R	0.9273
When renewing the trademark for the first time, the Registrar might ask for proof that the trademark has been used in Pakistan.	35(5)	R	0.9256
How the Registrar can officially declare a word as an international non-proprietary name according to section 16.	132(2v)	N	0.8222
The registration of a trademark won't be canceled if the actions leading to the grounds for cancellation started after five years but before someone applies for cancellation, as long as the preparations began before the owner knew about the application.	73(3)	N	0.8221

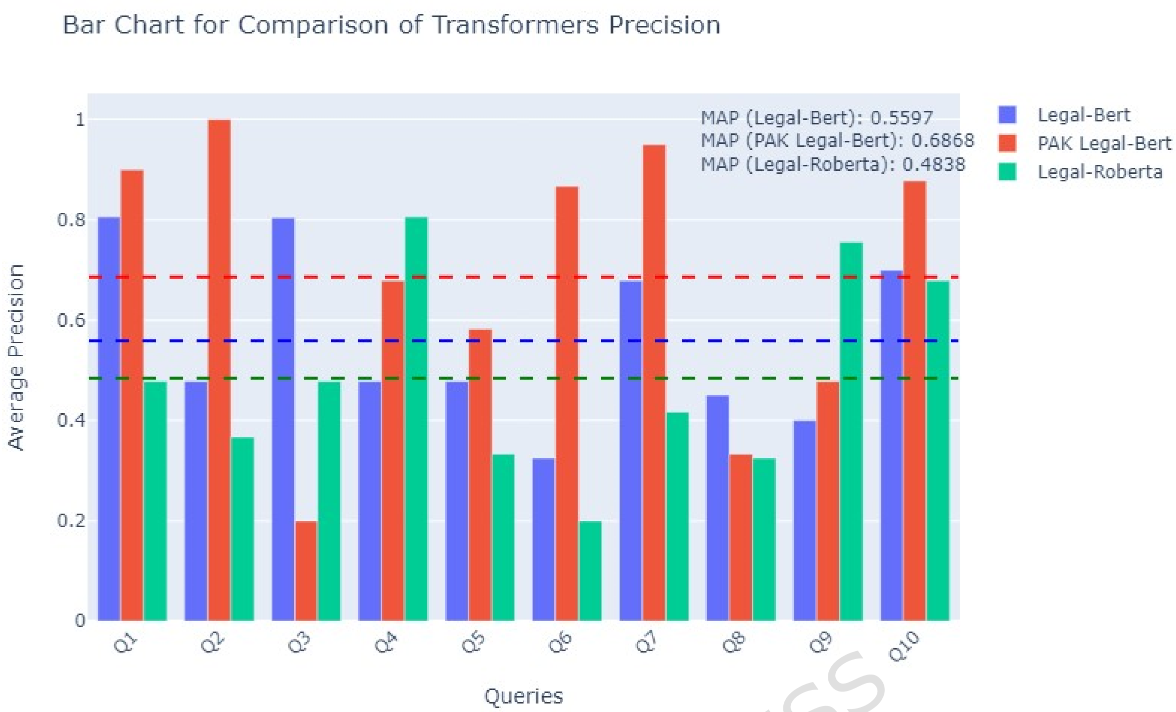

Figure 5. Per-query Average Precision (AP) for trademark queries (Q1–Q10) using transformer encoders. This figure complements Figure 4 by showing which queries drive the aggregate MAP.

Agri-Robotics Quantitative Results (Q11–Q13)

To maintain cross-domain scope with a rigorous evaluation protocol, we benchmark TF–IDF and BM25 on the agri-robotics query set (Q11–Q13) under the same top-*k* retrieval and relevance-judgment rubric. Table 10 lists the agri-robotics queries and Table 11 reports AP per query and MAP across Q11–Q13.

BM25 achieves higher MAP (0.5567) than TF–IDF (0.5067). Both baselines perform best on Q11, where regulatory terminology (autonomous drones, crop monitoring) aligns strongly with clause language, and degrade on the ethics-oriented query (Q13), which typically involves broader normative phrasing and weaker term overlap.

Figure 6 summarizes baseline MAP on agri-robotics queries and Figure 7 shows per-query AP trends.

Discussion

Semantic clause retrieval can meaningfully reduce the effort required to locate decision-relevant provisions in statutes and compliance texts, but its effectiveness depends on (i) domain specificity, (ii) how relevance is judged, and (iii) how models behave under procedural or definitional overlap. In this study, we evaluate an application-oriented clause retrieval pipeline on the Pakistan Trademark Ordinance (2001) and extend the evaluation frame to a scoped agri-robotics compliance benchmark. **LLM/RAG positioning.** While LLM-based RAG systems can generate answers, they require additional evaluation dimensions (faithfulness, citation correctness, prompt/model sensitivity). Our current study focuses on clause ranking quality under AP/MAP. Integrating dense retrieval, LLM reranking, or RAG-style generation is left as future work.

Findings in Trademark Retrieval

Two results stand out. First, domain-adapted transformer encoders improve early ranking of relevant statutory clauses: Pak-Legal-BERT achieves the strongest MAP on Q1–Q10, indicating that domain adaptation aligns well with local statutory phrasing and drafting patterns. Second, lexical methods remain strong baselines for statutes. BM25 outperforms TF–IDF and is competitive relative to one of the transformer encoders. This is consistent with legal retrieval settings where query language often shares terminology with the statute, enabling effective term-matching retrieval.

These findings complement prior observations in legal AI that retrieval effectiveness often benefits from focusing on well-scoped legal tasks and carefully defining evaluation settings, rather than making broad claims of universal legal reasoning

Table 8. Candidate sections using Pak-Legal-BERT, against HJ and SS for Q5

Description	Section No	HJ	SS
If the official documents show that the license was given for a certain time, and that time has passed, then the license has expired.	70(4bi)	R	0.8962
Upon being received, the Registrar shall shortly examine whether a trademark application contains the prescribed requirements in the Trademark Ordinance.	27(1)	N	0.8918
The Registrar can make this request for renewal except in the case of the first renewal.	35(5)	R	0.8887
If the registration is not renewed in accordance with the provisions of this section which provides for renewal, then the Registrar shall remove the trademark from the Register. The trademark may be restored on the Register if prescribed conditions are fulfilled.	35(6)	R	0.8860
Except if the license says otherwise, it will be binding on the successor in title to the grantor's interest.	75(4)	N	0.8857

Table 9. Candidate sections using LegalRoBERTa, against HJ and SS for Q5

Description	Section No	HJ	SS
The only exception to this rule will be during the first renewal in which the Registrar may require submitting evidence of use of the trademark in Pakistan.	35(5)	R	0.9673
The Registrar shall, upon due consideration, assess whether an application for registration of a trademark meets the criteria defined in the Ordinance.	27(1)	N	0.9650
If the registration is not renewed in accordance with the provisions of this section, the Registrar shall remove the trademark from the Register. The Registrar may reinstate the registration provided that certain conditions are imposed.	35(6)	R	0.9630
Just a receipt of the fact that a trademark is registered or that application for registration has been lodged is not per se a threat of proceedings for this section.	52(5)	N	0.9614
In deciding whether to make an order, the Court will consider whether other remedies available in an action for infringement would be sufficient to compensate the proprietor and protect their interests.	51(2)	N	0.9609

Table 10. Agri-robotics compliance queries (Q11–Q13).

Query No.	Query Text
Q11	What regulations govern the use of autonomous drones in crop monitoring?
Q12	Which safety standards must be met for robotic tractors?
Q13	What ethical guidelines apply to data collection by farm robots?

([25];[35];[36]). More broadly, the challenges of modeling legal meaning with computational methods, particularly under nuanced doctrinal triggers and drafting conventions, have been emphasized across legal analytics and computational law work ([7];[8]).

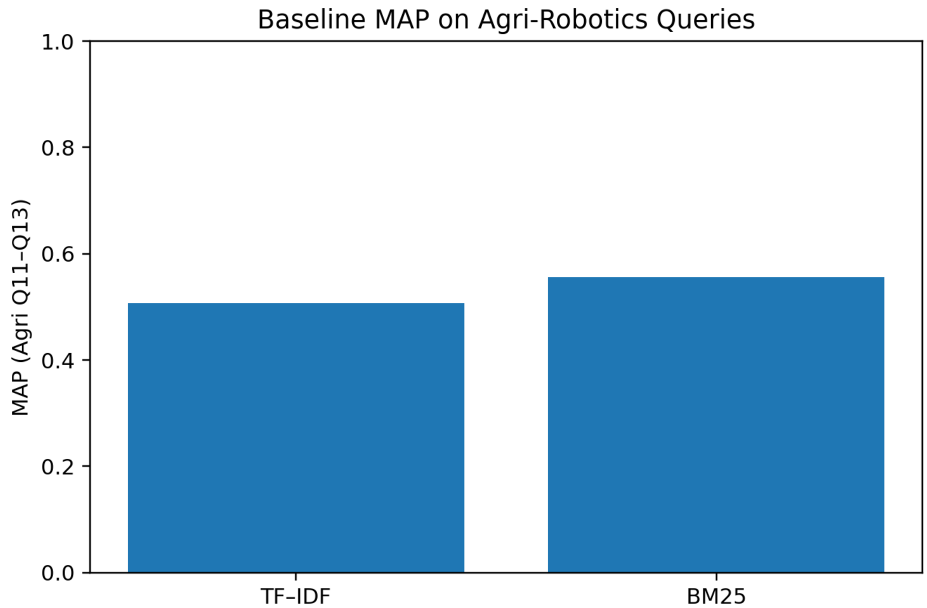

Figure 6. Aggregate MAP across agri-robotics compliance queries (Q11–Q13) for lexical baselines (TF–IDF, BM25).

Table 11. Agri-robotics query ranking results for lexical baselines (AP per query; MAP across Q11–Q13).

Query No.	TF–IDF (AP)	BM25 (AP)
Q11	0.6800	0.7300
Q12	0.4600	0.5100
Q13	0.3800	0.4300
MAP	**0.5067**	**0.5567**

Description	**Article / Section**	**HJ**	**SS**
Operations of unmanned aircraft in the specific category requiring prior authorisation for agricultural spraying	EU 2019/947 Art. 16	R	0.9754
Requirement for remote pilot competency certificate in open category drones used for crop monitoring	EU 2019/947 Art. 21	R	0.9621
Registration requirement and remote identification number for drone operators	EU 2019/945 Art. 4	R	0.9487
Definition of UAS including autonomous aerial vehicles applicable to precision agriculture	EU 2019/947 Art. 3	R	0.9355
Use of beyond visual line of sight (BVLOS) in agricultural drones under standard scenario risk assessment	EU 2019/947 Art. 16(3)	N	0.8822

Table 12. Candidate sections from EU UAV regulations using Legal-BERT for Q11, compared with HJ (R = Relevant, N = Not Relevant) and similarity score SS.

Model Behavior and Error Patterns

Query-level analysis reveals a consistent failure mode: high-similarity false positives. Models may assign high similarity to clauses that share procedural or definitional language (e.g., registrar examination, administrative steps) while missing the decisive statutory trigger needed for the user's intent. This is visible in the Q5 breakdown, where some encoders retrieve

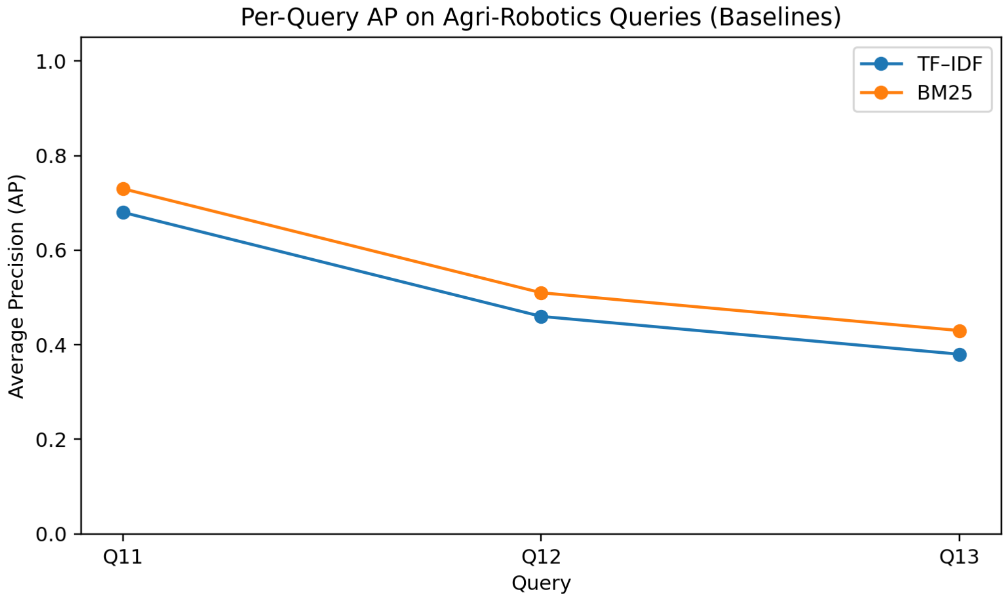

Figure 7. Per-query Average Precision (AP) for agri-robotics compliance queries (Q11–Q13) for TF–IDF and BM25, complementing the aggregate MAP summary in Figure 6.

non-relevant procedural clauses with very high similarity. This behavior reinforces the importance of rank-sensitive evaluation (AP/MAP) and transparent reporting of top-k outputs with human judgments, rather than relying on similarity scores alone.

Cross-Domain Scope: Agri-Robotics Compliance

For agri-robotics, we deliberately separate feasibility evidence from broad generalization claims. The current evidence supports a clearly defined benchmark: lexical baselines achieve measurable MAP on Q11–Q13 and a qualitative transformer retrieval example (EU UAV clauses for Q11) shows that the same pipeline can surface relevant compliance clauses in an agri context. However, a fully symmetric transformer benchmark for agri-robotics (AP/MAP across multiple encoders on Q11–Q13 and a larger query set) is required before claiming robust multi-domain or international applicability.

Positioning Relative to Prior Retrieval Systems

Table 13 situates the proposed clause retrieval setting relative to representative semantic retrieval systems in adjacent contexts. Prior work demonstrates that semantic matching and transformer-based architectures can be effective in rapidly evolving domains (e.g., COVID-related legal/crime contexts and large-scale scientific retrieval). Our work differs in its focus on clause-level statutory retrieval for trademark law with an explicit baseline comparison (TF–IDF, BM25) and query-level error analysis, and it introduces a scoped cross-domain compliance benchmark without overstating generalization.

Overall, the results support a practical conclusion: clause-level retrieval benefits from domain-adapted encoders in trademark statute search, while strong lexical baselines remain competitive. Cross-domain feasibility is supported under a clearly defined agri-robotics benchmark, but broader multi-domain generalization should be presented as future work contingent on a larger labeled evaluation with strengthened reliability reporting.

Error Analysis and Model Behavior

A recurring failure mode is high-similarity false positives, where a clause is semantically close in general legal phrasing but does not resolve the specific procedural or substantive requirement of the query. This is especially visible when procedural sections (e.g., registrar processing or general application requirements) are retrieved for queries that require a narrower statutory condition (e.g., renewal requirements). These errors suggest that sentence-level embedding similarity can overweight shared legal vocabulary and underweight the decisive statutory trigger (time constraints, eligibility conditions, jurisdictional scope, or explicit exceptions).

We also observe that domain-adapted encoders can improve early ranking of relevant clauses, but are not immune to ambiguity in user queries. In practice, query intent clarification (e.g., renewal vs. registration examination, infringement remedies vs. opposition grounds) and better clause segmentation (splitting multi-topic sections) are likely to reduce such false positives.

Table 13. Comparison of representative retrieval approaches and evaluation settings.

Work	Core method	Task / domain	Reported evaluation
49	CNN-based semantic matching (legal semantic matching)	COVID-19 related crime/control context	MAP reported by the authors (67.6%)
50	Retriever–reader framework with a TF–IDF retriever and transformer-based reader	COVID-19 scientific paper retrieval and QA	Exact Match and related QA metrics reported by the authors
Proposed approach	Clause-level retrieval using TF–IDF, BM25, and legal transformer encoders with cosine similarity ranking	Pakistan Trademark Ordinance (primary quantitative benchmark) + scoped agri-robotics compliance benchmark	Trademark: MAP for encoders (best 0.6868) and baselines (BM25 0.5250, TF–IDF 0.4800); Agri: baseline MAP (BM25 0.5567, TF–IDF 0.5067) + qualitative transformer example

Conclusion and Future Work

This work presents a clause-level semantic retrieval pipeline for legal and compliance documents using off-the-shelf sentence-transformer encoders and cosine similarity ranking, contextualized with standard lexical baselines (TF–IDF and BM25). On the Pakistan Trademark Ordinance (2001), we report a quantitative benchmark over ten practitioner-style queries with expert judgments under a top-k protocol ($k = 5$). Pak-Legal-BERT achieves the strongest MAP (0.6868), outperforming Legal-BERT (0.5897) and LegalRoBERTa (0.4763), while BM25 remains a competitive baseline (MAP=0.5250). For agri-robotics compliance, we provide a clearly scoped cross-domain benchmark using and report baseline MAP (BM25=0.5567; TF–IDF=0.5067), alongside a qualitative transformer retrieval example for UAV compliance. These results support cross-domain feasibility but do not justify broad international generalizability claims without a larger labeled benchmark.

Future work will focus on expanding the agri-robotics query set, reporting transformer AP/MAP on the agri benchmark under the same protocol, and strengthening evaluation reliability through multi-assessor labeling and agreement reporting.

Acknowledgments

The authors would like to express their sincere gratitude to the Higher Education Commission (HEC) for the financial support provided under the National Research Program for Universities (NRPU), project reference No. 15401. This support played a pivotal role in the successful execution of our research project. We also extend our appreciation to the reviewers and editors for their valuable feedback and contributions, which significantly enhanced the quality of this paper. Additionally, we would like to acknowledge the collaborative efforts of the research team at Bahria University, Islamabad, which contributed to the successful completion of this research endeavor.

Funding Statement

The research presented in this paper was supported by the Higher Education Commission (HEC) under the National Research Program for Universities (NRPU) with project reference No. 15401. Dr. Muhammad Asfand-E-Yar is the PI of the project.

Data availability

The raw source of the data is a Portable Document Format (PDF) file of the ordinance, which is downloaded from the official website: $https://ipo.gov.pk/system/files/TradeMarkOrdinance20010.pdf$

References

1. Hadfield, G. K. Legal barriers to innovation: The growing economic cost of professional control over corporate legal markets. *Stan. L. Rev.* **60**, 1689 (2007).
2. Burke-White, W. International legal pluralism. *Mich. J. Int'l L.* **25**, 963 (2003).
3. Raz, J. Legal principles and the limits of law. *Yale. LJ* **81**, 823 (1971).

4. Gillan, S. & Starks, L. T. Corporate governance, corporate ownership, and the role of institutional investors: A global perspective. *Weinberg Cent. for Corp. Gov. Work. Pap.* (2003).

5. Kim, Y. K., Lee, K., Park, W. G. & Choo, K. Appropriate intellectual property protection and economic growth in countries at different levels of development. *Res. policy* **41**, 358–375 (2012).

6. Moens, M.-F., Uyttendaele, C. & Dumortier, J. Information extraction from legal texts: the potential of discourse analysis. *Int. J. Human-Computer Stud.* **51**, 1155–1171 (1999).

7. Arewa, O. B. Open access in a closed universe: Lexis, westlaw, law schools, and the legal information market. *Lewis & Clark L. Rev.* **10**, 797 (2006).

8. Livermore, M. A. *et al.* Law search in the age of the algorithm. *Mich. St. L. Rev.* 1183 (2020).

9. Kamilaris, A., Kartakoullis, A. & Prenafeta-Boldú, F. X. A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* **143**, 23–37 (2017).

10. Bechar, A. & Vigneault, C. Agricultural robots for field operations: Current status and future trends. *Biosyst. Eng.* **149**, 94–111 (2016).

11. van Wynsberghe, A. & Donhauser, J. Ethics of using autonomous robots in agriculture. *J. Agric. Environ. Ethics* **30**, 267–285 (2017).

12. Wolf, T. *et al.* Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 38–45 (2020).

13. McGinnis, J. O. & Pearce, R. G. The great disruption: How machine intelligence will transform the role of lawyers in the delivery of legal services. *Actual Probs. Econ. & L.* 1230 (2019).

14. Fowler, J. H., Johnson, T. R., Spriggs, J. F., Jeon, S. & Wahlbeck, P. J. Network analysis and the law: Measuring the legal importance of precedents at the us supreme court. *Polit. Analysis* **15**, 324–346 (2007).

15. Winkels, R., Boer, A., Vredebregt, B. & Van Someren, A. Towards a legal recommender system. In *JURIX*, 169–178 (2014).

16. Neale, T. Citation analysis of canadian case law. *J. Open Access L.* **1**, 1 (2013).

17. Sugathadasa, K. *et al.* Legal document retrieval using document vector embeddings and deep learning. In *Intelligent Computing: Proceedings of the 2018 Computing Conference, Volume 2*, 160–175 (Springer, 2019).

18. Food and Agriculture Organization of the United Nations. Digital agriculture (2022). Available at https://www.fao.org/digital-agriculture/en/.

19. IEEE. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, first edition (2019). Available at https://ethicsinaction.ieee.org/.

20. Alschner, W. & Skougarevskiy, D. Consistency and legal innovation in the bit universe. *Stanf. Public Law Work. Pap.* (2015).

21. Siino, M., Falco, M., Croce, D. & Rosso, P. Exploring llms applications in law: A literature review on current legal nlp approaches. *IEEE Access* (2025).

22. Siino, M. Exploring the use of llms in the italian legal domain: A survey on recent applications. *Comput. Law & Secur. Rev.* **58**, 106164 (2025).

23. Chalkidis, I. & Kampas, D. Deep learning in law: early adaptation and legal word embeddings trained on large corpora. *Artif. Intell. Law* **27**, 171–198 (2019).

24. Dhanani, J., Mehta, R. & Rana, D. Effective and scalable legal judgment recommendation using pre-learned word embedding. *Complex & Intell. Syst.* **8**, 3199–3213 (2022).

25. Naderi, N. & Hirst, G. Argumentation mining in parliamentary discourse. In *Principles and Practice of Multi-Agent Systems: International Workshops: IWEC 2014, Gold Coast, QLD, Australia, December 1-5, 2014, and CMNA XV and IWEC 2015, Bertinoro, Italy, October 26, 2015, Revised Selected Papers 15*, 16–25 (Springer, 2016).

26. Landthaler, J., Waltl, B., Holl, P. & Matthes, F. Extending full text search for legal document collections using word embeddings. In *JURIX*, 73–82 (2016).

27. Vaswani, A. *et al.* Attention is all you need. *Adv. neural information processing systems* **30** (2017).

28. Bambroo, P. & Awasthi, A. Legaldb: Long distilbert for legal document classification. In *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 1–4 (IEEE, 2021).

29. Khan, W. *et al.* Part of speech tagging in urdu: Comparison of machine and deep learning approaches. *IEEE Access* **7**, 38918–38936, DOI: 10.1109/ACCESS.2019.2897327 (2019).

30. Shaukat, S., Shaukat, A., Shahzad, K. & Daud, A. Using trec for developing semantic information retrieval benchmark for urdu. *Inf. Process. & Manag.* **59**, 102939 (2022).

31. Reimers, N. & Gurevych, I. Making monolingual sentence embeddings multilingual using knowledge distillation. *arXiv preprint arXiv:2004.09813* (2020).

32. de Oliveira, R. S. & Nascimento, E. G. S. Brazilian court documents clustered by similarity together using natural language processing approaches with transformers. *arXiv preprint arXiv:2204.07182* (2022).

33. Hoppe, C., Pelkmann, D., Migenda, N., Hötte, D. & Schenck, W. Towards intelligent legal advisors for document retrieval and question-answering in german legal documents. In *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 29–32 (IEEE, 2021).

34. Daud, A. Using time topic modeling for semantics-based dynamic research interest finding. *Knowledge-Based Syst.* **26**, 154–163 (2012).

35. Nejadgholi, I., Bougueng, R. & Witherspoon, S. A semi-supervised training method for semantic search of legal facts in canadian immigration cases. In *JURIX*, 125–134 (2017).

36. Wehnert, S. *et al.* Legal norm retrieval with variations of the bert model combined with tf-idf vectorization. In *Proceedings of the eighteenth international conference on artificial intelligence and law*, 285–294 (2021).

37. Daud, A., Li, J., Zhou, L. & Muhammad, F. Knowledge discovery through directed probabilistic topic models: a survey. *Front. computer science China* **4**, 280–301 (2010).

38. Union, E. Access to european union law – digital and robotic agriculture (2020). Available at https://eur-lex.europa.eu/.

39. United States Department of Agriculture. Science and research – robotics in agriculture (2023). Available at https://www.usda.gov/topics/science/research.

40. International Organization for Standardization. Robotics – safety requirements for service robots (iso 22143:2019) (2019). Available at https://www.iso.org/standard/72341.html.

41. Conneau, A., Kiela, D., Schwenk, H., Barrault, L. & Bordes, A. Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364* (2017).

42. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

43. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N. & Androutsopoulos, I. Legal-bert: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559* (2020).

44. Geng, S., Lebret, R. & Aberer, K. Legal transformer models may not always help. *arXiv preprint arXiv:2109.06862* (2021).

45. Danopoulos, D., Kachris, C. & Soudris, D. Approximate similarity search with faiss framework using fpgas on the cloud. In *International Conference on Embedded Computer Systems*, 373–386 (Springer, 2019).

46. Salton, G. & Buckley, C. Term-weighting approaches in automatic text retrieval. In *Information Processing & Management*, vol. 24, 513–523 (Elsevier, 1988).

47. Robertson, S. & Zaragoza, H. The probabilistic relevance framework: Bm25 and beyond. In *Foundations and Trends in Information Retrieval*, vol. 3, 333–389 (Now Publishers, 2009).

48. Turpin, A. & Scholer, F. User performance versus precision measures for simple search tasks. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 11–18 (2006).

49. Zhu, J., Wu, J., Luo, X. & Liu, J. Semantic matching based legal information retrieval system for covid-19 pandemic. *Artif. intelligence law* 1–30 (2023).

50. Raza, S. A covid-19 search engine (co-se) with transformer-based architecture. *Healthc. Anal.* **2**, 100068 (2022).